Proceedings

(papers in English)

9th Iranian Statistics Conference

20-22 August 2008

Department of Statistics

University of Isfahan

Isfahan, Iran

Foreword

The Proceedings of the 9th Iranian Statistical Conference, published in two volumes, is the product of sustained endeavors made by members of the Scientific and Organizing Committees, scientific advisers, several students and personnel of the Department of Statistics on one hand, and the authors of papers on the other hand. Manuscripts appearing in this volume are invited and accepted contributed papers which are written in English, arranged in alphabetic order of their first author. Papers written in Farsi are appeared in the ad joint Farsi volume.

It should be noted that these papers are chosen by the referees and the Scientific Committee from 185 out of 600 papers whose abstracts had been accepted for oral presentation. Invited papers have not been refereed. However, due to time limitations, no attempts have been made to have the papers revised by their authors or the Scientific Committee, except for some minor editing corrections. Therefore, the authors of papers are responsible for the content of these manuscripts. Further, for the same reasons, some contributed papers that have not reached the conference office on time or have not fulfilled the conference requirements are not appeared in the Proceedings.

It is acknowledged that the Conference is held with joint cooperation of the Iranian Statistical Society, the Statistical Research Center of Iran, the Central Bank of Iran and the Municipality of Isfahan.

Among others, sincere thanks are due to Mr. Ehsan Karimzadeh for his effort and expertise that made the proceedings available on time.

Conference Organizers

Members of Organizing and Scientific Committees

Organizing Committee

1. Dr. Mohammad Hossein Alamatsaz	University of Isfahan
2. Dr. Majid Asadi (Chairman of the conference)	University of Isfahan
3. Dr. Mohammad Bahrami	University of Isfahan
4. Mr. Nasrollah Iranpanah	University of Isfahan
5. Dr. Iraj Kazemi	University of Isfahan
6. Dr. Manouchehr Kheradmandnia	University of Isfahan
7. Mr. Ahmad Noorolah	University of Isfahan
8. Dr. Afshin Parvardeh	University of Isfahan
9. Dr. Arash Shahin	University of Isfahan
10. Dr. Mahmood Taheri	Isfahan University of Technology
11. Dr. Hooshang Talebi	University of Isfahan

Scientific Committee

1. Dr. Mohammad Hossein Alamatsaz (chairman)	University of Isfahan
2. Dr. Naser Reza Arghami	Ferdowsi University of Mashad
3. Dr. Majid Asadi	University of Isfahan
4. Dr. Abbas Gerami	University of Isfahan
5. Dr. Iraj Kazemi	University of Isfahan
6. Dr. Nader Nematolahi	Allameh Tabatabaei University
7. Dr. Bahram Soleymani	Isfahan University of Medical Sciences
8. Dr. Houshang Talebi	University of Isfahan

Contents

Some Results on Dynamic Rényi Entropy,
M. Abbasnejad, N. R. Arghami **1**

Multistep Forecasting Non-Stationary Time Series Using Wavelets and Kernel Smoothing,
M. Aminghafari, and Jean-Michel Poggi **15**

Fisher information in record values and their concomitants under the Gumbel's bivariate exponential distribution,
M. Amini, J. Ahmadi **26**

Jarque Bera Normality Test and Its Modification,
S. M. Amir Jahanshahi, et. al **38**

Approximate Estimators for the Scaled Burr XII Distribution Under Progressive Type-II Censoring,
A. Asgharzadeh, R. Valiollahi **56**

Maximum Inactivity Entropy Models,
S. Ashrafi **68**

Probabilistic normed space and statistical convergent,
P. Azhdari **80**

On $M/(G_1, G_2)/1/G(BS)/V_s$ with optional second service,
A. Badamchi zadeh **88**

State Dependence and Feedback Effects in Individual Poverty Histories,
M. Biewen **102**

The Bayesian Testing for Weighted Distributions,
R. Chinipardaz, S. M. R. Alavi **111**

A Bayesian approach for analyzing influential factors on unemployment status using a random coefficient transitional

binary logistic model with non-monotone missing pattern,
S. Eftekhari Mahabadi, M. Ganjali **125**

The asymptotic properties of some discrete distributions generated by standard skewed stable laws
D. Farbod, K. V. Gasparian **138**

Step-Stress Accelerated Life Test Planning for Life Data Analysis,
N. Fard, Chenhua Li **145**

A Monte Carlo Method In Application To Estimate Extinction Probability of Bisexual Galton-Watson Process,
R. Farnoosh, Z. Zarabi Zadeh **157**

Third order asymptotic behavior of Rao score test for inhomogeneous Poisson processes,
Kh. Fazli **163**

Detection of Outliers and Leverage points and Robust Regression Analysis of Iranian Urban Household Income and Expenditure Data,
M. Ganjali, and S. Eftekhari Mahabadi **175**

Testing Independence: A Bayesian Predictive Approach,
A. Ghalamfarsa Mostofi **190**

On Some Statistical Aspects of Agreement Among Measurements,
S. Hedayat **200**

D_b-optimal Design for Rank-Order Nested Logit Models in discrete choice experiment,
H. Jafari **211**

Search when the lie depends on the target,
Gyula O.H. Katona **222**

Bayesian Estimation of the Parameters of the Bivariate Poisson Regression Models,
S. Khafri, et. al **233**

Inference Bayesian Nonparametric of densities under constrains of form,
S. Khazaei **245**

Exact likelihood equations for multivariate symmetric Kotz-type autoregressions,
Z. Khodadadi, B. Tarami **259**

Recurrence relations for distributions of skew-t and a linear combination of order statistics from bivariate-t,
M. Khosravi, A. Jamalizadeh **272**

Regression model diagnostics under long memory,
H. L. Koul **282**

Bayesian Model Selection for ARCH Models Using Iterated Importance Sampling,
F. Lak **299**

A.P.O. Rules are Asymptotically Non-Deficient for Estimation under Linex Loss Function,
E. Mahmoudi **312**

Large Efficiency Issues in Generalized Estimating Equations for Randomized Clinical Trials,
M. Mansourian, et. al **325**

A Modified Process for Modeling Wireless Communication Networks,
N. Modarresi, and S. Rezakhah **336**

Forward Moving Average Representation in Multivariate MA(1) Processes,

M. Mohammadpour, and A. R. Soltani **347**

Efficient Adaptive Sampling Designs for Assessing Marine Life Stacks,
M. Mohammad Salehi **359**

Notes Due to Lorenz Curve and Lorenz Ordering in View of Weighted Distributions,
G. R. Mohtashami Borzadaran , Z. Behdani **374**

Methods for Estimating Special Variations in Rotating Panel Surveys,
M. Namazi-Rad **387**

Runge–Kutta Methods For Numerical Solution Of Stochastic Differential Equations,
M. Namjoo, A. R. Soheili **398**

On using wavelet thresholding for independent component analysis,
V. Nassiri, M. Aminghafari **412**

Periodically correlated time series and their spectra,
A. Nematollahi **425**

Variational approximations for Logistic Mixed Models,
J.T. Ormerod, M.P. Wand **450**

On Nonparametric Density Estimation in Mixture Models,
R. Pakyari **468**

Licensing A New Treatment; A Frequentist Condition on a Bayesian Approach,
H. Pezeshk **473**

Comparing Quantile Regression and Linear Regression to Analysis the Risk Factors of Reflux Duration,

M. A. Pourhoseingholi, et. al **483**

A note on the mean past and the mean residual life of a $(n-k+1)-out-of-n$ system under multi monitoring,
M. H. Poursaeed **496**

A Nonparametric Method for Assessment of Interactions in a Multi-factor Design with Censored Data,
M. H. Rahbar, et al **510**

Asymptotic Behavior of Weighted Sums of Dependent Random Variables,
V. Ranjbar. Y et. al **524**

Influence diagnostic in partial least squares in measurement error models,
A. Rasekh, F. Mazarei **537**

Latend transition models for analyzing longitudinal ordinal response data with random dropout,
Z. Rezaei Ghahroodi, M. Ganjali **549**

Archimedean Copula and Multivariate Hazard rate Ordering,
M. Rezapour, M. H. Alamatsaz **563**

Point and interval estimators of fuzzy numbers,
A. Saeidifar **574**

A Characterization and Generalization for Two Sided Power Distributions and Adjusted Method of Moments,
A. R. Soltani H. Homei **585**

Minimum-Kolmogorov-distance estimator of the tail index of heavy- tailed distributions using V-statistics,
L. Tafakori, A. R. Nematollahi **606**

On the Inactivity Time of Lower k-Records and Associated

Stochastic Orders,
M. Tavangar, and Majid Asadi **614**

Characterization of Pareto distribution through coincidences of semiparamtric families,
Gh. Yari, M. Rajaei **626**

Assessment of Digit Preference in age data of national Iranian Census, 2005,
A. Yazdanparast, and et. al **634**

Almost Sure Results for a linear function of the top k order statistics,
A. Yousefi, R. Vasudeva **642**

Central Limit Theorem For Linearly Negative Dependent Fuzzy Random Variables,
R .Zaman, B. Sadeghpour Gildeh **655**

The Effect of Covariates on the Accuracy of Diagnostic or Screening Tests Using ROC Method,
N. Zare, et. al **664**

A study on the residual Rényi entropy of order statistics,
S. Zarezadeh **679**

Some Results on Dynamic Rényi Entropy

M. Abbasnejad, N. R. Arghami

Department of Statistics, Ferdowsi University of Mashhad

Abstract: This paper examines the role of Rényi entropy in ordering distributions and establishes a relation between Rényi entropy ordering of distributions and a strong ordering known as dispersion ordering. It is also shown that Rényi entropy of continuous transformations behave similar to variance and Shannon entropy. Further, we investigate Rényi entropy ordering of location-scale and shape-scale families. The results are conveniently tabulated in terms of distribution parameters.

If a system is known to have survived to age t, then Rényi entropy is no longer useful for measuring the uncertainty about the remaining lifetime of a system. Asadi *et al.* (2005) defined dynamic Rényi entropy which measures the uncertainty about the remaining lifetime of a system of age t. We present a bound for dynamic Rényi entropy of a system in terms of mean residual life function. Then some conditions are stated under which Rényi entropy of a residual lifetime distribution is monotone. We give sufficient conditions for a function $\phi(X)$ of a random variable X to have more (less) dynamic Rényi entropy than X itself.

Keywords: Dynamic Rényi Entropy; Life Distributions; Mean Residual Life Function; Ordering Relations.

1 Introduction

Let X be a continuous random variable with probability density function $f(x)$. Rényi entropy of order α is defined as (see Rényi, 1961)

$$H_\alpha(X) = -\frac{1}{\alpha - 1} \log \int_{-\infty}^{+\infty} f^\alpha(x) dx \qquad \alpha > 0 \quad (\alpha \neq 1)$$

where

$$H(X) = \lim_{\alpha \to 1} H_\alpha(X) = -\int_{-\infty}^{+\infty} f(x) \log f(x) dx$$

is Shannon entropy of X.

$H_\alpha(X)$ is known as the spectrum of Rényi information, when it is considered

as a function of α. For various properties and applications of $H_\alpha(X)$, one could refer to Rényi (1961), Morales *et al.* (1977) and Song (2001).

If we think of X as the life time of a new system, then $H(X)$ or $H_\alpha(X)$ can be useful for measuring the associated uncertainty. But in many cases, one has information about the current age of the system. Obviously $H(X)$ and $H_\alpha(X)$ are not suitable in such situations and they should be modified to take the current age into account. Accordingly, Ebrahimi (1996) introduced a measure of uncertainty known as residual entropy, for the residual life distribution. The residual entropy of X is defined as

$$H(X,t) = -\int_t^{+\infty} \frac{f(x)}{\bar{F}(t)} \log \frac{f(x)}{\bar{F}(t)} dx.$$

Where $\bar{F}(x)$ is the survival function of X. Indeed $H(X,t)$ is Shannon entropy of the Variable $Y = X - t|X > t$.
For applications of $H(X,t)$ one may refer to Ebrahimi and Kirmani (1996a-c), Sankaran and Gupta (1999), Asadi *et al.* (2000), Di Crescenzo and Longobardi (2002, 2004) and Belezunce *et al.* (2004).
Similarly, Asadi *et al.* (2005) and Abraham and Sankaran (2005) presented Rényi entropy of residual lifetime (dynamic Rényi entropy) and studied some properties of it.

The rest of the paper is organized as follows:
In Section 2, we examine the relation between dispersion ordering and Rényi entropy and variance orderings of random variables. Conditions for transformations of continuous random variables that preserve the equivalence of variance and Rényi entropy ordering are also presented. In Section 3, we order some well-known families based on their Rényi entropies. The results are tabulated in terms of distribution parameters. Finally in Section 4, more properties of dynamic Rényi entropy of a random variable X are discussed.

2 Rényi entropy ordering of random variables

Let X and Y be two random variables with distribution functions $F(x)$ and $G(y)$, probability density functions $f(x)$ and $g(y)$ and survival functions

$\bar{F}(x)$ and $\bar{G}(y)$, respectively.

Definition 1 *The random variable X is said to be less than or equal to Y in dispersion ordering, denoted by $X \overset{dis}{\leq} Y$, if and only if*

$$F^{-1}(v) - F^{-1}(u) \leq G^{-1}(v) - G^{-1}(u) \quad 0 < u \leq v < 1,$$

or equivalently,

$$g(G^{-1}(u)) \leq f(F^{-1}(u)) \quad \forall \ \ u \in (0,1).$$

Definition 2 *The random variable X is said to be less than or equal to Y in Rényi entropy ordering, denoted by $X \overset{Re}{\leq} Y$, if $H_\alpha(X) \leq H_\alpha(Y)$ for all $\alpha > 0$.*

Definition 3 *The random variable X is said to be less than or equal to Y in variance ordering, denoted by $X \overset{V}{\leq} Y$, if $V(X) \leq V(Y)$.*

Theorem 1 *If X and Y are two random variables such that $X \overset{dis}{\leq} Y$, then $X \overset{Re}{\leq} Y$.*

Proof: First note that the probability integral transformation provides the following useful representation of Rényi entropy of the random variable X

$$H_\alpha(X) = -\frac{1}{\alpha - 1} \int_0^1 f^{\alpha-1}(F^{-1}(u))du.$$

Since $X \overset{dis}{\leq} Y$ we have

$$\int_0^1 [f^{\alpha-1}(F^{-1}(u)) - g^{\alpha-1}(G^{-1}(u))]du \geq 0, \quad \forall \ \ \alpha > 1,$$

$$\int_0^1 [f^{\alpha-1}(F^{-1}(u)) - g^{\alpha-1}(G^{-1}(u))]du \leq 0, \quad \forall \ \ 0 < \alpha < 1.$$

So by noting that log is an increasing function, the result follows.

The next theorem states some conditions under which variance and Rényi entropy order similarly when continuous random variables are transformed.

Theorem 2 *Let X be a continuous random variable with probability density function $f(x)$. Let $Y = g(X)$, where $g(x)$ is a function with a continuous derivative $g'(x)$ in the support of $f(x)$ and $EY^2 < +\infty$. If $|g'(x)| \geq 1$ for all x in the support of $f(x)$, then $X \overset{V}{\leq} Y$ and $X \overset{Re}{\leq} Y$.*

Proof: $X \overset{V}{\leq} Y$ has been proved by Ebrahimi *et al.* (1999). Let $h(y)$ denote probability density function of Y. Then

$$\begin{aligned} H_\alpha(Y) &= -\frac{1}{\alpha-1}\log\int_{-\infty}^{+\infty} h^\alpha(y)dy \\ &= -\frac{1}{\alpha-1}\log\int_{-\infty}^{+\infty} f^\alpha(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)|^\alpha dy \\ &= -\frac{1}{\alpha-1}\log\int_{-\infty}^{+\infty} f^\alpha(x)|g'(x)|^{1-\alpha}dx \\ &\geq -\frac{1}{\alpha-1}\log\int_{-\infty}^{+\infty} f^\alpha(x)dx = H_\alpha(X), \end{aligned}$$

and the proof is completed.

Example 1 *Let X be a non-negative random variable with an absolutely distribution function $F(x)$. If $Y_i = \exp\theta_i X, \quad i = 1, 2$, where $0 < \theta_1 < \theta_2$, then $Y_1 \overset{Re}{\leq} Y_2$.*

It can be easily shown that

$$H_\alpha(Y_i) = \log\theta_i - \frac{1}{\alpha-1}\log\int_0^{+\infty} f^\alpha(x)e^{-(\alpha-1)\theta_i x}dx.$$

So

$$H_\alpha(Y_2) - H_\alpha(Y_1) = \log\frac{\theta_2}{\theta_1} - \frac{1}{\alpha-1}\log\frac{\int_0^{+\infty} f^\alpha(x)e^{-(\alpha-1)\theta_2 x}dx}{\int_0^{+\infty} f^\alpha(x)e^{-(\alpha-1)\theta_1 x}dx} > 0, \quad \forall\ \alpha > 0.$$

Example 2 *Let X be a non-negative random variable with an absolutely distribution function $F(x)$. If $Y_i = X_i^{\theta_i}, \quad i = 1, 2$, where $1 < \theta_1 < \theta_2$, then $Y_1 \overset{Re}{\leq} Y_2$.*

We have

$$H_\alpha(Y_2)-H_\alpha(Y_1)=\log\frac{\theta_2}{\theta_1}-\frac{1}{\alpha-1}\log\frac{\int_0^{+\infty}f^\alpha(x)x^{-(\alpha-1)(1-\theta_2)}dx}{\int_0^{+\infty}f^\alpha(x)x^{-(\alpha-1)(1-\theta_1)}dx}>0,\quad\forall\ \alpha>0.$$

3 Rényi entropy and variance of some well-known distributions

3.1 Location-Scale family

Definition 4 *A distribution is said to be in a location-scale family with location parameter θ and scale parameter β if its density is in the form of*

$$f(x;\theta,\beta)=\frac{1}{\beta}f_0(\frac{x-\theta}{\beta}).$$

We can write a random variable X with a distribution in a location-scale family as $X=\beta Z+\theta$, where Z has a distribution free from θ and β. Thus, the variance and Rényi entropy of all location-scale distributions are independent of the location parameter θ and can be written as

$$V(X)=\beta^2V(Z),\qquad\qquad H_\alpha(X)=\log\beta+H_\alpha(Z),$$

where $V(Z)$ and $H_\alpha(Z)$ are constants independent of θ and β. Thus, within each location-scale family variance and Rényi entropy are increasing functions of β. Table 1 shows several well-known location-scale families of distributions with their variance and Rényi entropy.

3.2 shape-scale family

Definition 5 *A distribution is said to be in a shape-scale family with shape parameter θ and scale parameter β if its density is in the form of*

$$f(x;\theta,\beta)=\frac{1}{\beta}f_0(\frac{x}{\beta};\theta).$$

Table 1. Rényi entropy and variance ordering of location-scale distributions.

family and density	variance and Rényi entropy	orderings
Normal:	$V(X)=\beta^2$	$V\nearrow\beta$
$f(x)=\frac{1}{\sqrt{2\pi}\beta}e^{-\frac{(x-\theta)^2}{2\beta^2}}$	$H_\alpha(X)=\log\beta+\frac{1}{2}\log 2\pi$	$H_\alpha\nearrow\beta$
$-\infty<x,\theta<+\infty,\quad \beta>0$	$+\frac{\log\alpha}{2(\alpha-1)}$	
Extreme value:	$V(X)=\frac{\pi^2\beta^2}{6}$	$V\nearrow\beta$
$f(x)=\frac{1}{\beta}e^{-\frac{(x-\theta)}{\beta}-e^{-\frac{(x-\theta)}{\beta}}}$	$H_\alpha(X)=\log\beta-\frac{1}{\alpha-1}\log\Gamma(\alpha)$	$H_\alpha\nearrow\beta$
$-\infty<x,\theta<+\infty,\quad \beta>0$	$+\frac{\alpha}{\alpha-1}\log\alpha$	
Laplace:	$V(X)=2\beta^2$	$V\nearrow\beta$
$f(x)=\frac{1}{2\beta}e^{-\frac{\lvert x-\theta\rvert}{\beta}}$	$H_\alpha(X)=\log\beta+\frac{1}{\alpha-1}\log\alpha+\log 2$	$H_\alpha\nearrow\beta$
$-\infty<x,\theta<+\infty,\ \beta>0$		
Logistic:	$V(X)=\frac{\pi^2\beta^2}{3}$	$V\nearrow\beta$
$f(x)=\frac{1}{\beta}\frac{e^{-\frac{(x-\theta)}{\beta}}}{(1+e^{-\frac{(x-\theta)}{\beta}})^2}$	$H_\alpha(X)=\log\beta-\frac{1}{\alpha-1}\log B(\alpha,\alpha)$	$H_\alpha\nearrow\beta$
$-\infty<x,\theta<+\infty,\ \beta>0$		
Uniform:	$V(X)=\frac{\beta^2}{12}$	$V\nearrow\beta$
$f(x)=\frac{1}{\beta}$	$H_\alpha(X)=\log\beta$	$H_\alpha\nearrow\beta$
$\theta-\frac{\beta}{2}<x<\theta+\frac{\beta}{2},\quad \beta>0$		
Cauchy:	variance does not exist	$H_\alpha\nearrow\beta$
$f(x)=\frac{1}{\pi\beta\{1+(\frac{x-\theta}{\beta})^2\}}$	$H_\alpha(X)=\log\beta+\frac{\alpha}{\alpha-1}\log\pi$	
$-\infty<x,\theta<+\infty,\ \beta>0$	$-\frac{1}{\alpha-1}\log B(\frac{1}{2},\alpha-\frac{1}{2}),\quad \forall\alpha>\frac{1}{2}$	
Levy:	variance does not exist	$H_\alpha\nearrow\beta$
$f(x)=\sqrt{\frac{\beta}{2\pi}}\frac{1}{(x-\theta)^{\frac{3}{2}}}e^{-\frac{\beta}{2(x-\theta)}}$	$H_\alpha(X)=\log\beta-\log 2$	
$x>\theta,\ \beta>0$	$-\frac{1}{\alpha-1}\{\log\frac{\Gamma(\frac{3}{2}\alpha-1)}{\Gamma(\frac{1}{2})}-(\frac{3}{2}\alpha-1)\log\alpha\}$	
	$\forall\alpha>\frac{3}{2}$	

Ebrahimi *et al.* (1999).
The variance and Rényi entropy of these distributions are increasing in β, but for most of these families, variance and Rényi entropy are complicated functions of θ. The results of Section 2 can be useful for establishing the orderings. Table 2 shows some shape-scale families with their Rényi entropy and variance.
Shaked (1982) showed that the scaled ($\beta = 1$) gamma family has dispersion ordering with respect to θ, that is, if $X_i \sim \Gamma(\theta_i, 1), \quad i = 1, 2$, where $\theta_1 < \theta_2$, then $X_1 \overset{dis}{\leq} X_2$ and so by Theorem 2.1 it follows that $X_1 \overset{V,Re}{\leq} X_2$.
If $X \sim \Gamma(\theta, \beta)$, then $Y = \frac{1}{X}$ has an Inverse gamma distribution. Also, $h(x) = \frac{1}{x}$ is convex and decreasing. Thus, from Theorem 2.B.5 of Shaked and Shantikumar (1994) if $X_1 \overset{dis}{\leq} X_2$, then $Y_1 \overset{dis}{\geq} Y_2$. So we have $Y_1 \overset{V,Re}{\geq} Y_2$.
From Theorem 2.6 of Shaked (1982), it can be shown that the generalized normal distribution is dispersion ordered in θ. Thus, for generalized normal family, $(V, H_\alpha) \nearrow \theta$.
The orderings for the inverse generalized normal distribution are obtained similar to inverse gamma. The orderings for the log-normal, Pareto and triangular distribution are obtained easily by taking derivatives. For Weibull and log-logestic distributions, we get orderings using the relation between dispersion ordering and variance and Rényi entropy orderings.

Remark 1 *Rényi entropy expressions of most of these distributions are tabulated by Song (2001).*

Remark 2 *Shannon entropy ordering of most of these distributions have been given by Ebrahimi et al. (1999).*

4 Dynamic Rényi entropy

Let X be non-negative random variable representing the lifetime of a system and $f(x)$ and $F(x)$ denote the probability density function and distribution function of X, respectively. Let $\bar{F}(x) = 1 - F(x)$ be the survival function with $\bar{F}(0) = 1$.

Table 2. Rényi entropy and variance ordering of shape-scale distributions.

family and density	variance and Rényi entropy	orderings
Exponential: $f(x)=\frac{1}{\beta}e^{-\frac{x}{\beta}}$ $x>0,\ \beta>0$	$V(X)=\beta^2$ $H_\alpha(X)=\log\beta+\frac{1}{\alpha-1}\log\alpha$	$(V,H_\alpha)\nearrow\beta$
Gamma: $f(x)=\frac{1}{\Gamma(\theta)\beta^\theta}x^{\theta-1}e^{-\frac{x}{\beta}}$ $x>0,\ \theta,\beta>0$	$V(X)=\theta\beta^2$ $H_\alpha(X)=\log\beta+\frac{1}{\alpha-1}[\alpha\log\Gamma(\theta)$ $-\log\Gamma(\alpha(\theta-1)+1)$ $+(\alpha(\theta-1)+1)\log\alpha]$	$(V,H_\alpha)\nearrow\beta$ $(V,H_\alpha)\nearrow\theta$
Log-Normal: $f(x)=\frac{1}{\sqrt{2\pi}\theta x}e^{-\frac{1}{2\theta^2}[\log x-\log\beta]^2}$ $x>0,\ \theta,\beta>0$	$V(X)=\beta^2[e^{2\theta^2}-e^{-\theta^2}]$ $H_\alpha(X)=\log\beta+\log\theta$ $+\frac{\log\alpha}{2(\alpha-1)}-\frac{(\alpha+1)^2}{2\alpha(\alpha-1)}\theta^2+\frac{1}{2}\log2\pi$	$(V,H_\alpha)\nearrow\beta$ $V\nearrow\theta$ $H_\alpha\searrow\theta\ \ (\alpha>1)$ $H_\alpha\nearrow\theta\ \ (\alpha<1)$
Pareto: $f(x)=\frac{\theta\beta^\theta}{x^{\theta+1}}$ $x>\beta,\ \theta,\beta>0$	$V(X)=\frac{\beta^2}{\theta(\theta-1)^2(\theta-2)}\ \ (\theta>2)$ $H_\alpha(X)=\log\beta-\frac{\alpha}{\alpha-1}\log\theta$ $+\frac{1}{\alpha-1}\log\{(1+\theta)\alpha-1\}$	$(V,H_\alpha)\nearrow\beta$ $(V,H_\alpha)\searrow\theta$
Weibull: $f(x)=\theta(\frac{x}{\beta})^{\theta-1}e^{-(\frac{x}{\beta})^\theta}$ $x>0,\ \theta,\beta>0$	$V(X)=\beta^2[\Gamma(1+\frac{2}{\theta})-\Gamma^2(1+\frac{1}{\theta})]$ $H_\alpha(X)=\log\beta-\frac{1}{\alpha-1}\log\Gamma(\alpha-\frac{\alpha-1}{\theta})$ $-\frac{\alpha+\theta(\alpha-1)}{\alpha-1}\log\alpha-\log\theta$	$(V,H_\alpha)\nearrow\beta$ $(V,H_\alpha)\searrow\theta$
Triangular: $f(x)=\frac{2}{\theta\beta}\frac{(x)}{\beta}\quad 0\le x\le\theta\beta$ $\frac{2}{(1-\theta)\beta}(1-\frac{(x)}{\beta})\quad \theta\beta\le x\le\beta$	$V(X)=\frac{\beta^2(\theta^2-\theta+1)}{18}$ $H_\alpha(X)=\log\beta-\frac{\alpha}{\alpha-1}\log2$ $+\frac{1}{\alpha-1}\log(\alpha+1)$	$(V,H_\alpha)\nearrow\beta$ $V\searrow\theta<\frac{1}{2}$ $V\nearrow\theta>\frac{1}{2}$
Inverse Gamma: $f(x)=\frac{1}{\Gamma(\theta)\beta}(\frac{x}{\beta})^{-\theta-1}e^{-(\frac{x}{\beta})^{-1}}$ $x>0,\ \theta,\beta>0$	$V(X)=\frac{\beta^2}{(\theta-1)^2(\theta-2)}\ \ (\theta>2)$ $H_\alpha(X)=\log\beta+\frac{1}{\alpha-1}[\alpha\log\Gamma(\theta)$ $-\log\Gamma(\alpha(\theta+1)-1)$ $+(\alpha(\theta+1)-1)\log\alpha]$	$(V,H_\alpha)\nearrow\beta$ $(V,H_\alpha)\searrow\theta$
Generalized Normal: $f(x)=\frac{1}{2\Gamma(\theta/2)\beta}(\frac{x}{\beta})^{\theta-1}e^{-(\frac{x}{\beta})^2}$ $x>0,\ \theta,\beta>0$	$V(X)=\beta^2\frac{\theta-2\Gamma^2(\frac{\theta}{2}+\frac{1}{2})}{2\Gamma^2(\frac{\theta}{2})}$ $H_\alpha(X)=\log\beta+\frac{1}{\alpha-1}[\alpha\log\Gamma(\frac{\theta}{2})$ $-\log\Gamma(\frac{\alpha(\theta-1)+1}{2})+\frac{(\alpha(\theta-1)+1)}{2}\log\alpha$ $+(\alpha+1)\log2]$	$(V,H_\alpha)\nearrow\beta$ $(V,H_\alpha)\nearrow\theta$
Inverse Generalized Normal: $f(x)=\frac{1}{2\Gamma(\theta/2)\beta}(\frac{x}{\beta})^{-\theta-1}e^{-(\frac{x}{\beta})^{-2}}$ $x>0,\ \theta,\beta>0$	$V(X)=\beta^2[\frac{1}{\theta-2}-\frac{\theta\Gamma^2(\theta-1)}{2\Gamma^2(\frac{\theta}{2})}]$ $H_\alpha(X)=\log\beta+\log2+\frac{1}{\alpha-1}[\alpha\log\Gamma(\frac{\theta}{2})$ $-\log\Gamma(\frac{\alpha(\theta+1)-1}{2})$ $+\frac{(\alpha(\theta+1)-1)}{2}\log\alpha]\ \ \alpha>\frac{1}{\theta+1}$	$(V,H_\alpha)\nearrow\beta$ $(V,H_\alpha)\searrow\theta$
Log-logistic: $f(x)=\frac{\frac{\theta}{\beta}(\frac{x}{\beta})^{\theta-1}}{[1-(\frac{x}{\beta})^\theta]^2}$	$V(X)=\beta^2[\Gamma(1+\frac{2}{\theta})\Gamma(1-\frac{2}{\theta})$ $-\Gamma^2(1+\frac{1}{\theta})\Gamma^2(1-\frac{1}{\theta})]\ \ (\theta>2)$ $H_\alpha(X)=\log\beta-\log\theta-\frac{1}{\alpha-1}\log B$ $[(\alpha-1)(1+\frac{1}{\theta})+1,\alpha(1-\frac{1}{\theta})+\frac{1}{\theta}]$	$(V,H_\alpha)\nearrow\beta$ $(V,H_\alpha)\searrow\theta$

Rényi entropy of the residual life density is defined by

$$\begin{aligned} H_\alpha(f,t) &= -\frac{1}{\alpha-1}\log\int_t^{+\infty}\left[\frac{f(x)}{\bar{F}(t)}\right]^\alpha dx \\ &= \frac{\alpha}{\alpha-1}\log\bar{F}(t)-\frac{1}{\alpha-1}\log\int_t^{+\infty} f^\alpha(x)dx. \end{aligned} \tag{1}$$

In the following theorem we give an upper bound for $H_\alpha(f,t)$ in terms of mean residual life function (MRL).

Theorem 3 *If X is a life time of a system with probability density function $f(x)$, survival function $\bar{F}(x)$ and mean residual life function $\delta_F(t)$, then*

$$H_\alpha(f,t)\le\frac{\log\alpha}{\alpha-1}+\log\delta_F(t),\qquad\forall\ t\ge 0. \tag{2}$$

Where $\delta_F(t)=E(X-t|X>t)=\int_t^{+\infty}x\frac{f(x)}{\bar{F}(t)}dx-t=\frac{\int_t^{+\infty}\bar{F}(x)dx}{\bar{F}(t)}$.

Proof: Let $Y_t\overset{d}{=}Y|Y>t$ and $g_t(y)$ be its probability density function. Then

$$g_t(y)=\frac{d}{dy}P(Y_t\le y)=\frac{d}{dy}\big[P(Y\le y|Y>t)\big]=\frac{d}{dy}\frac{F(y)}{\bar{F}(t)}=\frac{f(y)}{\bar{F}(t)}$$

If we define $Z_t=Y_t-t$, then its probability density function is $h_t(\eta)=g_t(\eta+t)$ and $E(Z_t)=\delta_F(t)$. Thus,

$$\begin{aligned} H_\alpha(Z_t) &= -\frac{1}{\alpha-1}\log\int_0^{+\infty}h_t^\alpha(\eta)d\eta=-\frac{1}{\alpha-1}\log\int_0^{+\infty}g_t^\alpha(\eta+t)d\eta \\ &= -\frac{1}{\alpha-1}\log\int_t^{+\infty}g_t^\alpha(\eta)d\eta=H_\alpha(f,t). \end{aligned}$$

Under assumption $\delta_F(t)<\infty$, if the support of a random variable is $[0,\infty)$, exponential distribution with mean $\delta_F(t)$ has the maximum entropy. So,

$$H_\alpha(f,t)=-\frac{1}{\alpha-1}\log\int_0^{+\infty}h_t^\alpha(\eta)d\eta\le\frac{\log\alpha}{\alpha-1}+\log\delta_F(t),$$

and the proof is completed.

Definition 6 *A distribution function $F(x)$ is said to have decreasing (increasing) dynamic Rényi entropy (DDRE (IDRE)) if $H_\alpha(f,t)$ is decreasing (increasing) in $t \geq 0$.*

Definition 7 *A distribution function $F(x)$ is said to be decreasing (increasing) mean residual life (DMRL (IMRL)) if its mean residual life function is decreasing (increasing) in $t \geq 0$.*

Theorem 4 *If F is DDRE, then*

$$H_\alpha(f,t) \leq \frac{\log \alpha}{\alpha - 1} + \log \mu,$$

where $\mu = \delta_F(0)$.

Proof: $H_\alpha(f,t)$ is a decreasing function of t. So the result follows easily by inequality (2).

In the following theorem we identify conditions under which Rényi entropy of a residual lifetime distribution is monotone.

Theorem 5 *(a) If F is DMRL, then it is DDRE.*
(b) If F is IDRE, then it is IMRL.

Proof: It is easy to see that

$$\begin{aligned}
\frac{\partial}{\partial t} H_\alpha(f,t) &= -\frac{\alpha}{\alpha-1}\frac{f(t)}{\bar{F}(t)} + \frac{1}{\alpha-1}\frac{f^\alpha(t)}{\int_t^{+\infty} f^\alpha(x)dx} \\
&= -\frac{\alpha}{\alpha-1} r_F(t) + \frac{1}{\alpha-1} r_F^\alpha(t) e^{(\alpha-1)H_\alpha(f,t)} \\
&\leq -\frac{\alpha}{\alpha-1} r_F(t) + \frac{1}{\alpha-1} r_F^\alpha(t) e^{(\alpha-1)[\frac{\log\alpha}{\alpha-1} + \log \delta_F(t)]} \\
&= -\frac{\alpha}{\alpha-1} r_F(t) + \frac{1}{\alpha-1} r_F^\alpha(t)[\alpha \delta_F^{\alpha-1}(t)] \\
&= -\frac{\alpha}{\alpha-1} r_F(t)\big\{ -1 + [r_F(t)\delta_F(t)]^{\alpha-1}\big\} \\
&= -\frac{\alpha}{\alpha-1} r_F(t)\big\{ -1 + [1 + \delta_F'(t)]^{\alpha-1}\big\}.
\end{aligned}$$

Where the above inequality comes from inequality (2) and the last equality holds because $r_F(t)\delta_F(t) = 1 + \delta'_F(t)$. Using above inequality the result follows.

In many cases of practical interest one would like to know whether the DDRE (IDRE) property of X is inherited by a transformation of X. The next theorem provides a partial answer.

Theorem 6 *(a) If X is IDRE, and if ϕ is non-negative, increasing and convex, then $\phi(X)$ is also IDRE.*
(b) If X is DDRE, and if ϕ is non-negative, increasing and concave, then $\phi(X)$ is also DDRE.

Proof: (a) The probability density function of $Y = \phi(X)$ is $g(y) = \frac{f(\phi^{-1}(y))}{\phi'(\phi^{-1}(y))}$. Thus,

$$H_\alpha(g,t) = -\frac{1}{\alpha-1}\log\left\{\frac{1}{\bar{F}^\alpha(\phi^{-1}(t))}\int_t^{+\infty}\frac{f^\alpha(\phi^{-1}(y))}{\phi'^\alpha(\phi^{-1}(y))}dy\right\}$$

by taking $x = \phi^{-1}(t)$ we have

$$H_\alpha(g,t) = -\frac{1}{\alpha-1}\log\left\{\frac{1}{\bar{F}^\alpha(\phi^{-1}(t))}\int_{\phi^{-1}(t)}^{+\infty} f^\alpha(x)\phi'^{1-\alpha}(x)dx\right\} \qquad (3)$$

and

$$\begin{aligned}\frac{\partial}{\partial t}H_\alpha(g,t) &= -\frac{\alpha}{\alpha-1}\frac{-\frac{1}{\phi'(t)}f(\phi^{-1}(t))}{\bar{F}(\phi^{-1}(t))} \\ &\quad -\frac{1}{\alpha-1}\frac{-\frac{1}{\phi'(t)}f^\alpha(\phi^{-1}(t))\phi'^{1-\alpha}(\phi^{-1}(t))}{\int_{\phi^{-1}(t)}^{+\infty} f^\alpha(x)\phi'^{1-\alpha}(x)dx} \\ &= -\frac{\alpha}{\alpha-1}\frac{1}{\phi'(t)}r_F(\phi^{-1}(t)) \\ &\quad + \frac{1}{\alpha-1}.\frac{1}{\phi'(t)}r_F^\alpha(\phi^{-1}(t)).\frac{\phi'^{1-\alpha}(\phi^{-1}(t))}{\int_{\phi^{-1}(t)}^{+\infty}\frac{f^\alpha(x)}{\bar{F}^\alpha(\phi^{-1}(t))}\phi'^{1-\alpha}(x)dx},\end{aligned}$$

Let $\alpha > 1$. $\phi'(x)$ is an increasing function because $\phi(x)$ is a convex function. So $\phi'^{1-\alpha}(x)$ is a decreasing function, that is,

$$\phi'^{1-\alpha}(x) \leq \phi'^{1-\alpha}(\phi^{-1}(t)), \qquad \forall \;\; x > \phi^{-1}(t).$$

Hence,

$$\begin{aligned}
\frac{\partial}{\partial t}H_\alpha(g,t) &\geq -\frac{\alpha}{\alpha-1}\frac{1}{\phi'(t)}r_F(\phi^{-1}(t)) \\
&+ \frac{1}{\alpha-1}.\frac{1}{\phi'(t)}r_F^\alpha(\phi^{-1}(t)).\frac{1}{\int_{\phi^{-1}(t)}^{+\infty}\left[\frac{f(x)}{\bar{F}(\phi^{-1}(t))}\right]^\alpha dx} \\
&= \frac{1}{\phi'(t)}\left[-\frac{\alpha}{\alpha-1}r_F(\phi^{-1}(t)) + \frac{1}{\alpha-1}r_F^\alpha(\phi^{-1}(t))e^{(\alpha-1)H_\alpha(f,\phi^{-1}(t))}\right] \\
&= \frac{1}{\phi'(t)}\left[H'_\alpha(f,\phi^{-1}(t))\right] \geq 0.
\end{aligned}$$

A similar result follows for $0 < \alpha < 1$.
(b) The proof is similar to that of (a).

Example 3 *Let X have the exponential distribution with failure rate θ. Then $Y = X^{\frac{1}{\beta}}$ $(\beta > 0)$, has the Weibull distribution with survival function $\bar{G}(t) = \exp(-\theta t^\beta)$ $t > 0$. The non-negative increasing function $\phi(x) = x^{\frac{1}{\beta}}$ $x > 0, \beta > 0$ is convex (concave) if $0 < \beta < 1$ $(\beta > 1)$. Hence, Weibull distribution is IDRE (DDRE) if $0 < \beta < 1$ $(\beta > 1)$.*

In the next theorem we give sufficient conditions for a function $\phi(X)$ of a random variable X to have more (less) dynamic Rényi entropy than X itself.

Theorem 7 *(a) If ϕ is non-negative and increasing on $[0, +\infty)$ with $\phi'(x) \geq 1$ for all $x \geq 0$ and if F is DDRE, then*

$$H_\alpha(\phi(X), t) \geq H_\alpha(X, t).$$

(b) If ϕ is increasing on $[0, +\infty)$ with $\phi(0) = 0$ and $\phi'(x) \leq 1$ for all $x \geq 0$ and if F is IDRE, then

$$H_\alpha(\phi(X), t) \leq H_\alpha(X, t).$$

Proof: Letting $g(x)$ denote the probability density function of $\phi(X)$ and using (3) we have

$$H_\alpha(g,t) = \frac{\alpha}{\alpha-1}\log \bar{F}^\alpha(\phi^{-1}(t)) - \frac{1}{\alpha-1}\log \int_{\phi^{-1}(t)}^{+\infty} \frac{f^\alpha(x)}{\phi'^{\alpha-1}(x)}dx.$$

For $\alpha > 1$, by noting that $\phi'(x) \geq 1$ it follows that

$$H_\alpha(g,t) \geq \frac{\alpha}{\alpha-1}\log \bar{F}^\alpha(\phi^{-1}(t)) - \frac{1}{\alpha-1}\log \int_{\phi^{-1}(t)}^{+\infty} f^\alpha(x)dx = H_\alpha(f,\phi^{-1}(t)),$$

Also, note that if $\phi'(x) \geq 1, \quad x \geq 0$, then $\phi(x) - \phi(0) \geq x$. So that $x \geq \phi^{-1}(x)$ for non-negative increasing ϕ. Thus,

$$H_\alpha(g,t) \geq H_\alpha(f,\phi^{-1}(t)) \geq H_\alpha(f,t).$$

We have a similar result for $0 < \alpha < 1$.
(b) The proof is similar to that of (a).

References

[1] Abraham, B. and Sankaran, P.G. (2005). Renyi's entropy for residual lifetime distributions. *Stat. Papers*, **46**, 17-30.

[2] Asadi, M., Ebrahimi, N., Hamedani, G.G., Soofi, E.S. (2000). Maximum dynamic entropy models. *J. Appl. Prob.*, **41**, 379-390.

[3] Asadi, M., Ebrahimi, N., Soofi, E.S. (2005). Dynamic generalized information measures. *Stat. Prob. Lett.*, **71**, 85-98.

[4] Belzunce, F., Navarro, J., Ruiz, J.M. (2004). Some results on residual entropy functions. *Metrika*, **59**, 147-161.

[5] Di Crescenzo, A. and Logobardi, M. (2002). Entropy-based measure of uncertainty in past lifetime distributions. *J. Appl. Prob.*, **39**, 434-440.

[6] Di Crescenzo, A. and Logobardi, M. (2004). A measure of discrimination between past lifetime distributions. *Stat. Prob. Lett.*, **67**, 173-182.

[7] Ebrahimi, N. (1996). How to measure uncertainty in the residual lifetime distributions. *Sankhya*, **58**, 48-57.

[8] Ebrahimi, N. and Kirmani, S.N.U.A (1996a). A characterization of the proportional hazards models through a measure of discrimination between two residual lifetime distributions. *Biometrika*, **83**, 233-235.

[9] Ebrahimi, N. and Kirmani, S.N.U.A (1996b). A measure of discrimination between two residual lifetime distributions and its applications. *ann. Inst. Statist. Math.*, **48**, 257-265.

[10] Ebrahimi, N. and Kirmani, S.N.U.A (1996c). Some results on ordering of survival functions through uncertainty. *Stat. prob. Lett.*, **29**, 167-176.

[11] Ebrahimi, N., Maasoumi, E., Soofi, E.S. (1999). Ordering univariate distributions by entropy and variance. *Econometrics*, **90**, 317-336.

[12] Morales, D. Pardo, L., Vajda, I. (1977). Some new statistics for testing hypotheses in parametric models. *Mult. Anal.*, **62**, 137-168.

[13] Renyi, A. (1961), On measures of entropy and information. Proc. Fourth. Berkeley Symp. Math. Stat. Prob. 1960, Vol I, University of California Press, Berkeley, 547-561.

[14] Sankaran, P.G. and Gupta, R.P. (1999). Characterization of lifetime distributions using measure of uncertainty. *Calcutta Stat. Assoc. Bull.*, **49**, 154-166.

[15] Shaked, M. (1982), Dispertive ordering of distributions. *J. App. Prob.*, **19**, 310-320.

[16] Shaked, M. and Shanthikumar, J.G. (1994), *Stochastic Orders and Their Applications.* Academic Press, New York.

[17] Song, K. (2001). Renyi information, loglikelihood and an intrinsic distribution measure. *Stat. Plan. Inf.*, **93**, 51-69.

Multistep Forecasting Non-Stationary Time Series Using Wavelets and Kernel Smoothing

Mina Aminghafari[1,2,*], and Jean-Michel Poggi[3], [†]

[1] *Department of Statistics, Faculty of Mathematics and Computer Science, Amirkabir University of Technology*

[2] *Institute for Studies in Theoretical Physics and Mathematics (IPM)*

[3] *Laboratoire de Mathématique – U.M.R. C 8628, Université Paris-Sud*

Abstract: This work deals with forecasting time series using wavelets and kernel smoothing. A forecasting procedure can be defined by estimating the prediction equation by direct regression of the process on the non-decimated wavelet coefficients depending on its past values. In the same context, after the seminal work of Renaud et al. [7], we study a generalization of the prediction procedure associating kernel smoothing and wavelets. We then illustrate the proposed procedure on non-stationary simulated and real data.

Keywords: Forecasting, Non-stationary, Time series, Wavelets, Kernel.

1 Introduction

For fifteen years, wavelets have been used for various purposes in statistics including denoising, nonparametric function estimation, data compression as well as process synthesis for example (see [2] and [8]). An interesting example for time series is given by Dahlhaus *et al.* considering a time-varying autoregressive process.

An approach considered by Renaud *et al.* [7] for time series forecasting using wavelets, estimates directly the prediction equation by direct regression of the process on the Haar non-decimated wavelet coefficients depending on its past values. Starting from this seminal work, [1] propose various extensions in different directions such as using more regular wavelets or extrapolating the low frequency component of a possibly non-stationary signal. In [1], low frequency component is extrapolated by local polynomial fitting. In this paper we propose to examine two new topics: the first is to usee kernel smoothing for the low-frequency component extrapolation an extension to multi-step prediction. For high frequency components

[*] The first author was in part supported by a grant from IPM (No. 86620037)

[†] Jean-Michel.Poggi@math.u-psud.fr

given by the details, we use reconstructed versions of the detail coefficients instead of the original signal. Finally, let us mention that an alternative method mixing more closely wavelet decomposition and kernel method is proposed in [3]) in a functional framework and should be considered in a future work.
This paper is organized as follows. In Section 2 the use of wavelets for time series forecasting is motivated. In Section 3 we recall some facts about non-decimated wavelet transform and kernel smoothing. In Section 4, the proposed approach for stationary process is recalled and then it is adapted to non-stationary series. Finally, the proposed procedure is illustrated on non-stationary simulated and real data. Then the proposed prediction method is applied for non-stationary data.

2 Why wavelets for time series forecasting?

Let us recall some key arguments for using wavelet decompositions for forecasting time series.

Wavelet transform decomposes the observed signal in a smooth low frequency component and a sum of details corresponding to increasing resolution levels. It turns out that if the observed series is of the form: $Y(t) = f(t) + X(t)$, where X is a stochastic process and f a deterministic smooth function. The deterministic part of series (trend or smooth part) can be estimated, for example, by polynomial fitting of the approximation coefficients of a suitably chosen decomposition level and the detail coefficients are used for the prediction of the purely stochastic part. Hence assuming the stationarity of X and a sufficiently regular wavelet, wavelet transform automatically filters the non-stationary component, instead of trying to detrend or suppress quasi-periodic smooth components as in the classical non-stationary ARIMA approach.

Complex multiscale structure of the observed signal can often be simplified using wavelets leading to signals of simpler structure. Wavelet transform produces coefficients which are organized according to two parameters related to time and scale respectively. The coefficients at a given scale, of a long memory stochastic signal are of short memory, and then easier to model.

A given wavelet coefficient depends on the values of the signal only

within an interval centered around the corresponding position and of length proportional to the associated scale. It follows that if the considered series is not too far from stationarity, procedures using wavelets are naturally localized.

3 Preliminary material

For basics on wavelets the reader can refer to [4]. Let us recall some facts about the non-decimated wavelet transform and kernel smoothing.

3.1 Non-decimated wavelet transform

The so-called discrete wavelet transform (DWT) is often used for estimation purposes since it is associated to orthogonal or biorthogonal wavelet bases. But this transform is not translation invariant, *i.e.* the coefficients of a translated version of a signal are, in general, not equal to the translated coefficients of the same signal. So it turns out that, in the time series context, if we observe one more observation or delete the first value of the past data, the new discrete wavelet coefficients could be different from the previous original coefficients. This lack of translation invariance of the usual discrete wavelet transform (DWT) has both practical and theoretical consequences: the computations need to be performed again when a new observation is available and trying to fit a regression model between a future value of a series and some DWT coefficients depending the past values, is ill posed since the explanatory variables change with the instant of the last observed value.

A classical way to circumvent this drawback is to use the non-decimated wavelet transform (NDWT) (see [5]). The non-decimated detail and approximation coefficients of $(X_0, X_1, ..., X_{N-1})$ are defined for levels $j \geq 1$ by:

$$w_{j,t} = \sum_{k=0}^{L_j-1} \tilde{g}_{j,k} X_{(t-k) \bmod N} \quad \text{and } c_{j,t} = \sum_{k=0}^{L_j-1} \tilde{h}_{j,k} X_{(t-k) \bmod N} \tag{1}$$

where $\{\tilde{h}_{j,k}\}_{j,k}$ and $\{\tilde{g}_{j,k}\}_{j,k}$ are the filters corresponding to the NDWT wavelet and obtained by convolving upsampled versions of $\tilde{h}$, low-pass and $\tilde{g}$, high-pass filters associated with the orthogonal wavelet, normalized in

l^1 and supposed to be of length L (see [5] p. 169). Starting from the observations, we can write:

$$X = \mathcal{A}_J + \sum_{j=1}^{J} \mathcal{D}_j. \tag{2}$$

where $\mathcal{A}_J$ and D_j are suitably chosen reconstructed versions associated with the NDWT coefficients. In addition, for the Haar wavelet, the reconstruction step is trivial and then, for any t the following relation holds: $X_t = c_{J,t} + \sum_{j=1}^{J} w_{j,t}$.

Let us remark that the NDWT decomposition preserves the total energy of signal i.e. the l^2 norm $\|X\|^2 = \sum_{j=1}^{J} \|W_j\|^2 + \|C_J\|^2$ where X, W_j, C_J are the wavelet coefficients of level j and the approximation coefficients of the signal respectively, stored columnwise (see [5], page 169).

3.2 Non-decimated wavelet transform

The NDWT is a redundant transformation. So there is not a unique way to define an inversion formula. The inverse NDWT can be computed via an inverse pyramid algorithm (see [5] p. 177) described as follows

$$c_{j-1,t} = \sum_{l=0}^{L-1} \tilde{h}_l c_{j,(t+2^{j-1}l) \bmod N} + \sum_{l=0}^{L-1} \tilde{g}_l w_{j,(t+2^{j-1}l) \bmod N} \tag{3}$$

where $\tilde{h}$ and $\tilde{g}$ are the scaling filter and wavelet filter normalized in l^1.
The inverse formula (3) involves circular extension but we use for prediction a simple left 'zero-padding' extension.

3.3 Kernel Smoothing

Consider the nonparametric extrapolation of a discrete signal, $X(t)$ using kernel smoothing method. Let us denote the vector of lagged variables $X_{n,(r)} = (X_n, X_{n-1}, \cdots, X_{n-r+1})$ were r is the number of the past data used in the estimation. It is well known that the autoregression function f plays an important role for the extrapolation in time series context:

$$f(x) = E(X_{n+s} | X_{n,(r)} = x)$$

for $x \in \mathbb{R}^r$ and s is extrapolation horizon. The kernel estimator $\hat{f}_n$ of f based on $X_n, ..., X_2, X_1$ is defined as follow:

$$\hat{f}_n(x) = \sum_{t=r,...,n-s} w_{n,t} X_{t+s}$$

where weight sequence $w_{n,t}$ for kernel smoothers is defined by:

$$w_{n,t} = \frac{K((x - X_{t,(r)}/h_n)}{\sum_{m=r,...,n-s} K((x - X_{m,(r)}/h_n)}$$

and K is the kernel with scale factor h_n and called as the bandwidth. This estimator is called also Nadaraya-Watson. Classical theoretical results about non parametric kernel-based estimation show that the choice of kernel function does not strongly influence the asymptotic behavior of estimator but the choice of bandwidth is crucial. Let us note that when we use this procedure, we apply it with locally centered versions of the signal to compute weights and recentered versions for predictions (see [6]).

4 Prediction by regression on wavelet coefficients

4.1 Stationary case: the procedure

Aminghafari and Poggi [1] present an approach to predict X_{N+1} based on wavelet coefficients, starting from N observations. In this paper, the first change is to consider the s-step prediction procedure. The prediction step can be modified as follows. To predict X_{N+s}, when N observations $X_1, ..., X_N$ are given, is of the following form:

$$\hat{X}_{N+s} = \sum_{j=1}^{J} \sum_{k=1}^{r_j} a_{j,k} w_{j,N-k+1} + \sum_{k=1}^{r_{J+1}} a_{J+1,k} c_{J,N-k+1}. \tag{4}$$

Denoting by D_t the explanatory variables vector and by α the parameter vector:

$$D_t = [w_{1,t}, ..., w_{1,t-2r_1}, ..., w_{J,t}, ..., w_{J,t-2^J r_J}, c_{J,t}, ..., c_{J,t-2^J r_{J+1}}]^T \tag{5}$$

$$\alpha = [a_{1,1}, ..., a_{1,r_1}, ..., a_{J,1}, ..., a_{J,r_J}, ..., a_{J+1,1}, ..., a_{J+1,r_{J+1}}]^T \tag{6}$$

the prediction equation can be written as $\hat{X}_{N+s} = {D_N}^T\alpha$. and the parameter α is estimated by minimizing the empirical mean square prediction error.

4.2 Stationary case: simulated examples

4.2.1 The considered models

In this section we consider two examples of stationary (see [1] for detailed definition): a high order $AR(14)$ to experiment longer short-dependence; and a highly nonlinear model i.e. a generalized thresholded autoregressive ($GTAR$).

4.2.2 The experimental framework

For each model, we simulate 50 realizations of time series of size $N = 2000$ denoted by $(x_1^k, \dots x_N^k)_{k=1,\dots,50}$. Each realization of size $N = 2000$ observations is divided in two groups (standing for the past and the future respectively) of size $n = 1950$ (for the past) and $N - n = 50$ (for the future). The observations $x_1, \dots, x_{n-s+1}$ is used to select the explanatory variables and to estimate α by $\hat{\alpha}$ in the prediction equation. The observations $x_{n+1}, \dots, x_{n+N}$ are predicted one by one. Then performance of the s-step forecasting procedure is evaluated by computing the standard deviation of the prediction errors on the test sample which is an estimate of the Root Mean Square Prediction Error (RMSE). We compute $\bar{R}_{pred}$ the mean over the 20 realizations of the RMSE and in addition $std(R_{pred})$ measuring the variability of the performance.

Table 1: $\mathcal{M}_{AR}$ and $\mathcal{M}_{GTAR}$: 10-Step Prediction Performance

	Wavelet	$\bar{R}_{pred}$	$std(R_{pred})$
$\mathcal{M}_{AR}$	db2	1.06	0.12
$\mathcal{M}_{GTAR}$	db2	1.09	0.14

The performance results for AR and $GTAR$ models are given in Table 1 and are not far from those obtained for $s = 1$ *i.e.* $\bar{R}_{pred} = 1.02, 1.04$ for $\mathcal{M}_{AR}$ and $\mathcal{M}_{GTAR}$ respectively (see [1]). This motivate us to formulate our proposed method for non-stationary procedure in the following section.

4.3 Non-stationary case: the procedure

Indeed, the prediction procedure sketched previously, is not designed to handle the time series involving a deterministic trend. Suppose that the observed time series are of the form $Y_t = X_t + f(t)$ when X_t is a purely stochastic time series and $f(t)$ is a deterministic component. From equation (2) we can write

$$Y_t = (\mathcal{A}_J(X))_t + (\sum_{j=1}^{J} \mathcal{D}_j(X))_t + f(t), \quad 1 \leq t \leq N \tag{7}$$

where $\mathcal{A}_j(X)$ and $\mathcal{D}_j(X)$ are suitably chosen reconstructed versions of approximation and detail of level j respectively, obtained from the NDWT coefficients of X. Let us note that we use in this paper the reconstruction version of detail coefficients obtained by inverse NDWT using zero padding at the boundary.

So, a procedure for the s-step prediction of a signal contaminated by a trend can be proposed by extending the previous one:

- Step 1: The term $\mathcal{D}_t = \left(\sum_{j=1}^{J} \mathcal{D}_j(X)\right)_t$ can be predicted using the previous procedure since the details are supposed to be free of f for a convenient wavelet choice. So first perform regression between reconstructed version of detail coefficients, $\mathcal{D}_t$, and the past wavelet coefficients $\{w_{j,k}\}_{k \leq t-s}$. Of course, this step must reject the approximation coefficients in the prediction equation by taking $r_{J+1} = 0$. Let us denote by $\hat{\mathcal{D}}_{N+s}$, the prediction of the high frequency stationary components delivered by this step.

- Step 2: Then, $(\mathcal{A}_J(X))_t + f(t)$ is estimated by

$$Z_t = Y_t - \widehat{(\sum_{j=1}^{J} \mathcal{D}_j(X))_t} = Y_t - \hat{\mathcal{D}}_t.$$

  This signal can be extrapolated following various deterministic or stochastic ways. In [1], the local polynomial fitting is used. In the sequel, we use kernel smoothing procedure (see [6]. Let us denote by

$\hat{Z}_{N+s}$, the extrapolation of the low frequency components given by this second step.

- Step 3: Finally, the prediction of the top level time series Y is given by:

$$\hat{Y}_{N+s} = \hat{\mathcal{D}}_{N+s} + \hat{Z}_{N+s}.$$

4.4 Non-stationary case: parameters choice

The Gaussian kernel is used and we focus on the choice of kernel bandwidth, h, which is of great importance. We choose an initial value for h say h_0 then we propose to select r, the number of lagged variable used in the extrapolation, by performing some kind of hold-out procedure on the past observations. Once r is selected, we select h according to a similar procedure.

4.5 Non-stationary case: two simulated examples

Let us consider a previously examined stationary model $\mathcal{M}_{AR}$ contaminated by a seasonality, for example a sinusoidal function. We compare the following methods:

- Method 1: Procedure especially designed to predict stationary process without any specific adaptation;
- Method 2: Method proposed in [1] i.e. using polynomial fitting to extrapolate low frequency components;
- Method 3: Method proposed in this paper using kernel smoothing to extrapolate low frequency components;
- Method 4: Direct Kernel smoothing without preprocessing (see [6]).

The performance results for $\mathcal{M}_{AR} + Sin.05t$ for 1-step and 10-step prediction are given in Table 2 and Table 3 respectively.

In Table 3, the result for method 3 is of good quality and only the fourth method reaches similar performance. But 1-step prediction using last method leads to degradation (see Table 2).

Table 2: $\mathcal{M}_{AR} + Sin.05t$: 1-Step Prediction Performance

Method	Wavelet	$\bar{R}_{pred}$	$stdR_{pred}$	parameters
Method 1	db2	1.02	.11	
Method 2	db2	1.03	.10	15
Method 3	db2	1.04	.11	h=3 r=100
Method 4		1.06	.13	h=3 r=100

Table 3: $\mathcal{M}_{AR} + Sin.05t$: 10-Step Prediction Performance

Method	Wavelet	$\bar{R}_{pred}$	$stdR_{pred}$	parameters
Method 1	db2	1.28	.17	
Method 2	db2	1.16	.14	15
Method 3	db2	1.09	.14	h=3 r=100
Method 4		1.09	.14	h=3 r=100

4.6 Non-stationary case: a real example

The data collected from yearly minimal water levels of Nile River for the years 654 to 1281 provide a standard benchmark data set. The performance results using for 1-step prediction: $r = 10$, $h = 70$, and for 10-step prediction: $r = 500$, and $h = 5000$, are given in Table 4 and Table 5 respectively.

Table 4: Nile: 1-Step Prediction Performance

Method	Wavelet	R_{pred}
Method 1	db2	50.09
Method 2	db2	48.50
Method 3	db2	49.4
Method 4	-	52.93

In Table 4, a better prediction performance is reached using method 3.

Table 5: Nile: 10-Step Prediction Performance

Method	Wavelet	R_{pred}
Method 1	db2	1172.7
Method 2	db2	67.03
Method 3	db2	58.1
Method 4		55.01

In Table 5, methods using kernel smoothing overcome others and, as expected, the worse result is obtained for first method designed for stationary process.

References

[1] M. Aminghafari and J.-M. Poggi (2007). Forecasting time series using wavelets. *International Journal of Wavelets, Multiresolution and Information Processing*, 5(5):709–724.

[2] A. Antoniadis (1997). Wavelet in statistics: a review. *J. Ital. Statist. Soc.*, 6:97–144.

[3] A. Antoniadis, E. Paparoditis, and T. Sapatinas (2006). A functional wavelet-kernel approach for time series prediction. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68(5):837–857.

[4] M. Misiti, Y. Misiti, G. Oppenheim, and J-M. Poggi (2003). *Les ondelettes et leurs applications*. Hermes.

[5] D. B. Percival and A. T. Walden (2000). *Wavelet methods for time series analysis*. Cambridge University Press.

[6] J.-M. Poggi (1994). Prévision non paramétrique de la consommation électrique. *Revue de Statistique Appliquée*, 42(4):83–98.

[7] O. Renaud, J.-L. Starck, and F. Murtagh (2003). Prediction based on a multiscale decomposition. *International journal of Wavelet, Multiresolution and Information Processing*, 1:217–232.

[8] B. Vidakovic (1999). *Statistical modeling by wavelets.* Wiley.

Fisher information in record values and their concomitants under the Gumbel's bivariate exponential distribution

Morteza Amini, J. Ahmadi

Department of Statistics, Ferdowsi University of Mashhad

Abstract: Suppose (X, Y) is distributed as one of the most famous bivariate distributions, namely Gumbel's bivariate exponential distribution Type II (GBVE). We investigate some properties of the Fisher Information (FI) contained in the sequence of record values and their concomitants $(\{(R_n, R_{[n]}),\ n \geq 1\})$ about the unknown parameters of model. Three schemes are considered and in each scheme one of the parameters of GBVE distribution is assumed to be unknown the Fisher information about the unknown parameter contained in record values and their concomitants is investigated. The behavior of information measure with respect to n and other known parameters is studied. Also, we calculate and tabulate the relative efficiency of the optimal estimator (i.e. the estimator with a variance equal to the Crame'r Rao lower bound) of unknown parameter based on the first n records and their concomitants with respect to the desired estimator on the basis of a sample of size n of pairs (X, Y) from the parent distribution.

Keywords: Bivariate distribution; Additivity; Relative efficiency; Farlie-Gumbel-Morgenstern model; Copula.

1 Introduction

One of the most applicable measures of information which is applied in many engineering and mathematical problems is the well-known Fisher information. It also plays an important role in statistical estimation and inference. In estimation problems this measure is concerned through the information (Cramér-Rao) inequality and its association with the asymptotic properties, specially the asymptotic variance of the maximum likelihood estimators. The Fisher information contained in X about the parameter θ is denoted by $I_X(\theta)$.

Let $\{(X_i, Y_i), i \geq 1\}$ be a sequence of pair-wise random variables from a continuous distribution with parameter θ. If observe the record values of X, then the corresponding Y-value of each record statistic is termed the concomitant of that record value. The reader is referred to Raqab and

Ahsanullah (2002) for a general review of concomitants of ordered random variables. Amini and Ahmadi (2007) considered the dependence and correlation parameters to investigate the properties of Fisher information contained in record values and their concomitants. Amini and Ahmadi (2008) studied the comparing of Fisher information in record values and their concomitants with random observations in a general view on parameters. Abo-Eleneen and Nagaraja (2002) investigated some properties of Fisher information in an order statistic and its concomitant. Recently, Abo-Eleneen (2007) developed the results of Abo-Eleneen and Nagaraja (2002) for generalized order statistics and its concomitant. Also Abo-Eleneen and Nagaraga (2008) studied some properties of Fisher information in bivariate censored samples.

The rest of this paper is organized as follows. Section 2 contains some preliminaries. In Section 3, we consider GBVE distribution and obtained the Fisher information contained in record values and their concomitants in three schemes. In addition the additivity property of Fisher information is used to calculate Fisher information contained in the sequence of first n records and their concomitants about θ and compared this information with that of i.i.d. sample of same size from the parent distribution. We derived a closed form for the pdf of $R_{[n]}$. For the case $\theta = \mathrm{E}(Y)$ we also calculate the Fisher information contained in $R_{[n]}$ about θ and compare it with the desired informations contained in a single Y. Using the results, we investigate the behavior of the values of Fisher information with respect to n and other known assumed parameters.

2 Preliminaries

Suppose $\{(X_i, Y_i), i \geq 1\}$ be a sequence of i.i.d. pair-wise random variables with the absolutely continuous cumulative distribution function (cdf) $F_{(X,Y)}(x, y; \theta)$ and the corresponding pdf $f_{(X,Y)}(x, y; \theta)$. Also $f_X(x; \theta)$ and $F_X(x; \theta)$ denote the marginal pdf and cdf of Xs, respectively and $\bar{F}_X(x; \theta) = 1 - F_X(x; \theta)$. The sequence of upper records and their concomitants is defined as follows $[(R_n, R_{[n]}) = (X_{T_n}, Y_{T_n}), \quad n \geq 1]$, where $(T_1 = 1)$ with probability one and for $n \geq 2$ $[T_n = \min j : j > T_{n-1},\ X_j > X_{T_{n-1}}.]$ The joint pdf of the first n upper records and their concomitants is (See

Arnold et al., 1998)

$$f_{(\mathbf{R}_n,\mathbf{R}_{[n]})}(\mathbf{r}_n,\mathbf{r}_{[n]}) = \prod_{i=1}^{n} f_{(X,Y)}(r_i,s_i;\theta) / \prod_{i=1}^{n-1}[\bar{F}_X(r_i;\theta)], \tag{1}$$

where $(\mathbf{R}_n,\mathbf{R}_{[n]}) = (R_1,\ldots,R_n,R_{[1]},\ldots,R_{[n]})$. Also $(R_n,R_{[n]})$ has the joint pdf

$$f_{R_n,R_{[n]}}(x,y;\theta) = \frac{1}{(n-1)!} f(x,y;\theta)\left[\bar{H}(x;\theta)\right]^{n-1}, \tag{2}$$

where $\bar{H}(x;\theta) = -\log(\bar{F}_X(x;\theta))$.

In order to obtain the new results in this paper, we need to recall the following results.

Lemma 1 (Lehmann 1989) *Let $I_X(\theta)$ be the Fisher information about θ contained in observing X. Suppose $\theta = z(\eta)$ where $z(.)$ is differentiable, then $[I^*_X(\eta) = z'(\eta)^2 I_X(z(\eta)),]$ where $I^*_X(\eta)$ is the information about η contained in X.*

Theorem 1 *(Amini and Ahmadi 2007) If F_X is free of θ, then $I_{(R_n,R_{[n]})}(\theta)$ is additive; that is*

$$I_{(\mathbf{R}_n,\mathbf{R}_{[n]})}(\theta) = \sum_{i=1}^{n} I_{(R_i,R_{[i]})}(\theta). \tag{3}$$

3 Gumbel's bivariate exponential distribution

There are two types of Gumbel's bivariate exponential distribution. Type I is a sub-family of the class of Farlie-Gumbel-Morgenstern (FGM) bivariate distribution with joint pdf

$$f_{X,Y}(x,y;\theta) = f_X(x)f_Y(y)[1+\theta(1-2F_X(x))(1-2F_Y(y))], -1<\theta<1,$$

where θ is the dependence parameter of this copula, such that the marginal pdf's of X and Y are kept exponential. Amini and Ahmadi (2007) showed that the formula of Fisher information contained in record values and their

concomitants in this class does not depend on the marginal distributions. In other word this information measure is distribution-free and only depends on the dependence model of copula. They obtained an explicit formula for the Fisher information contained in a record value and its concomitant about θ as

$$I_{(R_n,R_{[n]})}(\theta) = \sum_{j=0}^{\infty} \frac{\theta^{2j}}{2j+3} \sum_{k=0}^{2j+2} \left(\begin{array}{c} 2j+2 \\ k \end{array} \right) \frac{(-2)^k}{(k+1)^n}. \quad (4)$$

The authors also showed that for $n \geq 1$, $I_{(R_n,R_{[n]})}(\theta) > I_{(X,Y)}(\theta)$. Furthermore $I_{(R_n,R_{[n]})}(\theta) = I_{(R_n,R_{[n]})}(-\theta)$ and $I_{(R_n,R_{[n]})}(\theta)$ increases as $|\theta|$ increases.

In this paper, we consider Gumbel's bivariate exponential distribution type II (see Johnson and Kotz, 1972, p. 261), with the pdf

$$f_{(X,Y)}(x,y) = \left\{ \frac{(1+\alpha\frac{x}{\lambda_1})(1+\alpha\frac{y}{\lambda_2}) - \alpha}{\lambda_1\lambda_2} \right\} \exp\left\{ -\frac{x}{\lambda_1} - \frac{y}{\lambda_2} - \frac{\alpha xy}{\lambda_1\lambda_2} \right\}, \quad (5)$$

where $x,\ y > 0,\ \lambda_1,\ \lambda_2 > 0,\ 0 \leq \alpha \leq 1$. Our main aim is to investigate the Fisher information contained in $(\mathbf{R}_n, \mathbf{R}_{[n]})$ about each of the parameters of (5).

3.1 Fisher information about the mean of X

Let $\lambda_2 = 1$, $\lambda_1 = \theta^{-1}$ and $\alpha = a$ which is a known real number in (0,1). Then (5) can be rewritten as

$$f_{(X,Y)}(x,y) = \theta \left\{ (1+a\theta x)(1+ay) - a \right\} \exp\left\{ -(\theta x + y + a\theta xy) \right\}. \quad (6)$$

Using Lemma 1, without loss of generality we can investigate the Fisher information about θ instead of λ_1. The following Theorem presents a formula for calculation of the Fisher information contained in a record value and its concomitant about θ.

Theorem 2 *For GBVE distribution with the pdf in (6), we have*

$$\theta^2 I_{(R_n,R_{[n]})}(\theta) = n + \frac{1}{(n-1)!} \int_0^{\infty} J_a(x) x^{n-1} dx, \quad (7)$$

where

$$J_a(x) = \frac{a^3 e^{-x} x^2}{(1+ax)^3}\left(\frac{2a+1+ax}{a} + Ei\left(\frac{1+ax-a}{a}\right)\exp\left(\frac{1+ax-a}{a}\right)\right)$$

and

$$Ei(u) = \int_u^\infty \frac{e^{-t}}{t} dt.$$

Proof Using (8) we have

$$f_{R_n,R_{[n]}}(x,y;\theta) = f(x,y;\theta)\frac{x^{n-1}}{(n-1)!}.$$

So

$$\begin{aligned}\frac{\partial^2}{\partial\theta^2}\log f_{(R_n,R_{[n]})}(x,y) &= \tfrac{x^{n-1}}{(n-1)!}\frac{\partial^2}{\partial\theta^2}\log f_{(X,Y)}(x,y)\\ &= \tfrac{-1}{\theta^2}\left\{\left[\frac{a\theta x(1+ay)}{(1+a\theta x)(1+ay)-a}\right]^2 + n\right\}.\end{aligned} \tag{8}$$

Hence

$$J_a(x) = -\theta^2\int_0^\infty \frac{\partial^2}{\partial\theta^2}\log f_{(R_n,R_{[n]})}(x,y) f_{(X,Y)}(x,y)\, dy.$$

The proof is complete by integration with a change of variable of $t = (1+a\theta x)(1+ay) - a$ in the integral of the first term of (8). □

Table 2 shows the values of $\theta^2 I_{(R_n,R_{[n]})}(\theta)$ for $n = 3(2)7, 10$ and $a = 0.1(0.1)0.9$ calculated using Maple 11.0. This values increases by increasing n and a. But as it can be observed, the increment of information with respect to n is more distinct. Indeed the values of $\theta^2 I_{(R_n,R_{[n]})}(\theta)$ consist of an integer part which equals n and the decimal counterpart which equals the second term of right hand side of (7).

3.2 Fisher information about dependence parameter

Let $\lambda_1 = \lambda_2 = 1$ and $\theta = \alpha$ be the unknown parameter of (5). Hence we have

$$f_{(X,Y)}(x,y) = \{(1+\theta x)(1+\theta y) - \theta\}\exp\{-(x+y+\theta xy)\}, \tag{9}$$

The amount of Fisher information contained in a record value and its concomitant about θ is given by the next theorem.

Table 1: The values of $\theta^2 I_{(R_n,R_{[n]})}(\theta)$ from GBVE distribution for $\theta = (E(X))^{-1}$.

	a								
n	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3	3.065	3.171	3.277	3.376	3.467	3.549	3.624	3.693	3.756
5	5.125	5.288	5.424	5.536	5.628	5.706	5.771	5.828	5.877
7	7.186	7.382	7.527	7.637	7.721	7.788	7.842	7.887	7.925
10	10.268	10.490	10.632	10.729	10.800	10.853	10.895	10.928	10.956

Theorem 3 *The Fisher information contained in the nth record value and its concomitant about the dependence parameter θ for (9) equals*

$$I_{(R_n,R_{[n]})}(\theta) = \frac{1}{(n-1)!}\int_0^\infty k_\theta(x)x^{n-1}dx,$$

where

$$\begin{aligned} k_\theta(x) &= \frac{-2xe^{-x}}{(1+\theta x)^2} + \frac{1+\theta x}{\theta}e^{-x}\left\{\frac{\theta x^2}{1+\theta x} - \frac{2\theta x}{(1+\theta x)^2} - \frac{(1+\theta x)}{\theta}\right. \\ &\quad \left. + \exp\left\{\frac{1+\theta x-1}{\theta}\right\}Ei\left\{\frac{1+\theta x-1}{\theta}\right\}\right. \\ &\times \left.\left[\left(\frac{\theta x}{1+\theta x} - \frac{1+\theta x}{\theta}\right)^2 + 2\frac{1+\theta x}{\theta} - \frac{x^2}{1+\theta x} + 2\right]\right\}. \end{aligned}$$

Proof As in the proof of the last result since we have

$$\frac{\partial^2}{\partial\theta^2}\log f_{(R_n,R_{[n]})}(x,y) = \frac{2xy}{(1+\theta x)(1+\theta y)-\theta} - \left[\frac{x(1+\theta y)+y(1+\theta x)-1}{(1+\theta x)(1+\theta y)-\theta}\right]^2. \tag{10}$$

So

$$k_\theta(x) = -\int_0^\infty \frac{\partial^2}{\partial\theta^2}\log f_{(R_n,R_{[n]})}(x,y)f_{(X,Y)}(x,y)\,dy$$

and the result follows by integration with a change of variable of $t = (1+\theta x)(1+\theta y)-\theta$ in the integral of the second term of (10). □

Applying Theorem 1, we calculate the amount of Fisher information contained in the sequence of first n records and their concomitants about θ. We define the ratio of the Fisher information contained in $(\mathbf{R}_n, \mathbf{R}_{[n]})$

Table 2: The values of RE(RC,IID) from GBVE distribution about dependence parameter.

	θ								
n	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3	0.9612	1.0605	1.1788	1.2795	1.3547	1.4049	1.4326	1.4388	1.4199
4	0.9792	1.1767	1.3903	1.5656	1.6938	1.7787	1.8262	1.8392	1.8135
5	1.0234	1.3500	1.6811	1.9453	2.1351	2.2597	2.3292	2.3491	2.3148
7	1.1928	1.8678	2.4975	2.9802	3.3180	3.5353	3.6539	3.6864	3.6269
10	1.6568	3.0747	4.3054	5.2141	5.8330	6.2205	6.4243	6.4705	6.3512

about θ to the desired information contained in an i.i.d. sample of size n as

$$\text{RE(RC,IID)} = \frac{I_{(\mathbf{R}_n,\mathbf{R}_{[n]})}(\theta)}{nI_(X,Y)}.$$

This information ratio is also the relative efficiency of the optimal estimator of θ based on the first n records and concomitants with respect to that of an i.i.d. sample of the same size. Table 2 contains the values of RE(RC,IID) for $n = 3, 4, 5, 7, 10$ and $\theta = 0.1(0.1)0.9$. From these values we observe that for $n \leq 4$ an i.i.d. sample is more informative than first n records and concomitants while $\theta = 0.1$. But as n and θ increase the sequence of first n records and concomitants gets more informative than i.i.d. sample. The values of RE(RC,IID) increases as n increases and also have a maximum value for θ near 0.8 for considered values of n.

3.3 Fisher information about mean of Y

Let $\lambda_2 = \theta^{-1}$, $\alpha = a$ and $\lambda_1 = \lambda$ in (5), where a and λ are known values. So

$$f_{(X,Y)}(x,y) = \lambda\theta\left\{(1+a\lambda x)(1+a\theta y) - a\right\}\exp\left\{-\lambda x - \theta y - a\lambda\theta xy\right\}, \tag{11}$$

In this subsection, we are interested in the amount of Fisher information in the nth concomitant about θ in addition to the desired information contained in both nth record value and its concomitant. Hence, first of all, we need the marginal pdf of $R_{[n]}$.

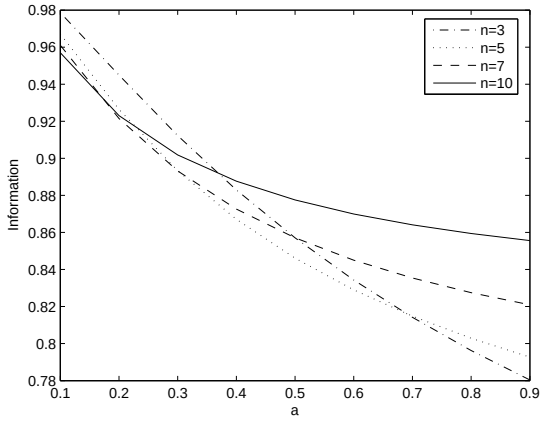


Figure 1: $\theta^2 I_{R_{[n]}}(\theta)$ for different values of a and n.

Lemma 2 *For Gumbel's bivariate exponential distribution with the pdf in (11), the pdf of* $R_{[n]}$ *is*

$$f_{R_{[n]}}(y;\theta) = \theta \frac{1+(n-1)a+a\theta y}{(1+a\theta y)^n} e^{-\theta y}. \tag{12}$$

Proof Using (8) the marginal pdf of $R_{[n]}$ is given by

$$f_{R_{[n]}}(y;\theta) = \int_0^\infty f(x,y) \frac{[\bar{H}(x;\theta)]^{(n-1)}}{(n-1)!} dx.$$

So by applying (11) we have

$$\begin{aligned} f_{R_{[n]}}(y;\theta) = \frac{1}{(n-1)!} \int_0^\infty & \exp\{-\lambda x - \theta y - a\lambda\theta xy\} \\ & \times \{\lambda\theta(1+a\lambda x)(1+a\theta y) - a\}(\lambda x)^{(n-1)} dx. \end{aligned}$$

Hence (12) is obtained by integration. □ The following result presents a formula for calculation of $I_{R_{[n]}}(\theta)$.

Theorem 4 *For GBVE distribution with the pdf in (11), we have*

$$\theta^2 I_{R_{[n]}}(\theta) = 1 + \Psi_1(n,a) - \Psi_2(n,a),$$

where

$$\Psi_1(n,a) = a^2 \int_0^\infty \frac{z^2 e^{-z}}{(1+(n-1)a+az)(1+az)^n}\, dz$$

and

$$\Psi_2(n,a) = na^2 \int_0^\infty \frac{z^2(1+(n-1)a+az)e^{-z}}{(1+az)^{n+2}}\, dz.$$

Proof Using Lemma 2, we get

$$\frac{\partial^2}{\partial\theta^2}\log f_{R_{[n]}}(y;\theta) = \frac{1}{\theta^2}\left\{-1-\left[\frac{a\theta y}{1+(n-1)a+a\theta y}\right]^2 + n\left[\frac{a\theta y}{1+a\theta y}\right]^2\right\}.$$

So the proof is complete by taking expectation with respect to (12). □
Figure 1 shows the values of $\theta^2 I_{R_{[n]}}(\theta)$ for different values of a and n. As it can be seen from the figure, the amount of information decreases as a increase. But the behavior of information with respect to n depends on the value of a. For a lower dependence, the first concomitants and for a higher dependence the further ones are more informative. Also the values of $\theta^2 I_{R_{[n]}}(\theta)$ stand for the ratio of $I_{R_{[n]}}(\theta)$ relative to $I_Y(\theta) = 1/\theta^2$. As it can be seen from Figure 1, these values are all less than 1.

In the next result, let for the sake of simplicity, $\lambda = 1$.

Theorem 5 *For GBVE distribution with the pdf in (11), we have*

$$\theta^2 I_{(R_n,R_{[n]})}(\theta) = 1 + \int_0^\infty h_a(x)\frac{x^{n-1}}{(n-1)!}dx,$$

where

$$h_a(x) = ae^{-x}(1+ax)^{-1}(1+(\frac{1+ax-a}{a})^2\exp(\frac{1+ax-a}{a}) \times Ei(\frac{1+ax-a}{a}) - (\frac{1+ax-a}{a})).$$

Proof For this case we have

$$\frac{\partial^2}{\partial\theta^2}\log f_{(R_n,R_{[n]})}(x,y) = \frac{-1}{\theta^2}\left\{\left[\frac{a\theta y(1+ax)}{(1+ax)(1+a\theta y)-a}\right]^2 + 1\right\}.$$

Table 3: The values of RE(RC,IID) from Gumbel's bivariate exponential distribution for $\theta = E(Y)$.

	a				
n	0.1	0.3	0.5	0.7	0.9
3	0.9981	0.9807	0.9539	0.9248	0.8958
5	0.9967	0.9687	0.9288	0.8873	0.8472
7	0.9955	0.9607	0.9132	0.8652	0.8196
10	0.9942	0.9526	0.8986	0.8452	0.7957

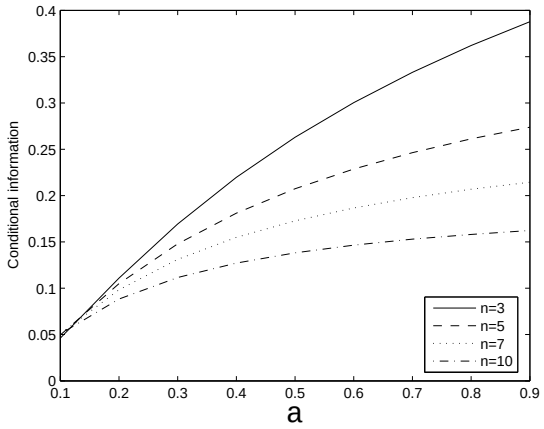


Figure 2: $\theta^2 I_{Rn|R_{[n]}}(\theta)$ for different values of r and n.

Hence

$$\begin{aligned} h_a(x) &= \int_0^\infty \left[\frac{ay(1+ax)}{(1+ax)(1+ay)-a}\right]^2 f_{(X,Y/\theta)}(x,y)\, dy \\ &= ae^{-x}(1+ax)^{-1}(1+(\frac{1+ax-a}{a})^2 \exp(\frac{1+ax-a}{a}) \\ &\times \quad Ei(\frac{1+ax-a}{a}) - (\frac{1+ax-a}{a})). \end{aligned}$$

□

Table 3 Shows the values of RE(RC,IID) $= I_{(\mathbf{R}_n,\mathbf{R}_{[n]})}(\theta)/nI_{(X,Y)}(\theta)$ for $n = 3(2)7, 10$ and $a = 0.1(0.2)0.9$. As it can be seen from Table 3, the i.i.d.

sample is more informative than the sequence of record values and their concomitants. Also these values decrease as n or a increase.

The next corollary formulate the increment of information when we consider the record value in addition to its concomitant, i.e.

$$I_{Rn|R_{[n]}}(\theta) = I_{(Rn,R_{[n]})}(\theta) - I_{R_{[n]}}(\theta).$$

Corollary 1 *The increment of Fisher information about the mean of Y when we consider the record value in addition to its concomitant in GBVE distribution can be calculated using the following equality*

$$\theta^2 I_{R_n|R_{[n]}}(\theta) = \Psi_1(n,a) + \Psi_2(n,a) + \int_0^\infty h_a(x)\frac{x^{n-1}}{(n-1)!}dx.$$

Figure 2 shows this increment for different values of n and a. As it can be seen, these values increase for higher dependence. But the behavior of this increment with respect to n depends on a. Indeed, except for lower values of a increasing n cause decreasing of information increment.

References

[1] Abo-Eleneen, Z. A. and Nagaraja, H. N., 2002. Fisher information in an order statistic and its concomitant. *Ann. Inst. Statist. Math.* **54**, No. 3, 667 – 680.

[2] Abo-Eleneen, Z. A., 2007. Fisher information in generalized order statistics and their concomitant. *J. Stat. Theory Appl.***6**, 156 – 173.

[3] Abo-Eleneen, Z. A. and Nagaraja, H. N., 2008. Fisher information in order statistics and their concomitants in bivariate censored samples. *Metrika* **67**, 327 – 347.

[4] Amini, M. and Ahmadi, J., 2007. Fisher information in record values and their concomitants about the dependence and correlation parameters. *Statist. Probab. Lett.* **77**, 964 – 972.

[5] Amini, M. and Ahmadi, J., 2008. Comparing Fisher information in record values and their concomitants with random observations. To appear in *Statistics*.

[6] Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N., 1998. *Records*, John Wiley & Sons, New York.

[7] Johnson, N. L. and Kotz, S., 1972. *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley, New York.

[8] Raqab, M. Z. and Ahsanullah, M., 2002. Concomitants of ordered random variables–A review. *J. Stat. Theory App.* **1**, 15 – 26.

Jarque Bera Normality Test and Its Modification

S. M. Amir Jahanshahi, H. Naderi, and R. Moayeripour

Department of Statistics, faculty of Mathematics, Sistan and Baluchestan University

Abstract: Many statistical tests have been proposed to find out whether a sample is drawn from a normal distribution or not. First we discuss the Jarque-Bera test (1981) which is based on the Karl Pearson's measure of skewness and kurtosis, then we write about problems associated with the JB test and we review modified JB (MJB) test proposed by Urzua (1996). Consequently, we write about Geary's skewness and kurtosis and G test proposed by Cho and Im (2002), Moreover, we review Fisher's measures of skewness and kurtosis and introduce a new JB test (FJB) based upon Fisher's measures. After that, we discuss about outlier poroblem and sensitivity of skewness and kurtosis to outlying values, and then, we review a Robust JB test (RJB) proposed by Gel and Gastwirth (2006). After that, we review Quantile measures of skewness (Bowley and Kelly) and Quantile measure of kurtosis (K_q) and two other Quantile measures proposed by Moors (1996), after all, we introduce two new JB test based on reviewed Quantile measures and finally we compare the power of all introduced tests via simulation.

Keywords: Normality, Skewness, Kurtosis, Jarque-Bera, Quantile, Outlier, Robust, Power, Simulation.

1 Introduction

Normality test is important for some theoretical and empirical research and The validity of several parametric statistical inference procedures depends on the underlying distributional assumptions. Several parametric statistical tests assumed to be distribution of data is normal, It is well known that departures from normality may lead to substantially incorrect statements in the analysis. Therefore, testing for normality is an important issue. see for example Bowman and Shenton (1975), Shenton and Bowman (1977), Jarque and Bera (1980, 1987), D'Agostino and Stephens (1986), Spiegelhalter (1980), Thode (2002). Jarque-Bera test (JB hereafter) is simple to compute, and the power has proved comparable to other powerful tests such as Shapiro-Wilk test (1965) or Shapiro-Francia test (1972) that are computationally more demanding. See Mardia (1980) for an exhaustive account of

the various normality tests. Simulation results comparing the power of the JB tests with other tests were reported by Pearson, D'Agostino and Bowman (1977), Jarque and Bera (1987), and Deb and Sefton (1996). The JB test was uncovered by Jarque and Bera as the Lagrangian multiplier (LM) test against the Pearson family distributions. Also, Geary (1947) showed that the third and the fourth sample moments used in the JB statistic are most sensitive asymptotically in testing for skewness and kurtosis assuming Gram-Charlier density. Therefore, the JB test shares the nice asymptotic optimal property against a wide class of non-normal distributions.

The sample skewness and kurtosis statistics are excellent descriptive and inferential measures for evaluating normality. Any test based on skewness or kurtosis is usually called an omnibus test. An omnibus test is sensitive to various forms of departure from normality. Among the commonly used tests for normality, the Jarque-Bera (1980) test (JB), D'Agostino's (1971) D test, and Pearson's (1900) χ^2 goodness of fit test (χ^2 test) are omnibus tests. Another famous test is Shapiro Wilk test (1965), which is based on order statistics, or tests based on the distance between the empirical distribution function and the normal cumulative distribution function such as the Kolmogorov Smirnov, the Cramer von Mises, or the Anderson Darling test have been proposed, A test based on L_2 distance between Gaussian and empirical characteristic functions has been introduced by Epps and Pulley (1983) and developed by Henze and others. For more details see Mardia (1980), Henze (1997), Epps (1999), and references therein. For the independent case, the omnibus tests are consistent, but it has been shown that their finite sample performance can be very poor (see Shapiro, Wilk, and Chen, 1968).

The most often discussed problem associated with the JB test would be the slow convergence of the test statistic to its limiting distribution, which makes the test under-sized even in a reasonably large sample. Another problem, which has not drawn much attention in the literature, is that the JB test has in general a weak power against platykurtic distributions. Although we tend to be more concerned with the situation of heavy-tailed population in practice, there will be many situations where we cannot exclude the possibility of short-tailed distribution.

The slow convergence of the JB statistic to its limiting distribution

would be associated with the estimation of the third and fourth moments: The average of the cubed and fourth powered normal variables converges very slowly to a normal distribution. The second problem, the low power property of the JB test against the platykurtic distributions would be explained as follows: The sample skewness and kurtosis of the JB statistic are standardized by the variances under normality, which are usually bigger than the actual variances when the distribution is platykurtic. Dividing by a bigger value than its actual variance makes both the non-centrality and the variance of the statistic smaller than those of usual noncentral chi-square distribution. Therefore, the statistic could be clustered below the critical value to make the power smaller than the size. However, as the sample size grows and the distribution gets away from the null distribution, we may expect a rapid increase of power since the variance of the statistic is small.

The opposite would be true for leptokurtic distributions; the skewness and kurtosis are divided by smaller values than the actual variance. Therefore, the distribution of the JB statistic would be centered even farther away from the null distribution but with larger variance. We may expect a good power property especially when the sample size is relatively small (see Cho and Im, 2002).

In this paper we develop some modified and robust JB normality tests, we introduce Fisher JB test (FJB) and three Quantile JB test (QJB) based upon quantiles measures of skewness and kurtosis.

In the following sections we describe details of JB test and the proposed tests, and conduct a Monte Carlo study to compare the tests. We conclude in the final section.

2 Pearson's Skewness and Kurtosis

Suppose we have a random sample $\{x_1, x_2, ..., x_n\}$, obtained from an independent population. The third and fourth moments of a distribution are called the skewness and kurtosis. For any distribution F with finite central moments μ_k (k ≤ 3), write $\mu_i = E(X-\mu)^i$, so that $\sigma^2 = \mu_i$. The pearson's *skewness* is defined as

$$\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}} \tag{1}$$

Skewness describes the asymmetry of a distribution. A symmetric distribution has zero skewness, an asymmetric distribution with the largest tail to the right has positive skewness, and a distribution with a longer left tail has negative skewness. For any distribution F with finite central moments μ_k (k ≤ 4), the pearson's *kurtosis* is defined as

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \tag{2}$$

There is no agreement on what it really measures. Strictly speaking, kurtosis measures both peakedness and tail heaviness of a distribution relative to that of the normal distribution. Consequently, its use is restricted to symmetric distributions. Finite-sample versions of β_1 and β_2 will be denoted by b_1 and b_2.

Let the j-th central sample moment be

$$m_j = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^j \quad , \quad for \quad j = 2, 3, \ldots \tag{3}$$

where $\bar{x}$ is the sample mean. Therefore, $\widehat{\sigma}^2 = m_2$. Sample counterparts are defined by

$$b_1^{1/2} = \frac{m_3}{m_2^{3/2}} = \frac{m_3}{\widehat{\sigma}^3} \quad , \quad b_2 = \frac{m_4}{m_2^2} = \frac{m_4}{\widehat{\sigma}^4} \tag{4}$$

3 Simple Jarque Bera

The Jarque-Bera (1980, 1987) Lagrange multiplier test is perhaps the most commonly used procedure for testing whether a univariate sample of n data points are drawn from a normal distribution. The test is based on a joint statistic using skewness and kurtosis coefficients. The JB statistic is

$$JB = n[\frac{(b_1^{1/2})^2}{6} + \frac{(b_2 - 3)^2}{24}] \tag{5}$$

If $\{x_i\}$ is normally distributed and iid, then

$$\sqrt{n}(b_1^{1/2} - 0) \xrightarrow{D} N(0, 6) \quad , \quad \sqrt{n}(b_2 - 3) \xrightarrow{D} N(0, 24) \tag{6}$$

and

$$Z_1 = \frac{\sqrt{n}(b_1^{1/2} - 0)}{\sqrt{6}} \xrightarrow{D} N(0,1) \quad , \quad Z_2 = \frac{\sqrt{n}(b_2 - 3)}{\sqrt{24}} \xrightarrow{D} N(0,1) \tag{7}$$

so we have

$$Z_1^2 + Z_2^2 \xrightarrow{D} \chi_2^2 \tag{8}$$

It is a joint test of the null hypothesis (of normality) that sample skewness equals 0 and sample kurtosis equals 3 (H_0: β_1=0 , β_2=3). Bera and Jarque (1981) showed that

$$JB = n[\frac{(b_1^{1/2})^2}{6} + \frac{(b_2 - 3)^2}{24}] \xrightarrow{D} \chi_{1-\alpha,2}^2 \tag{9}$$

The JB is asymptotically chi-squared distributed with two degrees of freedom ($n \geq 50$) because JB is just the sum of squares of two asymptotically independent standardized normals, see Bowman and Shenton, 1975). That means: H_0 has to be rejected at level α if JB $\geq \chi_{1-\alpha,2}^2$. The JB test is known to have good power properties in testing for normality, it is clearly easy to compute and it is commonly used for testing normality. One limitation of the test is that it is designed only for testing normality, while there exist some test can be applied to test for any types of underlying distribution. It has been noted that the small-sample tail quantiles of the JB statistic are quite different from their asymptotic counterparts; e.g. Deb and Sefton (1996) and Urzua (1996). The use of asymptotic critical values given even fairly large samples will distort the actual size of the test, and may lead to incorrect decisions in applied work.

4 Modified Jarque Bera

Simulation studies revealed that the size of the Jarque-Bera test is incorrect for small- and moderate-size samples (Poitras, 1992; Durfour, Farhat, Gardial and Khalaf, 1998; among other). Urzua (1996) introduced a modification of the JarqueBera test by standardizing the skewness and kurtosis in the formula of JB appropriately in the following way

$$MJB = [\frac{(\sqrt{b_1})^2}{var(b_1)} + \frac{(b_2 - E(b_2))^2}{var(b_2)}] \tag{10}$$

with

$$var(b_1) = \frac{6(n-2)(n+1)}{(n+1)(n+3)} \tag{11}$$

$$E(b_2) = \frac{3(n-1)}{(n+1)} \tag{12}$$

$$var(b_2) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)} \tag{13}$$

Notice, that JB and MJB are asymptotically equivalent, i.e. H_0 has to be rejected at level α if JB $\geq \chi^2_{1-\alpha,2}$. Moreover, compared with the Jarque-Bera test, the modified Jarque-Bera test slightly improves the size performance.

5 Geary's Skewness and Kurtosis

Define the moments of the product of the sign and the power of $(x-\mu)$;

$$s\mu_j = E[sgn(x-\mu).(x-\mu)^j] \quad , \quad j = 1, 2, ... \tag{14}$$

where μ is the population mean, and $sgn(x-\mu)$ denotes the sign of $(x-\mu)$. When j is odd, $s\mu_j = E|x-\mu|^j$. We have, for symmetric distribution of x, $s\mu_j = 0$ when j is even. Sample counterparts of $s\mu_j$ is

$$sm_j = \frac{1}{n}\sum_{i=1}^{n}[sgn(x-\mu).(x-\mu)^j] \quad , \quad j = 1, 2, ... \tag{15}$$

6 G test

Cho and Im (2002) proposed G normality test using lower moments studied earlier by R.C. Geary (1935, 1936, 1947). G test proposed based on the first two moments; $s\mu_1$ and $s\mu_2$. if let

$$a_j = \frac{sm_j}{\widehat{\sigma}^j} \quad , \quad j = 1, 2, ... \tag{16}$$

the G statistic is

$$G = n[\frac{(a_1 - \sqrt{\frac{2}{\pi}})^2}{(1 - \frac{3}{\pi})} + \frac{a_2^2}{(3 - \frac{8}{\pi})}] \tag{17}$$

As we will see, under normality, this G statistic approaches to the chi-square distribution with two degrees of freedom as the sample size grows.

The two statistics a_1 and a_2 have been studied, among others, by Geary (1935, 1936, 1947) and Gastwirth and Owens (1977). For example, $sm_1 = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$. Therefore, a_1 is the ratio of the mean deviation and standard deviation proposed by Geary (1935) as a statistic for normality test. Under normality a_1 converges strongly to the expected value of the mean deviation of standard normal, which is $\sqrt{\frac{2}{\pi}}$. Therefore, deviation of a_1 from $\sqrt{\frac{2}{\pi}}$ is viewed as an evidence of non-normality. Geary (1935) showed that when x is iid normal;

$$\sqrt{n}(a_1 - \sqrt{\frac{2}{\pi}}) \rightarrow N(0, 1 - \frac{3}{\pi}) \tag{18}$$

It is obvious that a_2 converges to zero under normality. Geary (1947) proved that

$$\sqrt{n}a_2 \rightarrow N(0, 3 - \frac{8}{\pi}) \tag{19}$$

Combining the results of (18) and (19), and noting that, as was shown by Gastwirth and Owens (1977), $\sqrt{n}a_1$ and $\sqrt{n}a_2$ are independent asymptotically under normality, the G statistic follows the chi-square distribution with two degrees of freedom in the limit. That means: H_0 has to be rejected at level α if G $\geq \chi^2_{1-\alpha,2}$.

The power comparison between G and JB depends on the shape of the distribution under test. However, the JB test suffers an extreme power deficit when the population distribution has short-tails. The G test has more stable power against various directions of the deviation from normality. All in all, it seems that the G test is a good alternative to the JB test especially when the sample size is moderate to small, unless we have a strong a priori belief that the population distribution of the sample we put to a normality test cannot be platykurtic.

7 Fisher's Skewness and Kurtosis

Natural estimators of the population skewness ($b_2^{1/2}$) and kurtosis (b_2) are biased. The unbiased estimators of the moments are:

$$K_2 = \frac{n}{(n-1)} \cdot m_2 \tag{20}$$

$$K_3 = \frac{n^2}{(n-1)(n-2)} \cdot m_3 \tag{21}$$

$$K_4 = \frac{n^2}{(n-1)(n-2)(n-3)} \cdot (n+1)m_4 - 3(n-1)m_2^2 \tag{22}$$

Measures based on the unbiased estimators of moments (The Fisher's estimators) are

$$G_1 = \frac{K_3}{K_2^{3/2}} = \frac{\sqrt{n(n-1)}}{n-2} \cdot g_1 \tag{23}$$

$$G_2 = \frac{K_4}{K_2^2} = \frac{(n-1)}{(n-2)(n-3)} \cdot \{(n+1)g_2 + 6\} \tag{24}$$

where $g_1 = b_1^{1/2}$ and $g_2 = b_2 - 3$ (see Joanes and Gill, 1998).

The Fisher's estimators of the population skewness and kurtosis are used in SAS, SPSS and SPLUS softwares. It's worthwhile to know that skewness and kurtosis in MINITAB are calculated by

$$b_1(M) = \frac{m_3}{s^3} = (\frac{n-1}{n})^{3/2} \cdot g_1 \tag{25}$$

$$b_2(M) = \frac{m_4}{s^4} - 3 = (\frac{n-1}{n})^2 \cdot g_2 - 3 \tag{26}$$

where $s^2 = \frac{n}{n-1} \cdot m_2$. Notice that, MINITAB kurtosis estimator (b_2) is biased, Joanes and Gill (1998) showed that for normal distribution

$$Bias(g_2) = -\frac{6}{n+1} \quad , \quad Bias(b_2) \simeq -\frac{12}{n+1} \tag{27}$$

But, in general (Lihua and Ejaz Ahmed, 2008)

$$var(b_2(M)) < var(b_2) < var(G_2) \tag{28}$$

Thus, MINITAB kurtosis estimator is biased but It's varicance is less than Fisher's (G_2) and general (b_2) kurtosis estimators.

Meanwhile, skewness and kurtosis in EXCEL are calculated by

$$b_1 = \frac{n^2}{(n-1)(n-2)} \cdot \frac{m_3}{s^3} \tag{29}$$

$$b_2 = \frac{n^2(n+1)}{(n-1)(n-2)(n-3)} \cdot \frac{m_4}{s^4} - \frac{3(n-1)^2}{(n-2)(n-3)} \tag{30}$$

8 FJB test

In this section we introduce a new form of JB test which is based upon fisher's skewness and kurtosis. As we saw in previous paragraph fisher's measure are unbias so we would have more accurate statistic based on them. So that, The proposed FJB test by using (23) and (24) is:

$$FJB = n[\frac{G_1^2}{6} + \frac{(G_2)^2}{24}] \tag{31}$$

As we know, the sampling distribution of skewness and kurtosis is known for large sample sizes (n>50). So that, under normality this FJB statistic approaches to the chi-square distribution with two degrees of freedom as the sample size grows.

9 Robust Jarque Bera

The classical skewness and kurtosis coefficient have some common disadvantages. They both have a zero breakdown value, and so they are very sensitive to outlying values. One single outlier can make the estimate become very large or small, making it hard to interpret. Another disadvantage is that they are only defined on distributions having finite moments.

The Robust Jarque-Bera (RJB) is the robust version of the Jarque-Bera (JB) test of normality. In particular, RJB utilizes the robust standard deviation (namely the Average Absolute Deviation from the Median (MAAD))

to estimate sample kurtosis and skewness. For more details see Gel and Gastwirth (2006). The RJB statistic is

$$RJB = n[\frac{(\sqrt{b_1})^2}{6} + \frac{(b_2 - 3)^2}{64}] \tag{32}$$

with

$$b_1^{1/2} = \frac{m_3}{j^3} \quad , \quad b_2 = \frac{m_4}{j^4} \tag{33}$$

where the unbiased robust estimate of standard deviation (j) is

$$j = \sqrt{\frac{\pi}{2}} \cdot \frac{1}{n} \sum_{i=1}^{n} |x_i - median(x)| \tag{34}$$

RJB isn't asymptotically chi-squared distributed, so for making decision about normality empirical critical values provided by Monte Carlo simulation are used.

10 Quantile Skewness and Kurtosis

As we know, Disadvantage of Skewness and kurtosis coefficients is sensitivity to outlier values. But, Robustness is the advantage of Quantile measures of skewness and kurtosis and they are not very sensitive to outlier values. We discuss about Quantile measures in below.

An alternative measure of skewness has been proposed by Professor Bowley. Bowley's measure is based on quartiles. In a symmetrical distribution first and third quartiles is the same distance even the median as the first quartile is below it, i.e. $Q_3 - median = median - Q_1$. If a distribution is positively skewed, the top 25 percent of the values will tend to be further from median than the bottom 25 percent, i.e. Q_3 will be further from median than Q_1 is from median (Q_2) and the reverse for negative skewness. The Bowley's measure of Skewness is defined by

$$SK_B = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)} = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} \tag{35}$$

where, $Q_1 = F^{-1}(0.25), Q_2 = F^{-1}(0.50), Q_3 = F^{-1}(0.75)$. The denominator is in fact twice the quartile deviation, so that the degree of skewness

is measured relative to the dispersion of the distribution. The measure is called the quartile measure of skewness, and the value of the coefficient obtained varies between ±1. Symmetry is indicated by zero in both Karl Pearson's and Bowley's method. They have in common only the general form that they are both derived from the difference between two measures of central tendency expressed as a ratio to measures of variation.

Bowley's measure disscused above neglects the two extreme quarters of the data. It would be better for measure to cover the entire data especially because in measuring skewness, we are often interested in the more extreme items. Bowley's measure can be extended by taking any two deciles equidistant from the median or any two percentiles equidistant from the median. Kelly has suggested the following formula for measuring skewness based upon the 10th and the 90th percentiles (or the first and ninth deciles):

$$SK_k = \frac{(P_{90} - Q_2) - (Q_2 - P_{10})}{(P_{90} - Q_2) + (Q_2 - P_{10})} = \frac{P_{10} + P_{90} - 2Q_2}{P_{90} - P_{10}} \tag{36}$$

The Bowley coefficient of skewness has been generalized by Hinkley (1975)

$$SK_3(\alpha) = \frac{F^{-1}(1-\alpha) + F^{-1}(\alpha) - 2F^{-1}(0.5)}{F^{-1}(1-\alpha) - F^{-1}(\alpha)} \tag{37}$$

for any α between 0 and 0.5. Note that the Bowley and Octile coefficients of skewness are a special case of Hinkley's coefficient when $\alpha = 0.25$ and $\alpha = 0.125$.

Moors (1988) showed that the conventional measure of kurtosis (g_2) can be interpreted as a measure of the dispersion of a distribution around the two values $\mu \pm \sigma$. Hence, g_2 can be large when probability mass is concentrated either near the mean m or in the tails of the distributions. Based on this interpretation, Moors (1988) proposed a robust alternative to g_2

$$w_2 = \frac{F^{-1}(7/8) - F^{-1}(5/8) + F^{-1}(3/8) - F^{-1}(1/8)}{F^{-1}(6/8) - F^{-1}(2/8)} \tag{38}$$

Moors justified this estimator on the ground that the two terms, $F^{-1}(7/8) - F^{-1}(5/8)$ and $F^{-1}(3/8) - F^{-1}(1/8)$, are large (small) if relatively little (much) probability mass is concentrated in the neighbourhood of $F^{-1}(6/8)$ and $F^{-1}(2/8)$, corresponding to large (small) dispersion around $\mu \pm \sigma$. The

denominator is a scaling factor ensuring that the statistic is invariant under linear transformation. It is easy to calculate that $F^{-1}(1/8) = -F^{-1}(7/8) = -1.15$, $F^{-1}(2/8) = -F^{-1}(6/8) = -0.68$, $F^{-1}(3/8) = -F^{-1}(5/8) = -0.32$ and $F^{-1}(4/8) = 0$ for $N(0,1)$ and therefore the Moors coefficient of kurtosis is 1.23.

Another Quantile measure of Kurtosis is defined by the ratio of Quartile Range and Percentile Range

$$K_q = \frac{\frac{Q_3 - Q_1}{2}}{P_{90} - P_{10}} = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} \tag{39}$$

For normal observation K_q is equal to 0.263, (see Gupta, 1985).

So that, another robust test based upon Quantiles proposed in Moors et al. (1996). This test (Moors) fits in the framework of the above generalization by using

$$w_1 = \frac{F^{-1}(0.75) + F^{-1}(0.25) - 2F^{-1}(0.5)}{F^{-1}(0.75) - F^{-1}(0.25)} \tag{40}$$

as a robust measure of skewness and

$$w_2 = \frac{F^{-1}(0.875) - F^{-1}(0.625) + F^{-1}(0.375) - F^{-1}(0.125)}{F^{-1}(0.75) - F^{-1}(0.25)} \tag{41}$$

as a robust measure of kurtosis. By using only quantiles of the data this test is resistant to 12.5 percent outliers in the data. As far as, $w_1 \sim N(0, \frac{2}{n})$ and $w_2 \sim N(1.23, \frac{3}{n})$ for large n, the Moors statistic is defined as below:

$$W = n[\frac{(w_1)^2}{2} + \frac{(w_2 - 1.23)^2}{3}] \tag{42}$$

The Moors statistic is asymptotically chi-square distributed with two degrees of freedom.

11 Two new robust JB test

Now, we can define our new proposed tests. They are belong to robust family tests because they are calculated by robust measures of skewness and kurtosis.

In this section we introduce two new robust version of JB test by using quantile skewness and kurtosis measures. As we know, Quantile skewness and kurtosis for large n (n > 50) are normally distributed, we used a Monte Carlo simulation with one million replication and we concluded that:

$$SK_B \sim N(0, \frac{2}{n}) \quad , \quad SK_k \sim N(0, \frac{4}{n}) \quad and \quad K_q \sim N(0.263, \frac{0.08}{n}) \tag{43}$$

So that, we can define the first one (QJB_1) by using (29), (31) and the second one (QJB_2) by using (30) and (31) as below:

$$QJB_1 = n[\frac{(SK_B)^2}{2} + \frac{(K_q - 0.263)^2}{0.08}] \tag{44}$$

$$QJB_2 = n[\frac{(SK_k)^2}{4} + \frac{(K_q - 0.263)^2}{0.08}] \tag{45}$$

It's trivial that two introduced statistic are asymptotically chi-square distributed with two degrees of freedom and normality has to be rejected at level α if QJB_1 and $QJB_2 \geq \chi^2_{1-\alpha,2}$. Meanwhile, the QJB_1 and QJB_2 statistic are resistant to 25 and 10 percent outliers in the data.

12 Monte Carlo Simulations and Results

This section discusses the Monte Carlo simulations applying for comparing the power of considered normality tests with 50, 100 and 200 simulated samples for calculating probability values of considered normality tests, the simulated samples are from 8 kinds of distributions. The results are tabulated in below table. Indeed, the values in the table are pvalues of considered normality tests for assessing normality in simulated data.

Dist	n	JB	MJB	G	FJB	W	QJB_1	QJB_2	RJB
normal(0,1)	50	0.650	0.529	0.471	0.326	0.378	0.437	0.346	0.342
normal(0,1)	100	0.417	0.612	0.699	0.593	0.524	0.452	0.436	0.640
normal(0,1)	200	0.769	0.790	0.733	0.820	0.539	0.563	0.756	0.656
lgnormal(0,1)	50	0.002	0.049	0.000	0.000	0.016	0.063	0.049	0.000
lgnormal(0,1)	100	0.000	0.000	0.000	0.000	0.028	0.002	0.000	0.000
lgnormal(0,1)	200	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
gamma(2,3)	50	0.000	0.015	0.020	0.021	0.075	0.057	0.029	0.030
gamma(2,3)	100	0.000	0.000	0.000	0.005	0.011	0.055	0.024	0.000
gamma(2,3)	200	0.000	0.000	0.000	0.000	0.018	0.009	0.001	0.000
cauchy(2,6)	50	0.000	0.000	0.000	0.000	0.009	0.018	0.002	0.000
cauchy(2,6)	100	0.000	0.000	0.000	0.000	0.000	0.013	0.000	0.000
cauchy(2,6)	200	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
uniform(2,4)	50	0.091	0.090	0.013	0.087	0.105	0.063	0.047	0.140
uniform(2,4)	100	0.014	0.011	0.004	0.012	0.095	0.034	0.036	0.065
uniform(2,4)	200	0.002	0.001	0.000	0.000	0.021	0.007	0.000	0.007
beta(1,3)	50	0.042	0.033	0.031	0.039	0.033	0.060	0.017	0.044
beta(1,3)	100	0.001	0.003	0.006	0.004	0.077	0.008	0.011	0.029
beta(1,3)	200	0.000	0.000	0.000	0.000	0.004	0.002	0.000	0.001
binomial(0.6)	50	0.251	0.231	0.081	0.293	0.161	0.038	0.013	0.161
binomial(0.6)	100	0.288	0.120	0.233	0.249	0.298	0.079	0.039	0.049
binomial(0.6)	200	0.423	0.288	0.301	0.251	0.088	0.042	0.026	0.112
geometric(0.3)	50	0.000	0.000	0.001	0.000	0.087	0.041	0.025	0.002
geometric(0.3)	100	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001
geometric(0.3)	200	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

13 Conclusion

In this paper we discussed several goodness-of-fit tests. The commonly used Jarque-Bera test of normality was extended to become a goodness-of-fit test. We noted JB test often fails to lead to a correct actual size, due to the slow rate of convergence towards the limiting distribution. Moreover, the test cannot be performed at distributions without finite moments, and as it is based on moments of the data it is strongly influenced by the presence of outlying values. Therefore Moors et al. (1996) proposed to replace the classical skewness and kurtosis coefficient by robust alternatives, leading to the MOORS test. We conducted a similar approach by using other quantile measures. The three new proposed tests are FJB, QJB_1, QJB_2 and the test statistics are asymptotically χ^2 distributed. In order to examine the size and power of the proposed testing procedures, we have conducted Monte Carlo simulation and compared the results. We have found that the sizes of our testing procedure are close to the corresponding nominal sizes, and that the power of our testing procedures are better than,

or in some cases as good as, the other competing tests, especially in presence of outlying values.

References

[1] Bowman, K. O. and L. R. Shenton, 1975, Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2, *Biometrika*, 62, 243-250.

[2] Bera, A. K. and C. M. Jarque, 1982, Model specification tests: A simultaneous approach, *Journal of Econometrics*, 20, 59-82.

[3] Bowley, A. L., 1920, Elements of Statistics, *New York, Charles Scribner's and Sons.*

[4] Cho, D. W. and K. S. Im, 2002, A Test of Normality Using Geary's Skewness and Kurtosis Statistics. Department of Economics, College of Business Administration, University of Central Florida, Orlando, pp: 1-11.

[5] Croux, C. and Rousseeuw, P. J., 1992, Time-Efficient Algorithms for Two Highly Robust Estimators of Scale, *Computational Statistics*, 1, 411 - 428.

[6] Durfour, J., Farhat, A., Gardial, L., and Khalaf, L., 1998, Simulation-based Finite Sample Normality Tests in Linear Regressions, *Econometrics Journal*, 1, C154-C173.

[7] Deb, P. and M. Sefton, 1996, The distribution of Lagrangian multiplier test of normality, *Economics Letters*, 51, 123-130.

[8] D'Agostino, R. B., 1971, An omnibus test of normality for moderate and large sample sizes. *Biometrika*, 58, 341-348.

[9] D'Agostino, R. and B. Roseman, 1974, The power of Geary's test of normality, 61, *Biometrika*, 181-184.

[10] D'Agostino, R. and M. Stephens, 1986, Goodness-of-Fit-Techniques, *New York: Marcel Dekker.*

[11] Dumonceaux, R., C. Antle, and G. Haas, 1973, Likelihood ratio test for discrimination between two models with unknown location and scale parameter, *Technometrics*, 15, 19-27.

[12] Epps, T. W. and L. B. Pulley, 1983, A test for normality based on the empirical characteristic function, *Biometrika*, 70, 723726.

[13] Epps, T. W., 1999, Limiting behavior of the integrated characteristic function test for normality under Gram-Charlier alternatives, *Statistics and Probability Letters*, 42, 175184.

[14] Fisher, RA., 1973, Statistical methods for research workers, *14th ed. New York: Hafner Publishing Co.*

[15] Gastwirth, J. L. and M.G.B. Owens, 1977, On classical test of normality, *Biometrika*, 64, 135-139.

[16] Geary, R. C., 1947, Testing for normality, *Biometrika*, 34, 209-242.

[17] Geary, R. C., 1935, The ratio of the mean deviation to standard deviation as a test of normality, *Biometrika*, 27, 310-332.

[18] Geary, R. C., 1936, Moments of the ratio of mean deviation to the standard deviation as a test of normality, *Biometrika*, 28, 295-305.

[19] Gupta, S. P., 1985, Statistical Method, *Sultan Chand and Sons, New Delhi.*

[20] Hinkley DV., 1975, On Power Transformations to Symmetry, *Biometrika*, 62, 101-111.

[21] Henze, N. 1997, A new approach to the BEHP tests for multivariate normality, *Journal of Multivariate Analysis* 62, 123.

[22] Jarque, C. M. and A. K. Bera, 1987, A test for normality of observations and regression residuals, *International Statistical Review*, 55, 163-172.

[23] Jarque, C. M. and Bera, A. K., 1980. Eficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economic Letters* 6, 255-259.

[24] Jarque, C. M., and Bera, A. K., 1981, Eficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals: Monte Carlo evidence, *Economics Letters*, 7, 313-318.

[25] Joanes, D. N. and Gill, C. A., 1998. Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society (Series D)*, The Statistician 47(1), 183-189.

[26] Lihua, A., Ejaz Ahmed, S., 2008. Improving the performance of kurtosis estimator. *Computational Statistics and Data Analysis*, 52, 26692681.

[27] Mardia, K. V., 1980, Test of univariate and multivariate normality, *Handbook of Statistics*, P. R. Krishnaiah, edition, 1, 279-320.

[28] Moors, J.J.A., Wagemakers, R.T.A., Coenen, V.M.J., Heuts, R.M.J., and Janssens, M.J.B.T., 1996. Characterizing systems of distributions by quantile measures. *Statistica Neerlandica*, 50, 417430.

[29] Moors, J. J. A., 1988, A Quantile Alternative for Kurtosis, *The Statistician*, 37, 25-32.

[30] Pearson, E.S., D'Agostino, R.B., and Bowman, K.O., 1977, Test for Departure from Normality: Comparison of Powers, *Biometrika*, 64, 2, 231-246.

[31] Poitras, G., 1992, Testing Regression Disturbances for Normality with Stable Alternatives: Further Monte Carlo Evidence, *Journal of Statistical Computation and Simulation*, 41, 109-123.

[32] Shapiro, S. S. and M. B. Wilk, 1965, An analysis of variance test for normality (complete samples), *Biometrika*, 52, 591-611.

[33] Shapiro, S. S. and R. S. Francia, 1972, An approximation analysis of variance test for normality, *Journal of the American Statistical Association*, 67, 215- 216.

[34] Shenton, L. R. and K. O. Bowman, 1977, A bivariate model for the distribution of $\sqrt{b_1}$ and b_2, *Journal of the American Statistical Association*, 72, 206-211.

[35] Spiegelhalter, D. J., 1980, An Omnibus Test for Normality for Small Samples, *Biometrika*, 67, 493-496.

[36] Thode, H., Jr., 2002, Testing Normality, *New York: Marcel Dekker.*

[37] Uthoff, V. A. 1973, The most powerful scale and location invariant test of the normal versus double exponential, *Annals of Statistics*, 1, 170-174.

[38] Urzu'a, C., 1996, On the correct use of omnibus tests for normality, *Economics Letters*, 53, pp. 247251.

Approximate Estimators for the Scaled Burr XII Distribution Under Progressive Type-II Censoring

A. Asgharzadeh, and R. Valiollahi

Department of Statistics, University of Mazandaran

Abstract: For the Burr XII distribution, the maximum likelihood method does not provide an explicit estimator for the scale parameter based on a progressively Type-II censored sample. In this paper, we first present a simple method of deriving an explicit estimator by approximating the likelihood function. We then examine through simulations the estimated risk (ER) of this estimator and show that this estimator is as efficient as the maximum likelihood estimator (MLE). An approximation based on a noninformative prior for the scale parameter are used for obtaining the Bayes estimates. Finally, a Monte Carlo simulation are carried out to compare the performance of the estimates.

Keywords: maximum likelihood estimation, Bayes estimation, Fisher information, Burr XII distribution, Monte Carlo simulation, Progressive Type-II censoring.

1 Introduction

The Burr XII distribution was first introduced in the literature by Burr (1942). This distribution is used in areas of quality control, reliability studies, duration and failure time modeling. The probability density function (pdf) and cumulative distribution function (cdf) of the three-parameter Burr XII (which we shall simply denote by Burr(c, k, θ)) are given, respectively, by

$$f(x,\theta) = \frac{ck}{\theta}(\frac{x}{\theta})^{c-1}(1+(\frac{x}{\theta})^{c})^{-(k+1)}, \qquad x>0, \quad c>0, \quad k>0, \quad \theta>0 \tag{1}$$

$$F(x,\theta) = 1-(1+(\frac{x}{\theta})^{c})^{-k}, \qquad x>0 \tag{2}$$

where c and k are shape parameters and θ is scale parameter.

Inferences for Burr XII model were discussed by many authors such as Popadopoulos (1978), Evans and Ragab (1983), Zimmer et al. (1998), Al-Hussaini and Jaheen (1992, 1994), Al-Hussaini (1992), Ali Mousa (1995) and Soliman (2005).

In this study, we consider a censoring scheme called progressive Type-II censoring. Suppose that n independent items, with continuous density $f(x)$, are put on a life test. Suppose, further, that a censoring scheme $(R_1, R_2, ..., R_m)$ is previously fixed such that immediately following the first failure, X_1, R_1 surviving items are removed from the experiment at random, and immediately following the second failure, X_2, R_2 surviving items are removed from the experiment at random. This process continues until at the time of the m^{th} observed failure X_m, the remaining R_m items are removed from the test. The m ordered observed failure times denoted by

$$X_{1:m:n}^{(R_1,R_2,...,R_m)}, X_{2:m:n}^{(R_1,R_2,...,R_m)}, ..., X_{m:m:n}^{(R_1,R_2,...,R_m)},$$

are called progressive Type II right censored order statistics of size m from a sample of size n with progressive censoring scheme $(R_1, R_2, ..., R_m)$. If the failure times of the n items originally on the test are from a continuous population with cdf $F(x)$ and pdf $f(x)$, the joint probability density function

$$X_1 = X_{1:m:n}^{(R_1,R_2,...,R_m)}, X_2 = X_{2:m:n}^{(R_1,R_2,...,R_m)}, ..., X_m = X_{m:m:n}^{(R_1,R_2,...,R_m)}$$

is given (Balakrishnan and Aggarwala (2000)) by

$$f_{X_1,X_2,...,X_m}(x_1, x_2, ..., x_m, \theta) = A \prod_{i=1}^{m} f(x_i, \theta)[1 - F(x_i, \theta)]^{R_i}, \qquad (3)$$

where

$$A = n(n-R_1-1)(n-R_1-R_2-2)...(n-R_1-R_2-...-R_{m-1}-m+1). \quad (4)$$

In this paper, the maximum likelihood (ML) and approximate maximum likelihood (AML) estimates for the scale parameter θ are obtained based on progressive Type-II censored samples. The approximate Bayes estimates of θ are obtained under the squared loss and LINEX loss functions.

These estimates are compared via a Mont Carlo simulation study. We also derive an approximate confidence interval based on asymptotic normality for parameter θ.

2 Point Estimation

In this section, the MLE and AMLE of the scale parameter θ are obtained based on progressively censored samples. The Bayes estimates for θ do not have explicit forms, so we use an approximation based on Lindley's method for finding the Bayes estimates.

2.1 Maximum Likelihood Estimation

Suppose that

$$\mathbf{X} = (X_{1:m:n}^{(R_1,R_2,...,R_m)}, X_{2:m:n}^{(R_1,R_2,...,R_m)}, ..., X_{m:m:n}^{(R_1,R_2,...,R_m)})$$

is a progressively Type-II censored sample from a life test on items whose life times have Burr (c, k, θ) distribution with density (1) and censoring scheme $(R_1, R_2, ..., R_m)$. The likelihood function based on this sample is given by

$$L(\theta; \underline{x}) = A\theta^{-mc} \prod_{i=1}^{m} ckx_i^{c-1}(1 + (\frac{x_i}{\theta})^c)^{-(k(R_i+1)+1)}, \tag{5}$$

where

$$\underline{x} = (x_1, x_2, ..., x_m), \quad x_i = x_{i:m:n}^{(R_1,R_2,...,R_m)}, \ i = 1, 2, ..., m.$$

The log likelihood function is given by

$$\ln L(\theta; \underline{x}) \propto -mc \ln(\theta) - T, \tag{6}$$

where

$$T = \sum_{i=1}^{m} \left[(k(R_i + 1) + 1) \ln(1 + (\frac{x_i}{\theta})^c) \right]. \tag{7}$$

The ML estimate $\hat{\theta}$ of the parameter θ is the solution of

$$-m + \sum_{i=1}^{m}(k(R_i+1)+1)\frac{x_i^c}{\theta^c + x_i^c} = 0. \tag{8}$$

Some numerical methods such as the Newton-Raphson method are employed to solve (8).

2.2 Approximate Maximum Likelihood Estimation

The likelihood equation in (8) does not an admit explicit solution for θ. Consider the random variable $Y = (\frac{X}{\theta})^c$. Then Y has a pareto distribution with pdf

$$g(y) = k(1+y)^{-(k+1)}, \tag{9}$$

and cdf

$$G(y) = 1 - (1+y)^{-k} \tag{10}$$

By using (8)and (10), the likelihood equation can be written as have

$$\frac{d}{d\theta}\ln\ L(\theta;\underline{x}) = nk - \sum_{i=1}^{m}\left[k(R_i+1)+1\right]\left[1-G(y_i)\right]^{\frac{1}{k}}. \tag{11}$$

We approximate the term $h(y_i) = [1-G(y_i)]^{\frac{1}{k}}$ by expanding it in a Taylor series around $E(Y_{i:m:n}) = \nu_{i:m:n}$. From Balakrishnan and Sandhu (1995), it is known that

$$G(Y_{i:m:n}) \overset{d}{=} U_{i:m:n},$$

where $U_{i:m:n}$ is the i-th progressively Type-II censored order statistic from the uniform $U(0,1)$ distribution. We then have

$$Y_{i:m:n} \overset{d}{=} G^{-1}(U_{i:m:n}),$$

$$\nu_{i:m:n} = E(Y_{i:m:n}) \approx G^{-1}(\alpha_{i:m:n}),$$

where $\alpha_{i:m:n} = E(U_{i:m:n})$. From Balakrishnan and Aggarwala (2000), it is known that

$$\alpha_{i:m:n} = 1 - \prod_{j=m-i+1}^{m}\frac{j+R_{m-j+1}+\cdots+R_m}{j+1+R_{m-j+1}+\cdots+R_m}, \quad i = 1,\cdots,m.$$

Since, for the pareto distribution, we have

$$G^{-1}(u) = (1-u)^{-1/k} - 1,$$

we can approximate $\nu_{i:m:n}$ by $(1-\alpha_{i:m:n})^{-1/k} - 1$. Now, upon expanding the function $h(y_i)$ around the point $\nu_{i:m:n}$ and keeping only the first two terms, we get

$$\begin{aligned} h(y_i) &\approx h(\nu_{i:m:n}) + (y_i - \nu_{i:m:n})h'(\nu_{i:m:n}) \\ &= \gamma_i + \delta_i y_i, \end{aligned} \tag{12}$$

where

$$\gamma_i = h(\nu_{i:m:n}) - \nu_{i:m:n}h'(\nu_{i:m:n}) = \frac{2\nu_{i:m:n}+1}{(1+\nu_{i:m:n})^2}, \quad i = 1, \cdots m \tag{13}$$

and

$$\delta_i = h'(\nu_{i:m:n}) = -\frac{1}{(1+\nu_{i:m:n})^2}, \quad i = 1, \cdots m. \tag{14}$$

Using the expression in (12), we approximate the likelihood equation in (11) by

$$\frac{d\ln L(\theta;\underline{x})}{d\theta} \approx nk - \sum_{i=1}^{m} [k(R_i+1)+1]\,(\gamma_i + \delta_i y_i) = 0,$$

which can be rewritten as

$$nk - \sum_{i=1}^{m} [k(R_i+1)+1]\,\gamma_i - \sum_{i=1}^{m} [k(R_i+1)+1]\,\delta_i (\frac{x_i}{\theta})^c = 0. \tag{15}$$

By solving the equation in (15) for θ, we obtain the approximate MLE of θ as

$$\hat{\theta}_{AML} = \left\{ \frac{\sum_{i=1}^{m} [k(R_i+1)+1]\,\delta_i x_i^c}{nk - \sum_{i=1}^{m}(k(R_i+1)+1)\gamma_i} \right\}^{\frac{1}{c}}. \tag{16}$$

The approximate MLE in (16) may provide us with a good starting value for the iterative solution of the likelihood equation in (8).

2.3 Bayes Estimation

Now, we consider Bayes estimate of θ under the squared error (SE) and LINEX loss functions. The LINEX loss function for $u = u(\theta)$ can be expressed as

$$L(\Delta) \propto e^{a\Delta} - a\Delta - 1; \quad a \neq 0$$

where $\Delta = (\tilde{u} - u)$, $\tilde{u}$ is an estimate of u. The sign, and magnitude of 'a' represent the direction and degree of symmetry, respectively. ($a > 0$ means overestimation is more serious than underestimation, and $a < 0$ means the opposite). For 'a' closed to zero, the LINEX loss function is approximately the squared error loss.

The Bayes estimator ($\hat{u}_{BL}$) of u under the LINEX loss function is

$$\hat{u}_{BL} = -\frac{1}{a}\ln(E(e^{-au}|x)).$$

We consider the noninformative prior

$$\pi(\theta) \propto \frac{1}{\theta}. \tag{17}$$

Combining the likelihood function (5) and prior (17) the posterior density of θ is

$$\pi(\theta|\underline{x}) \propto \theta^{-(mc+1)}e^{-T},$$

where T is given by (7).

The Bayes estimator of a function U of the unknown parameter θ, under squared error loss is the posterior mean

$$\hat{U}_{BS} = E(U|\underline{x}) = \frac{\int U(\theta)\pi(\theta|\underline{x})d\theta}{\int \pi(\theta|\underline{x})d\theta}. \tag{18}$$

The ratio of integral in equation (18) does not seem to take a closed form and so we use the following method of approximation.

Lindley's procedure: Lindley's method has been used by some authors to obtain Bayes estimators of the parameters of some distributions. See for example, Sinha (1985) and Soliman (2001). Lindley (1980) developed approximate procedure for evaluation the posterior expectation of $U(\theta)$ as

$$E(U(\theta)|\underline{x}) = \frac{\int U(\theta)e^{l(\theta)+\rho(\theta)}d\theta}{\int e^{l(\theta)+\rho(\theta)}d\theta},$$

which is the Bayes estimate of $U(\theta)$ under squared error loss function, where $\rho(\theta) = \ln(p(\theta))$, $p(\theta)$ is arbitrary function of θ, and $l(\theta)$ is the logarithm of the likelihood function. For more details see Lindley (1980).

In the one parameter case, Lindley's approximation form reduces to the following form:

$$E(U(\theta)|\underline{x}) = U(\theta) + \frac{1}{2}[U_2 + 2U_1\rho_1]\sigma^2 + \frac{1}{2}l_3U_1(\sigma^2)^2, \tag{19}$$

where $l_i = \frac{d^i}{d\theta^i}l(\theta), for\ i = 1,2,3,\quad U_j = \frac{d^j}{d\theta^j}U(\theta),\ for\ j = 1,2,\quad \rho_1 = \frac{d}{d\theta}\rho(\theta)$.
To apply Lindley's form (19), we first obtain

$$l_1 = \frac{c}{\theta}(s_1 - m),\quad l_2 = -\frac{c}{\theta^2}(cs_2 + s_1 - m),\quad l_3 = \frac{2c}{\theta^3}(c^2s_3 + cs^2 + s_1 - m),$$

where

$$s_1 = \sum_{i=1}^{m}(k(R_i+1)+1)\frac{x_i^c}{\theta^c + x_i^c}$$
$$s_2 = \sum_{i=1}^{m}(k(R_i+1)+1)\frac{(\theta x_i)^c}{(\theta^c + x_i^c)^2}$$
$$s_3 = \sum_{i=1}^{m}(k(R_i+1)+1)\frac{(\theta^2 x_i)^c}{(\theta^c + x_i^c)^3}.$$

From prior density(17), we observe that $\rho_1 = \frac{d}{d\theta}\rho(\theta) = -\frac{1}{\theta}$. Substitution of the above value in (19), we obtain the Bayes estimate of $U(\theta)$ using Lindley's method as

$$\hat{U}_B = E(U(\theta)|\underline{x}) = U(\theta) + \Psi_1U_1 + \Psi_2U_2, \tag{20}$$

where

$$\Psi_1 = \frac{c\theta s_3}{(cs_2 + s_1 - m)^2}$$
$$\Psi_2 = \frac{\theta^2}{2c(cs_2 + s_1 - m)}.$$

All functions of the right-hand side of (20) are to be evaluated at $\hat{\theta}_{ML}$. Substituting $U(\theta) = \theta$ in (20) we obtain the Bayes estimate of θ under squared error loss function given by

$$\hat{\theta}_{BS} = \theta + \Psi_1. \tag{21}$$

Substituting $U(\theta) = e^{-a\theta}$ in (20) we obtain the Bayes estimate of θ under LINEX loss function given by

$$\hat{\theta}_{BL} = \theta - \frac{1}{a} ln(a^2\Psi_2 - a\Psi_1 + 1). \tag{22}$$

3 Approximate Confidence Interval

In this section, we derive the observed and expected Fisher information based on the likelihood as well as the approximate likelihood functions. These enable the construction of confidence intervals based on pivotal quantities using the limiting normal distribution. From the log-likelihood function, we obtain

$$\begin{aligned} \frac{d^2 \ln L(\theta)}{d\theta^2} &= \frac{mc}{\theta^2} - \frac{c}{\theta^2} \sum_{i=1}^{m} [k(R_i + 1) + 1] \frac{x_i^c}{\theta^c + x_i^c} \\ &- \frac{c^2}{\theta^2} \sum_{i=1}^{m} [k(R_i + 1) + 1] \frac{(\theta x_i)^c}{(\theta^c + x_i^c)^2}. \end{aligned}$$

From this expression, we obtain the expected Fisher information as

$$\begin{aligned} E\left[-\frac{d^2 \ln L(\theta)}{d\theta^2}\right] &= -\frac{mc}{\theta^2} + \frac{c}{\theta^2} \sum_{i=1}^{m} [k(R_i + 1) + 1] E(\frac{x_i^c}{\theta^c + x_i^c}) \\ &+ \frac{c^2}{\theta^2} \sum_{i=1}^{m} [k(R_i + 1) + 1] E\left[\frac{(\theta x_i)^c}{(\theta^c + x_i^c)^2}\right]. \end{aligned}$$

Unfortunately, the exact mathematical expressions for the above expectations are very difficult to obtain. Therefore, we use the following approximation which is obtained by dropping the expectation operator E

$$E(\frac{x_i^c}{\theta^c + x_i^c}) \approx \frac{x_i^c}{\theta^c + x_i^c}.$$

$$E(\frac{(\theta x_i)^c}{(\theta^c + x_i^c)^2}) \approx \frac{(\theta x_i)^c}{(\theta^c + x_i^c)^2}.$$

Therefor

$$\begin{aligned} E\left[-\frac{d^2 \ln L(\theta)}{d\theta^2}\right] &\approx -\frac{mc}{\theta^2} + \frac{c}{\theta^2}\sum_{i=1}^{m}(k(R_i+1)+1)\frac{x_i^c}{\theta^c + x_i^c} \\ &+ \frac{c^2}{\theta^2}\sum_{i=1}^{m}(k(R_i+1)+1)\frac{(\theta x_i)^c}{(\theta^c + x_i^c)^2}. \end{aligned} \tag{23}$$

Similarly, from approximate likelihood equation, we obtain

$$\begin{aligned} E\left[-\frac{d^2 \ln L^*(\theta)}{d\theta^2}\right] &\approx -\frac{mc}{\theta^2} + \frac{c}{\theta^2}\sum_{i=1}^{m}(k(R_i+1)+1)\gamma_i(1+c\gamma_i)(\frac{x_i}{\theta})^c \\ &+ \frac{c}{\theta^2}\sum_{i=1}^{m}(k(R_i+1)+1)\delta_i(1+2c\gamma_i)(\frac{x_i}{\theta})^{2c} \\ &+ \frac{c^2}{\theta^2}\sum_{i=1}^{m}(k(R_i+1)+1)\delta_i^2(\frac{x_i}{\theta})^{3c}. \end{aligned} \tag{24}$$

To construct confidence intervals or to conduct tests of hypotheses for the scale parameter θ, one must construct a pivotal quantity. Since $\hat{\theta}$ and $\tilde{\theta}$ are asymptotically normally distributed with the corresponding variances

$$var(\widehat{\theta}) = \frac{1}{-E[\frac{d^2}{d\theta^2}\ln L]} \qquad \text{and} \qquad var(\widetilde{\theta}) = \frac{1}{-E[\frac{d^2}{d\theta^2}\ln L^*]},$$

we have the asymptotic distribution of

$$P_1 = \frac{\hat{\theta} - \theta}{\sqrt{var(\hat{\theta})}} \qquad \text{and} \qquad P_2 = \frac{\tilde{\theta} - \theta}{\sqrt{var(\tilde{\theta})}} \tag{27}$$

to be standard normal. Therefore,

$$(\hat{\theta} - z_{\alpha/2}\sqrt{var(\hat{\theta})},\ \hat{\theta} + z_{\alpha/2}\sqrt{var(\hat{\theta})}),$$

and

$$(\tilde{\theta} - z_{\alpha/2}\sqrt{var(\tilde{\theta})},\ \tilde{\theta} + z_{\alpha/2}\sqrt{var(\tilde{\theta})})$$

form $100(1-\alpha)\%$ confidence intervals for θ based on the pivotal quantities P_1 and P_2, respectively. Here $z_{\alpha/2}$ is a standard s-normal variate.

4 Monte Carlo Simulation

A simulation study was carried out to compare the performance of the approximate MLE, MLE and Bayes estimates. Using the algorithm presented in Balakrishnan and Sanhu (1995), we generated random progressively Type-II censored samples of various sizes and censoring schemes from BURR XII with ($\theta = 1$). We then computed the approximate MLE and the MLE of the parameter θ from (8) and (16) respectively. Bayes estimations was then obtained by using (21) and (22). Table 1 provides the average values of the estimates and their estimated risk (ER). All the averages were computed over 5000 simulations. From Table 1, we show that for both small and large samples the Bayes estimates relative to LINEX loss function have the smallest estimated risk (ER) as compared with quadratic Bayes estimates or the MLE's and AMLE's. 2. From Table 1, as the effective sample m/n decreases, the estimated risk (ER) of estimators increase.

References

[1] Al-Hussaini, E. K. and Jaheen, Z. F. (1992). "Bayesian estimation of the parameters, reliability and failure rate function of the Burr failure model." *J. Statist. Comput. Simul*, **41**, 31-40.

[2] Al-Hussaini, E. K. and Jaheen, Z.F. (1994). "Approximate Bayes estimators applied to the Burr model." *Commun. Statist.- Theory. Meth*, **23(1)**, 99-121.

[3] Ali Mousa, M. A. M. (1995). "Empirical Bayes estimators for the Burr Type-XII accerelated life testing model." *Commun. Statist.- Theory. Meth*, **52**, 95-103

[4] Burr, I. W. (1942). "Cumulative frequency functions." *Ann. Math. Statist.*, **13**, 215-232.

[5] Balakrishnan, N. and Aggarwala R. A. (2000). "Progressive censoring: Theory, Methods and Applications." *Boston: Birkhäuser*

[6] Balakrishnan, N. and Sandhu. R. A. (1995). "A simple simulation algorithm for generating progressive Type-II censored samples." *American Statistician*, **49(2)**, 229-230.

[7] Evans, I. G. and Ragab, A. S. (1983). "Bayesian inference givan a Type-II censored sample from Burr distribution." *Commun. Statist.- Theory. Meth*, 2,1569-1580.

[8] papadopolous, A. S. (1978). "The Burr distribution as a failure model from a Bayesian approach.", *IEEE. Trans. Rel.* ,**27**, 367-371.

[9] Lindley, D. V. (1980). "Approximate Bayesian method." *Trabajos de stadistica*, **31**, 223-237.

[10] Sinha, S. K. (1985). "Bayes estimation of reliability function of normal distribution." *IEEE. Trans. Rel.* **34**, 360-362.

[11] Soliman, A. A. (2005). "Estimation of parameters of life from progressively censored data using Burr XII madel." , *IEEE. Trans. Rel.*. **54(1)**, 34-42.

[12] Soliman, A. A. (2001). "LINEX and quadratic approximate Bayes estimators applied to the Pareto model." *Commun. Statist.-Simul.* , **30(1)**, 47-62.

[13] Zimmer, W. J. and Bert Keats. J. and Wang. F. K (1998). "The Burr XII distribution in reliability analysis." *J. Qual. Technol.*, **39**,386-394.

Table 1
Point Estimates and Estimated Risks (ER) for $\theta(\text{when} c = 3, k = 2.5)$

n	m	*Scheme*	$\hat{\theta}_{ML}$	$\tilde{\theta}$	$\hat{\theta}_{BS}$	$\hat{\theta}_{BL}$
30	10	$(10,8^0,10)$	1.0241	0.9903	1.0271	0.9895
			(0.0170)	(0.0154)	(0.0172)	(0.0143)
30	10	$(15,8^0,5)$	1.0280	0.9972	1.0346	0.9933
			(0.0179)	(0.0161)	(0.0185)	(0.0148)
30	10	$(5,8^0,15)$	1.0332	0.9951	0.0320	0.9956
			(0.0170)	(0.0148)	(0.0169)	(0.0138)
50	20	$(19^0,30)$	1.0342	0.9952	1.0127	0.9954
			(0.0086)	(0.0069)	(0.0073)	(0.0067)
50	20	$(30,19^0)$	1.0275	0.9936	1.0305	0.9927
			(0.0168)	(0.0151)	(0.0171)	(0.0140)
50	20	$(15,18^0,15)$	1.0228	0.9974	1.0171	0.9956
			(0.0096)	(0.0086)	(0.0093)	(0.0083)
100	50	$(49^0,50)$	1.0314	0.9996	1.0064	0.9992
			(0.0041)	(0.0029)	(0.0030)	(0.0029)
100	50	$(50,49^0)$	1.0713	0.9991	1.0069	0.9968
			(0.0122)	(0.0040)	(0.0041)	(0.0039)
100	50	$(30,48^0,20)$	1.0249	0.9988	1.0074	0.9981
			(0.0046)	(0.0038)	(0.0039)	(0.0037)

Maximum Inactivity Entropy Models

S. Ashrafi

Department of Statistics, University of Isfahan

Abstract: Let X be a random variable denoting the lifetime of a system for which the inactivity time is the conditional random variable $(t - X|X < t)$. In this paper we introduce the maximum inactivity entropy (MIE) models. The MIE model is the distribution with the density that maximizes the past entropy for all $t > 0$ under some constraints. This model provides an approximation for the lifetime distribution based on information about reversed hazard function which corresponds to the past lifetime. However, several well-known distributions are shown to be MIE under different constraints.

Keywords: Past lifetime distribution, Inactivity time, Reversed hazard function, Partial knowledge.

1 Introduction

In reliability analysis usually one deals with a system for which the lifetime distribution function is unknown but we may have some information about the lifetime of the system through the physical characterizations of the system. Having said that, we wish to derive a model that best approximates the corresponding lifetime distribution based on this information. To produce a model for the lifetime distribution in such a situation, the maximum entropy (ME) paradigm (Jaynes, 1982) can be employed. In this paradigm, partial Knowledge about the lifetime distribution function is formulated in terms of some information constraints.

Let X be a nonnegative random variable representing the lifetime. Suppose that X has the cumulative distribution function (CDF) F, $F(x) = P(X \leq x)$, the survival function $\bar{F} = 1 - F$, and the probability density function (PDF) $f(.)$. The entropy (Shannon, 1948) of a nonnegative continuous random variable X is given by

$$\begin{aligned} H(f) &= -\int_0^\infty f(x) \log f(x) dx \\ &= 1 - \int_0^\infty f(x) \log \lambda_F(x) dx, \end{aligned} \tag{1}$$

where$\lambda_F(x) = f(x)/\bar{F}(x)$ denotes the hazard function of X.

the ME model is the distribution whose PDF $f^*(.)$ maximizes $H(f)$ subject to some information constraints. The ME procedure in terms of moment constraints is well-known. Ebrahimi et al. (1996) studied a developing ME lifetime distribution subject to monotonicity constrains on the failure rate. Recently Ebrahmi et al. (2008) have studied the multivariate ME models.

Frequently, in reliability the problem of intrest is the lifetime beyond an age t. In such caces, the the distribution of intrest for computing uncertainty and information is the residual distribution with density

$$f_t(x) = \begin{cases} \frac{f(x)}{\bar{F}(t)} & x \geq t, \\ 0 & \text{otherwise.} \end{cases}$$

The entropy of the residual lifetime is given by

$$\begin{aligned} H(f;t) &= -\int_t^\infty f_t(x) \log f_t(x) dx \\ &= 1 - \frac{1}{\bar{F}(t)} \int_t^\infty f(x) \log \lambda_F(x) dx, \end{aligned} \tag{2}$$

Ebrahimi (2000) suggested to maximize the residual entropy and derived four lifetime models that maximize residual entropy subject to constraints on the growth and curvature of the hazard function. These models are maximum dynamic entropy (MDE). Asadi et al. (2004) obtained a result on the relationship between the hazard rate ordering and residual entropy ordering for distributions with monotone PDFs. Based on this result they identified classes of distributions in which some well-known distributions are the MDE models.

Using various results on ME and MDE models, we can find the approximation for the lifetime distribution through information about the hazard function. It has to be pointed out that hazard rate deals with the residual lifetime of X. Sometimes, in reliability analysis the problem of interest is the inactivity time of the system, when the system has failed until time t (t is the time that the system is inspected for the first time and it is found to be 'down'). However, we are interested in getting inference on inactivity random variable. The inactivity time of the system, which we denote by

$X_{[t]}$ is $X_{[t]} = [t - X|X < t]$. In such cases, the distribution of interest for computing uncertainty and information is the past distribution with density

$$f_{[t]}(x) = \begin{cases} \frac{f(x)}{F(t)} & 0 < x < t, \\ 0 & \text{otherwise.} \end{cases}$$

The entropy of the inactivity time is given by

$$\begin{aligned} H(f;[t]) &= -\int_0^t f_{[t]}(x) \log f_{[t]}(x) dx \\ &= 1 - \frac{1}{F(t)} \int_0^t f(x) \log r_F(x) dx, \end{aligned} \tag{3}$$

where $r_F(x) = f(x)/F(x)$ denotes the reversed hazard function of X.

The present paper aims to explore a variant of maximum dynamic entropy models which is obtained by maximizing the past entropy (3). We call these models as the maximum inactivity entropy (MIE) models. These models provide approximations for the lifetime distribution based on information about reversed hazard function which corresponds to the inactivity time $X_{[t]}$. The paper is organized as follows. In section 2, we provide a theorem on the relationship between the reversed hazard ordering and past entropy ordering for distributions with monotone PDFs. Using this theorem we compare uncertainty between the past lifetimes of parallel and series systems with the past lifetime of their components. In section 3, we obtain constraints in terms of differential inequalities and identify the set of distributions (maximal set) Ω_F in which some well-known distributions maximize the past entropy. Then we tabulate the MIE distributions based on their PDF and the constraints for the maximal set Ω_F. It is also shown that the mixture of two power distributions is MDI with the given maximal Ω_F.

2 Past entropy ordering

The following theorem has various direct applications and is useful to obtain the MIE models in the subsequent sections.

Theorem 1 *Let X and Y be two nonnegative continuous random variables having distribution functions F and G with PDFs f and g and reversed hazard functions r_F and r_G respectively. Suppose that $r_F(t) \leq r_G(t)$ for all $t > 0$. Then*

(a) *$H(f;[t]) \leq H(g;[t])$ if $g(x)$ is decreasing for all $x \geq 0$.*

(b) *$H(f;[t]) \geq H(g;[t])$ if $f(x)$ is increasing for all $x \geq 0$.*

Proof: Let $X_{[t]}$ and $Y_{[t]}$ denote the inactivity time with densities $f_{[t]}$ and $g_{[t]}$ respectively. since $r_F(t) \leq r_G(t)$, one can easily show that $X_{[t]} \leq_{st} Y_{[t]}$,where $\leq_{st}$ stands for stochastic ordering. Also, since $g_{[t]}(x)$ is decreasing in x, we get

$$E_{f_{[t]}}(\log g_{[t]}(X)) \geq E_{g_{[t]}}(\log g_{[t]}(X)) \tag{4}$$

Now consider the discrimination information between the two past distributions:

$$\begin{aligned} K(f_{[t]}; g_{[t]}) &= \int_0^t f_{[t]}(x) \log \frac{f_{[t]}(x)}{g_{[t]}(x)} dx \\ &= -H(f;[t]) - E_{f_{[t]}}(\log g_{[t]}(X)) \geq 0. \end{aligned} \tag{5}$$

Using this and (1) we have

$$H(f;[t]) \leq H(g;[t])$$

This completes the proof of part (a). To prove part (b), let $f(x)$ be increasing. Using essentially similar arguments as above, we have $E_{f_{[t]}}(\log f_{[t]}(X)) \leq E_{g_{[t]}}(\log f_{[t]}(X))$. This inequality and the fact that $K(g_{[t]}; f_{[t]}) \geq 0$ gives the required result.

The following examples give direct applications of theorem 1.

Example 1 *(System of components.)* Let $X_1, X_2, ..., X_n$ be independent and identically distributed (i.i.d) lifetimes of components of a system having a common PDF f, CDF F, and survival function $\bar{F}$.

(a) The lifetime of a series system with n components is determined by $Y = min(X_1, ..., X_n)$. Let g denote the PDF and G the CDF of Y. Then $r_G(x) = \theta(x) r_F(x)$, where

$$\theta(x) = \frac{n(\bar{F}(x))^{n-1}}{\sum_{j=0}^{n-1}(\bar{F}(x))^j} \leq 1$$

.

In this case, $r_G(x) \leq r_F(x)$. Thus, if $f(.)$ is a decreasing function, Then $H(f;[t]) \geq H(g;[t])$. If $g(.)$ is an increasing function, then $H(f;[t]) \leq H(g;[t])$.

(b) The lifetime of a parallel system with n components is determined by $Z = max(X_1, ..., X_n)$. Let g denote the PDF and G the CDF of Z. It can be shown that $r_G(x) = n r_F(x) \geq r_F(x)$. Thus, if the density $f(.)$ is an increasing function , then $H(f;[t]) \geq H(g;[t])$. If the density $g(.)$ is a decreasing function , then $H(f;[t]) \leq H(g;[t])$.

Example 2 *(Size-biased variable.)* Let X be a nonnegative continuous random variable with PDF f, CDF F, reversed hazard function r_F, and a finite mean. Then W is said to be a size-biased random variable corresponding to X if its density function is given by

$$g(w) = \frac{w}{E(X)} f(w), \qquad w \geq 0.$$

The hazard function of W is given by

$$r_G(x) = \frac{x}{\mu_{[F]}(x)} r_F(x), \qquad x > 0.$$

where

$$\mu_{[F]}(t) = E(t - X | X \leq t) = \begin{cases} t - \frac{\int_0^t F(y)dy}{F(t)} & F(t) > 0 \\ 0 & F(t) = 0 \end{cases}$$

is the mean past lifetime function. Since, $r_G(x) \geq r_F(x)$, we have $H(f;[t]) \leq H(g;[t])$ if $g(w)$ is decreasing in w and $H(f;[t]) \geq H(g;[t])$ if $f(x)$ is increasing in x.

3 Maximum inactivity entropy distribution

The reversed hazard function uniquely determines the lifetime distribution and hence its PDF $f(.)$. Thus, we can use the theorem 1 to identify the set of distributions $\Omega_F = \{f(.)\}$ in which $f^*(.)$ is the MIE; i.e. $H(f;[t]) \leq H(f^*;[t])$ for all $t > 0$. In this section, the differential equations, that describe the dynamic behavior of reversed hazard function of a distribution, are used to determine the MIE models because they have more simple closed form than reversed hazard functions.

For any distribution with a decreasing PDF $f^*(.)$, we can identify a set of distributions $\Omega_F = \{f(.)\}$ such that

$$r_F(t') = r_{F^*}(t') \qquad \exists t' > 0, \tag{6}$$

$$\frac{r'_F(t)}{r_F(t)} \leq \frac{r'_{F^*}(t)}{r_{F^*}(t)} \qquad for\, t \geq t', \tag{7}$$

$$\frac{r'_F(t)}{r_F(t)} \geq \frac{r'_{F^*}(t)}{r_{F^*}(t)} \qquad for\, t \leq t'. \tag{8}$$

Solution of the differential inequalities (5), (8) subject to (6) gives $r_F(t) \leq r_{F^*}(t)$ for all $t > 0$ for distributions in Ω_F. Thus by, theorem 1, $H(f^*;[t]) \geq H(f;[t])$. This is a general procedure that can be used to identify a set of distributions in which a given $f^*(.)$ is the MIE model. For distributions with increasing PDFs the first inequalities in (5), (8) is reversed. The set of distributions $\Omega_F = \{f(.)\}$ is referred to as the *maximal set* because it consists all distributions that are true in inequality (6), (5), (8).

For the distributions listed in table 2 the differential inequalities are in terms of $\rho(t)$, where $\rho(t) = h(t)r_F(t), h(t) > 0$ for all $t > 0$, because the constraints in terms of $r_F(t)$ are messy and not easily interpretable. Such as the inequality constraint (5),(8), for any distribution with a decreasing PDF $f^*(.)$, by theorem 1, the constraint $r_F(t') = r_{F^*}(t')\exists t' > 0, \frac{\rho'_F(t)}{\rho_F(t)} \leq \frac{\rho'_{F^*}(t)}{\rho_{F^*}(t)}$ for $t \geq t'$, $\frac{\rho'_F(t)}{\rho_F(t)} \geq \frac{\rho'_{F^*}(t)}{\rho_{F^*}(t)}$ for $t \leq t'$ gives $H(f^*;[t]) \geq H(f;[t])$.

Tables 1 and 2 display the constraints that define the maximal sets of distributions Ω_F in which some well-known distributions are MIE. In these

Table 1: Maximum Inactivity entropy distributions.

MIE density and support	constraints for maximal set
Uniform,$[0,\beta]$: $f^*(x)=\frac{1}{\beta}$	None
Exponential,$[0,\infty)$: $f^*(x)=\beta e^{-\beta x},\beta>0$	$\begin{cases} \exists t'>0: r_F(t')=r_{F^*}(t'), \\ \frac{r'_F(t)}{r_F(t)} \leq -(\beta+r_F(t)) \qquad for\, t\geq t', \\ \frac{r'_F(t)}{r_F(t)} \geq -(\beta+r_F(t)) \qquad for\, t\leq t'. \end{cases}$
Power,$[0,1]$ $f^*(x)=\alpha x^{\alpha-1},\alpha>0$	$\begin{cases} r_F(1)=\alpha \\ \frac{r'_F(t)}{r_F(t)} \geq \frac{-1}{\alpha} r_F(t) \quad \text{for } \alpha\leq 1 \\ \frac{r'F(t)}{r_F(t)} \leq \frac{-1}{\alpha} r_F(t) \quad \text{for } \alpha\geq 1. \end{cases}$
Half-Cauchy,$[0,\infty)$ $f^*(x)=\frac{2}{\pi(1+x^2)}$	$\begin{cases} \exists t'>0: r_F(t')=r_{F^*}(t'), \\ \frac{r'_F(t)}{r_F(t)} \leq -(2t\arctan t+1) r_F(t) \qquad for\, t\geq t', \\ \frac{r'_F(t)}{r_F(t)} \geq -(2t\arctan t+1) r_F(t) \qquad for\, t\leq t'. \end{cases}$
Half-logistic,$[0,\infty)$ $f^*(x)=\frac{\beta(\alpha+1)e^{\beta x}}{(\alpha+e^{\beta x})^2},\alpha,\beta>0$	$\begin{cases} \exists t'>0: r_F(t')=r_{F^*}(t'), \\ \frac{r'_F(t)}{r_F(t)} \leq \frac{-1}{\alpha+1}(e^{\beta t}+\alpha e^{-\beta t}) r_F(t) \qquad for\, t\geq t', \\ \frac{r'_F(t)}{r_F(t)} \geq \frac{-1}{\alpha+1}(e^{\beta t}+\alpha e^{-\beta t}) r_F(t) \qquad for\, t\leq t'. \end{cases}$

Table 2: Maximum Inactivity entropy distributions.

MIE density and support	constraints for maximal set
Weibull,$[0,\infty)$ $f^*(x)=k\alpha(x+x_0)^{\alpha-1}e^{-(x+x_0)^\alpha}$, $x_0>0, 0<\alpha<1$	$\begin{cases} \exists t'>0: r_F(t')=r_{F^*}(t'), \\ \frac{\rho'_F(t)}{\rho_F(t)} \le -(t+x_0)^{\alpha-1}(\alpha+\rho_F(t)) \quad for\ t\ge t', \\ \frac{\rho'_F(t)}{\rho_F(t)} \ge -(t+x_0)^{\alpha-1}(\alpha+\rho_F(t)) \quad for\ t\le t' \\ \text{for } \rho_F(t)=(x_0+t)^{1-\alpha}r_F(t) \end{cases}$
Linear failure rate,$[0,\infty)$ $f^*(x)=(\alpha+\beta x)e^{-(\alpha x+(\beta/2)x^2)}$ $0<\beta\le\alpha^2$	$\begin{cases} \exists t'>0: r_F(t')=r_{F^*}(t'), \\ \frac{\rho'_F(t)}{\rho_F(t)} \le -(1+\rho_F(t))(\alpha+\beta t) \quad for\ t\ge t', \\ \frac{\rho'_F(t)}{\rho_F(t)} \ge -(1+\rho_F(t))(\alpha+\beta t) \quad for\ t\le t \\ \text{for } \rho_F(t)=\frac{r_F(t)}{\alpha+\beta t} \end{cases}$
Extreme value,$[0,\infty)$ $f^*(x)=\alpha e^{(\beta/\alpha)x}\exp(\frac{\alpha^2}{\beta}(1-e^{(\beta/\alpha)x}))$, $0<\beta\le\alpha^2$	$\begin{cases} \exists t'>0: r_F(t')=r_{F^*}(t'), \\ \frac{\rho'_F(t)}{\rho_F(t)} \le -(\alpha+\rho_F(t))e^{(\beta/\alpha)t} \quad for\ t\ge t', \\ \frac{\rho'_F(t)}{\rho_F(t)} \ge -(\alpha+\rho_F(t))e^{(\beta/\alpha)t} \quad for\ t\le t' \\ \text{for } \rho_F(t)=e^{-(\beta/\alpha)t}r_F(t) \end{cases}$
Generalized Pareto,$[0,\infty)$ $f^*(x)=\frac{\beta+1}{\alpha}(1+\frac{\beta}{\alpha}x)^{-1/\beta-2}$, $\alpha,\beta>0$	$\begin{cases} \exists t'>0: r_F(t')=r_{F^*}(t'), \\ \frac{\rho'_F(t)}{\rho_F(t)} \le -\frac{1}{(1+(\beta/\alpha)t)}(\frac{\beta+1}{\alpha}+\rho_F(t)) \quad for\ t\ge t', \\ \frac{\rho'_F(t)}{\rho_F(t)} \ge -\frac{1}{(1+(\beta/\alpha)t)}(\frac{\beta+1}{\alpha}+\rho_F(t)) \quad for\ t\le t', \\ \text{for } \rho_F(t)=(1+\frac{\beta}{\alpha}t)r_F(t) \end{cases}$
Pareto(Type II),$[0,\infty)$ $f^*(x)=\alpha\beta^\alpha(\beta+x)^{-\alpha-1}$, $\alpha,\beta>0$	$\begin{cases} \exists t'>0: r_F(t')=r_{F^*}(t'), \\ \frac{\rho'_F(t)}{\rho_F(t)} \le -\frac{1}{(\beta+t)}(\alpha+\rho_F(t)) \quad for\ t\ge t', \\ \frac{\rho'_F(t)}{\rho_F(t)} \ge -\frac{1}{(\beta+t)}(\alpha+\rho_F(t)) \quad for\ t\le t' \\ \text{for } \rho_F(t)=(\beta+t)r_F(t) \end{cases}$
Beta,$[0,1]$ $f^*(x)=\beta(1-x)^{\beta-1}, \beta>0$	$\begin{cases} \frac{\rho'_F(t)}{\beta+\rho_F(t)} \ge -(1-x)^{\beta-1}(\beta+\rho_F(t)) \\ \quad \text{for } \beta\le 1, \rho_F(t)=(1-t)r_F(t) \\ \\ \frac{\rho'_F(t)}{\beta+\rho_F(t)} \le -(1-x)^{\beta-1}(\beta+\rho_F(t)) \\ \quad \text{for } \beta\ge 1, \rho_F(t)=(1-t)r_F(t) \end{cases}$

tables, the distributions are arranged according to the type of differential equations (complexity of the inequality constraints) for their maximal sets. The first constraint $\exists t' > 0 : r_F(t') = r_{F^*}(t')$ says that the maximal set Ω_F consists the distributions which their reversed hazard function plot meets the reversed hazard function plot of MIE distribution in at least a time $t > 0$. For the distribution listed in table 1, constraints are described in terms of the relative growth of the reversed hazard function, $\frac{r'_F(t)}{r_F(t)}$, and in table 2, constraint for maximal set are described in terms of the relative growth of $\rho(t)$, $\frac{\rho'_F(t)}{\rho_F(t)}$.

The list in table 1 begins with the uniform distribution, which its past entropy is maximum in the set of all distribution with no constraints.

The proof for the exponential distribution is given by the following theorem.

Theorem 2 *The exponential distribution is the MIE model in the set of distributions Ω_F with differentiable PDFs $f(.)$ over the support $[0, \infty)$ and the reversed hazard function $r_F(t)$,where*

(a) $\exists t' > 0 : r_F(t') = r_{F^*}(t')$,

(b) $\frac{r'_F(t)}{r_F(t)} \leq -(\beta + r_F(t)) \qquad for t \geq t'$,

(c) $\frac{r'_F(t)}{r_F(t)} \geq -(\beta + r_F(t)) \qquad for t \leq t'$.

Proof: from (b) we get

$$\frac{r'_F(t)}{r_F(t)(\beta + r_F(t))} \leq -1 \qquad for t \geq t'.$$

Multiplying both sides by β and decomposing the fraction, we obtain that

$$\frac{\beta r'_F(t)}{r_F(t)(\beta + r_F(t))} = \frac{r'_F(t)}{r_F(t)} - \frac{r'_F(t)}{(\beta + r_F(t))} \leq -\beta. \qquad for t \geq t'$$

Integrating both sides over $[t', t]$, $t' > 0$ is the time that $r_F(t') = r_{F^*}(t')$, and solving for $r_F(t)$ give

$$r_F(t) \leq \frac{\beta}{e^{\beta t} - 1} = r_{F^*}(t) \qquad for t \geq t',$$

where $r_F(t)$ is the reversed hazard function of F and $r_{F^*}(t)$ is the reversed hazard function of exponential with PDF shown in table 1. Using similar arguments as above for part (c), we get $r_F(t) \leq r_{F^*}(t)$ for $t \leq t'$. Then $r_F(t) \leq r_{F^*}(t)$ for $t > 0$. Since the exponential PDF is decreasing in x, by Theorem 1, $H(f;[t]) \leq H(f^*;[t])$.
The proof for the other distributions in table 1 is similar.

Table 2 lists the maximal sets for several other distribution that their constraints are described in terms of the relative growth $\rho(t) = h(t)r_F(t), h(t) > 0$ for all $t > 0$. The result for the weibull distribution is given by the following theorem. The proof for other distributions listed in table 2 essentially follows the same steps.

Theorem 3 *The weibull distribution is the MIE model in the set of distributions Ω_F with differentiable PDFs $f(.)$ over the support $[0, \infty)$ and the reversed hazard function $r_F(t)$,where*

(a) $\exists t' > 0 : r_F(t') = r_{F^*}(t')$

(b) $\frac{\rho'_F(t)}{\rho_F(t)} \leq -(t+x_0)^{\alpha-1}(\alpha + \rho_F(t)) \qquad for\, t \geq t',$

(c) $\frac{\rho'_F(t)}{\rho_F(t)} \geq -(t+x_0)^{\alpha-1}(\alpha + \rho_F(t)) \qquad for\, t \leq t'.$

with $\rho_F(t) = (x_0+t)^{1-\alpha} r_F(t)$.

Proof: From the condition (b), we have

$$\frac{\rho'_F(t)}{\rho_F(t)(\alpha + \rho_F(t))} \leq -(t+x_0)^{\alpha-1} \qquad for\, t \geq t',$$

Multiplying both sides by α and decomposing the fraction, we obtain that

$$\frac{\alpha\rho'_F(t)}{\rho_F(t)(\alpha + \rho_F(t))} = \frac{\rho'_F(t)}{\rho_F(t)} - \frac{\rho'_F(t)}{(\alpha + \rho_F(t))} \leq -\alpha(t+x_0)^{\alpha-1} \qquad for\, t \geq t',$$

Integrating both sides over $[t', t]$, $t' > 0$ is the time that $r_F(t') = r_{F^*}(t')$, and solving for $r_F(t)$ using the fact that $\rho_F(t) = (x_0+t)^{1-\alpha} r_F(t)$ give

$$r_F(t) \leq \frac{\alpha(x+x_0)^{\alpha-1}}{e^{(x+x_0)^\alpha - x_0^\alpha} - 1} = r_{F^*}(t) \qquad for\, t \geq t',$$

where $r_F(t)$ is the reversed hazard function of F and $r_{F^*}(t)$ is the reversed hazard function of weibull with PDF shown in table 2. Using similar arguments as above for part (c), we get $r_F(t) \leq r_{F^*}(t)$ for $t \leq t'$. Then $r_F(t) \leq r_{F^*}(t)$ for $t > 0$. Since the weibull PDF is decreasing in x, by Theorem 1, $H(f;[t]) \leq H(f^*;[t])$.

The next theorem shows the constraints that define the maximal sets Ω_F in which the mixture of two power distributions is a MIE distribution. It should be noted that the moment constraints that characterize this mixure model is not available.

Theorem 4 *The mixture of two power distributions with PDF* $f^*(x) = \alpha\beta_1 x^{\beta_1-1} + (1-\alpha)\beta_2 x^{\beta_2-1}, 0 < x < 1, 0 < \alpha < 1, 0 < \beta_2 < \beta_1 < 1 (1 < \beta_2 < \beta_1)$, *is the MIE model in the set of distributions* Ω_F *with differentiable PDFs* $f(.)$ *over the support [0,1] and reversed hazard functions* $r_F(t)$, *where*

(a) $r_F(1) = \alpha\beta_1 + (1-\alpha)\beta_2, 0 < \alpha < 1, 0 < \beta_2 < \beta_1 < 1 (1 < \beta_2 < \beta_1)$,

(b) $\beta_2 < \rho_F(t) < \beta_1$,

(c) $\frac{t\rho'_F(t)}{\rho_F(t)-\beta_2} \geq -(\rho_F(t)-\beta_1) (\frac{t\rho'_F(t)}{\rho_F(t)-\beta_2} \leq -(\rho_F(t)-\beta_1))$,

with $\rho_F(t) = tr_F(t)$.

Proof: from (c) we get $\frac{\rho'_F(t)}{(\rho_F(t)-\beta_2)(\beta_1-\rho_F(t))} \geq 1/t$. Multiplying both sides by $\beta_1 - \beta_2$ and decomposing the fraction, we obtain that

$$\frac{\rho'_F(t)}{\beta_1 - \rho_F(t)} + \frac{\rho'_F(t)}{\rho_F(t) - \beta_2} \geq \frac{(\beta_1 - \beta_2)}{t}.$$

Integrating both sides over $[t, 1]$ using the initial condition (a), and solving for $\rho_F(t)$ using the fact that $\rho_F(t) = tr_F(t)$ give

$$r_F(t) \leq \frac{\alpha\beta_1 x^{\beta_1-1} + (1-\alpha)\beta_2 x^{\beta_2-1}}{\alpha x^{\beta_1} + (1-\alpha)x^{\beta_2}} = r_{F^*}(t),$$

where $r_F(t)$ is the reversed hazard function of F and $r_{F^*}(t)$ is the reversed hazard function of mixture of two powers. Since the mixture of two powers PDF with parameters $\beta_1, \beta_2 < 1$ is decreasing in x, by Theorem 1, $H(f;[t]) \leq H(f^*;[t])$.

References

[1] Asadi, M., Ebrahimi, N., Hamedani, G.G. and Soofi, E.S. (2004). Maximum dynamic entropy models. *J. Appl. Prob.* **41,** 379-390.

[2] Ebrahimi, N. (2000). The Maximum entropy method for lifetime distributions. *Sankhyā, A* **62,** 223-235.

[3] Ebrahimi, N., Hamedani, G.G. and Soofi, E.S. (1996). Maximum entropy modeling with partial information on failure rate. School of Business Administration, University of Wisconsin-Milwaukee.

[4] Ebrahimi, N., Soofi, E.S. and Soyer, R. (2008). Multivariate maximum entropy identification, transformation, and dependence. To appear in *J. multivariate Analysis.*

[5] Jaynes, E.T. (1982). On the rationale of maximum entropy methods. *Proc. IEEE* **70,** 939-952.

[6] Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* **27,** 379-423.

Probabilistic normed space and statistical convergent

P. Azhdari

Department of Statistics, Islamic Azad University of north Branch

Abstract: In this paper after defining the concepts of statistical convergence and statistical Cauchy on probabilistic normed spaces, we give a characterization for statistically convergent sequences and some theorems about this kind of convergency.

Keywords: Probabilistic normed space, statistically convergent sequences, statistically Cauchy.

1 Introduction

The concept of probabilistic metric spaces introduced in 1942 by Menger [6]. Also the theory of probabilistic normed spaces is an important generalization of normed spaces [9]. It is interested to extend statistical convergence to probabilistic normed spaces. This idea was introduced by Steinhaus [8] but rapid developments started after the paper of Connor [2] and Fridy [3].

2 Preliminaries

A mapping $F : (-\infty, \infty) \to [0, 1]$ is called a distribution function if it is nondecreasing and left continuous on $(0, \infty)$ with $F(0) = 0$. The class of all distribution functions is denoted by Δ_+.A Menger space (probabilistic metric space) is an order pair (X, F) where X is a nonempty set and F is a mapping from $X \times X$ to Δ_+ is denoted by $F_{pq}(.)$ satisfies the following conditions:

1) $F_{pq}(x) = 1$ for all $x > 0$ iff $p = q$ $\forall\, p, q \in X$,
2) $F_{pq}(x) = F_{qp}(x)$ $\forall\, p, q \in X,\ x \in R$,
3) If $F_{pq}(t) = 1\ ,\ F_{qr}(s) = 1$ then $F_{pr}(t+s) = 1$ for $p, q, r \in X\ , x, y \in R$.

A binary operation $T : [0, 1] \times [0, 1] \to [0, 1]$ is called a triangular norm (abbreviated t-norm) if the following conditions are satisfied:

1) $T(a,1)=a$ forevery $a\in[0,1]$,
2) $T(a,b)=T(b,a)$ forevery $a,b\in[0,1]$,
3) $a\geq b,\ c\geq d\ \Rightarrow\ T(a,c)\geq T(b,d)\ \ a,b,c\in[0,1]$.
4) $T(T(a,b),c)=T(a,T(b,c))$.

Definition 1.2(Schweizer and Sklar): A generalized Menger space is a triple (X,F,T) where (X,F) is a probabilistic metric space, T is a t-norm and the following inequlity hold:

$$F_{pr}(x+y)\geq\ T(F_{pq}(x),F_{qr}(y))\ \ \forall\ p,q,r\in X,\ \forall\ x,y\geq 0$$

Definition 2.2: A triple (X,F,T) is called a probabilistic normed space if X is a real vector space,F is a mapping X into Δ_+ and T is a t-norm satisfying the following the following conditions:
1) $F_x(0)=0$
2) $F_x(t)=1\ \ \forall t>0\Leftrightarrow x=0$
3) $F_{\alpha x}(t)=F_x(\frac{t}{|\alpha|})\ \ \forall\ \alpha\in\ R-\{0\}$
4) $F_{x+y}(s+t)\geq\ T(F_x(s),F_y(t))\quad\forall\ x,y\in\ X,\ s,t\in\ R^+$.

Definition 3.2: **[1]** Let (X,F,T) be a probabilistic normed space. A sequence (x_n) of elements of X is said to be convergent to $\xi\ \in\ X$ with respect to the probabilistic norm F if for every $\epsilon\ >0$ and $\lambda\in\ (0,1)$ there exists a positive integer m_0 such that $F_{x_n-\xi}(\epsilon)>1-\lambda,\quad\forall\ n\geq\ m_0$ and we write $F-\lim\ x_n=\xi$ or $x_n\rightarrow\ \xi$ with respect to F. A sequence (x_n) of elements of X is said to be Cauchy with respect to the probabilistic norm F if $F_{x_n-x_m}(\epsilon)>1-\lambda,\quad\forall\ n,m\ \geq\ m_0$.

Definition 4.2: **[4]** Let (X,F,T) be a probabilistic normed space. A double sequence $x=(x_{ij})$ of elements of X is said to be convergent to $\xi\ \in\ X$ with respect to the probabilistic norm F if for every $\epsilon\ >0$ and $\lambda\in$ $(0,1)$ there exists a positive integer m_0 such that $F_{x_{ij}-\xi}(\epsilon)>1-\lambda$,whenever $i,j\geq\ m_0$ and we write $F_2-\lim\ x=\xi$ or $x_{ij}\rightarrow\ \xi$ as $i,j\rightarrow\infty$ with respect to F. A double sequence $x=(x_{ij})$ of elements of X is said to be Cauchy with respect to the probabilistic norm F if for every $\epsilon>0$ and $\lambda\ \in(0,1)$, there exist m_0 and n_0 such that $F_{x_{ij}-x_{kl}}(\epsilon)>1-\lambda$, for all

$i, k \geq m_0$, $j, l \geq n_0$.

3 Main Results

The natural density of a subset A of N is denoted by $\delta(A)$ and is defined by

$$\delta(A) = \lim_{n\to\infty} \frac{1}{n}|\{k < n : k \in A\}|$$

where the vertical bar denotes the cardinality of the respective set. The natural density may not exist for each set A but the upper density $\overline{\delta}$ always exists and defined as follows:

$$\overline{\delta}(A) = \limsup_n \frac{1}{n}|\{k \leq n : k\}|.$$

A sequence $x = (x_n)$ of numbers is statistically convergent to ξ if

$$\delta(\{k \in \mathbb{N} : |x_k - \xi| \geq \epsilon\}) = 0$$

for each $\epsilon > 0$ and we write $st - \lim x = \xi$.

Definition 1.3: Let (X, F, T) be a probabilistic normed space. A sequence (x_n) of elements of X is said to be statistically convergent [4] to $\xi \in X$ with respect to the probabilistic norm F if for every $\epsilon > 0$ and $\lambda \in (0, 1)$ we have

$$\delta(\{n \in N : F_{x_n-\xi}(\epsilon) \leq 1 - \lambda\}) = 0.$$

or equivalently

$$\lim_n \frac{1}{n}|\{k \leq n : F_{x_n-\xi}(\epsilon) \leq 1 - \lambda\}| = 0.$$

The element ξ is called statistical limit of the sequence(x_n) with respect to the probabilistic norm F and we write $st_F - \lim x_n = \xi$.

Definition 2.3: A sequence (x_n) of elements of X is said to be statistically Cauchy [4] with respect to the probabilistic norm F if for every $\epsilon > 0$ and $\lambda \in (0, 1)$ we have

$$\delta(\{n \in N : F_{x_n-x_m}(\epsilon) \leq 1 - \lambda\}) = 0.$$

Lemma 1.3: Let (X, F, T) be a probabilistic normed space. Then for every $\epsilon > 0$ and $\lambda \in (0, 1)$ the following statements are equivalent:
1) $st_F - \lim\ x_n = \xi$,
2) $\delta(\{n \in N :\ F_{x_n - \xi}(\epsilon) \leq\ 1 - \lambda\}) = 0$,
3) $\delta(\{n \in N :\ F_{x_n - \xi}(\epsilon) >\ 1 - \lambda\}) = 1$,
4) $st - \lim\ F_{x_n - \xi}(\epsilon) = 1$.

Theorem 1.3: Let (X, F, T) be a probabilistic normed space. If a sequence (x_n) is statistically convergent with respect to the probabilistic norm F, then $st_F - \lim\ x_n = \xi$ is unique.

Proof: Assume that $st_F - \lim x = \xi_1$ and $st_F - \lim x = \xi_2$. For a given $\lambda > 0$ choose $\delta \in (0, 1)$ such that $T((1-\delta), (1-\delta)) > 1 - \lambda$. Then for every $\epsilon > 0$ define the following sets:

$$A_{F,1}(\delta, \epsilon) = \{k \in N : F_{x_k - \xi_1}(\epsilon) \leq 1 - \delta\},$$

$$A_{F,2}(\delta, \epsilon) = \{k \in N : F_{x_k - \xi_2}(\epsilon) \leq 1 - \delta\}.$$

Since $st_F - \lim x = \xi_1$, $\delta\{A_{F,1}(\delta, \epsilon)\} = 0$ for all $\epsilon > 0$. Furthermore, using $st_F - \lim x = \xi_2$, we get $\delta\{A_{F,2}(\delta, \epsilon)\} = 0$ for all $\epsilon > 0$. Now let $A_F(\delta, \epsilon) = A_{F,1}(\delta, \epsilon) \bigcap A_{F,2}(\delta, \epsilon)$. Then $\delta\{A_F(\delta, \epsilon)\} = 0$ which implies $\delta\{\mathbb{N} - A_F(\delta, \epsilon)\} = 1$. When $k \in \mathbb{N} - A_F(\delta, \epsilon)$ then we have

$$F_{\xi_1 - \xi_2}(\epsilon) \geq F_{x_k - \xi_1}(\frac{\epsilon}{2}) * F_{x_k - \xi_2}(\frac{\epsilon}{2}) > (1 - \delta) * (1 - \delta)$$

Since $(1 - \delta) * (1 - \delta) > 1 - \lambda$ therefore

$$F_{\xi_1 - \xi_2}(\epsilon) > 1 - \lambda$$

Since $\lambda > 0$ and by the last relation $F_{\xi_1 - \xi_2}(\epsilon) = 1$ for each $\epsilon > 0$ we obtain $\xi_1 = \xi_2$ and $st_F - limit$ is unique.

Theorem 2.3: Let (X, F, T) be a probabilistic normed space. If $F - \lim\ x_n = \xi$, then $st_F - \lim\ x_n = \xi$.

proof: for all $\lambda \in (0,1)$ and $\epsilon > 0$, there is a number $j \in \mathbb{N}$ such that $F_{x_k-\xi}(\epsilon) \leq 1-\delta\}$ for every $n > j$, therefore the set $\{n \in \mathbb{N} : \ F_{x_n-\xi}(\epsilon) > 1-\lambda\}$ has at most finitely many terms. Since every finite subset has density zero,$\delta(\{n \in \mathbb{N} : \ F_{x_n-\xi}(\epsilon) > 1-\lambda\}) = 0$.

Theorem 3.3: Let (X, F, T) be a probabilistic normed space. Then $st_F - \lim \ x_n = \xi$ if and only if there exists an increasing index sequence $K = \{k_n\}_{n\in N}$ of natural numbers such that $\delta\{K\} = 1$ and $F - \lim_{n\in \ K} \ x_n = \xi$, i.e., $F - \lim_n \ x_{k_n} = \xi$.

The statistical convergence of double sequences can be defined as follows:

Definition 3.3:[5] Let $A \subset N \times N$ be a two dimensional of positive integers and $A(n,m)$ be the numbers of (i,j) in A such that $i \leq n$ and $j \leq m$. The two dimensional natural density is the lower asymptotic density of set A is defined as

$$\underline{\delta_2}(A) = \liminf_{n,m} \frac{A(n,m)}{nm}$$

When the sequence $\frac{A(n,m)}{nm}$ has a limit in $Pringsheim's$ sense [7] then A has a double natural density and is defined as

$$\delta_2(A) = \lim_{n,m} \frac{A(n,m)}{nm}$$

Definition 4.3: A real double sequence $x = (x_{i,j})$ is said to be statistically convergent to number ξ if for every $\epsilon > 0$ the set

$$\{(i,j), i \leq n, j \leq m : \ |x_{i,j} - \xi| \geq \epsilon\}$$

has double natural density zero and we write $st_2 - \lim_{i,j} x_{i,j} = \xi$.

Definition 5.3: A real double sequence $x = x_{i,j}$ is said to be statistically Cauchy if for every $\epsilon > 0$ there exist $n = n_0(\epsilon)$ and $m = m_0(\epsilon)$ such that for all $i, k \geq n_0, j, l \geq m_0$ the set

$$\{(i,j), i \leq n, j \leq m : \ |x_{ij} - x_{kl}| \geq \epsilon\}$$

has double natural density zero.

Definition 6.3: Let (X, F, T) be a probabilistic normed space. A real double sequence $x = (x_{i,j})$ is said to be statistically convergent to number $\xi \in X$ with respect to the probabilistic normed F if for every $\epsilon > 0$ and $\lambda \in (0, 1)$,

$$A = \{(i, j), i \leq n, j \leq m : \ F_{x_{i,j}-\xi}(\epsilon) \leq 1 - \lambda\}$$

has double natural density zero, that is, if $A(n, m)$ become the numbers of (i, j) in A

$$\lim_{n,m} \frac{A(n, m)}{nm} = 0.$$

In this case we writes $st_{F_2} - \lim_{i,j} x_{i,j} = \xi$ where ξ is called to be $st_{F_2} - limit$, and the set of all statistically convergent double sequences with respect to probabilistic norm F by st_{F_2}.

Lemma 2.3: Let (X, F, T) be a probabilistic normed space.Then for every $\epsilon > 0$ and $\lambda \in (0, 1)$ the following statements are equivalent:
1) $st_{F_2} - \lim_{i,j} x_{i,j} = \xi$,
2) $\delta_2(\{(i, j),\ i \leq n,\ j \leq m : \ F_{x_{i,j}-\xi}(\epsilon) \leq \ 1 - \lambda\}) = 0$,
3) $\delta_2(\{(i, j),\ i \leq n,\ j \leq m : \ F_{x_{i,j}-\xi}(\epsilon) > \ 1 - \lambda\}) = 1$,
4) $st_2 - \lim \ F_{x_{i,j}-\xi}(\epsilon) = 1$.

Proof: By Definition 6.3 the first three statement are equivalent. We will show 2,4 are equivalent.By Definition we have
$\{(i, j), i \leq n, j \leq m : \ |F_{x_{i,j}-\xi}(\epsilon) - 1| \geq \lambda\} =$
$\{(i, j), i \leq n, j \leq m : \ F_{x_{i,j}-\xi}(\epsilon) \geq 1+\lambda\} \cup \{(i, j), i \leq n, j \leq m : \ F_{x_{i,j}-\xi}(\epsilon) \leq 1 - \lambda\}$
it follows 2,4 are equivalent.

Theorem 4.3: Let (X, F, T) be a probabilistic normed space. If a double sequence $x = (x_{ij})$ is statistically convergent with respect to the probabilistic norm F, then $st_{F_2} - limit$ is unique.

Proof: Assume that $st_{F_2} - \lim x = \xi_1$ and $st_{F_2} - \lim x = \xi_2$. For a given $\lambda > 0$ choose $\delta \in (0, 1)$ such that $T((1 - \delta), (1 - \delta)) > 1 - \lambda$. Then

for every $\epsilon > 0$ define the following sets:

$$A_{F,1}(\delta, \epsilon) = \{(i,j) \in N \times N : F_{x_{ij}-\xi_1}(\epsilon) \leq 1 - \delta\},$$

$$A_{F,2}(\delta, \epsilon) = \{(i,j) \in N \times N : F_{x_{ij}-\xi_2}(\epsilon) \leq 1 - \delta\}.$$

Since $st_{F_2} - \lim x = \xi_1$, $\delta_2\{A_{F,1}(\delta, \epsilon)\} = 0$ for all $\epsilon > 0$. Furthermore, using $st_{F_2} - \lim x = \xi_2$, we get $\delta_2\{A_{F,2}(\delta, \epsilon)\} = 0$ for all $\epsilon > 0$. Now let $A_F(\delta, \epsilon) = A_{F,1}(\delta, \epsilon) \bigcap A_{F,2}(\delta, \epsilon)$. Then $\delta_2\{A_F(\delta, \epsilon)\} = 0$ which implies $\delta_2\{\mathbb{N} \times \mathbb{N} - A_F(\delta, \epsilon)\} = 1$. When $(i,j) \in \mathbb{N} \times \mathbb{N} - A_F(\delta, \epsilon)$ then we have

$$F_{\xi_1-\xi_2}(\epsilon) \geq F_{x_{ij}-\xi_1}(\frac{\epsilon}{2}) * F_{x_{ij}-\xi_2}(\frac{\epsilon}{2}) > (1-\delta) * (1-\delta)$$

Since $(1-\delta) * (1-\delta) > 1 - \lambda$ therefore

$$F_{\xi_1-\xi_2}(\epsilon) > 1 - \lambda$$

Since $\lambda > 0$ and by the last relation $F_{\xi_1-\xi_2}(\epsilon) = 1$ for each $\epsilon > 0$ we obtain $\xi_1 = \xi_2$ and $st_F - limit$ is unique.

Theorem 2.3: Let (X, F, T) be a probabilistic normed space. If for a double sequence $x = (x_{i,j})$, $F_2 - \lim\ x = \xi$, then $st_{F_2} - \lim\ x = \xi$.

proof: for all $\lambda \in (0,1)$ and $\epsilon > 0$, there is a number $k \in \mathbb{N}$ such that $F_{x_{ij}-\xi}(\epsilon) \leq 1 - \delta\}$ for every $i > k$ and $j > k$, therefore the set $\{(i,j) \in \mathbb{N} \times \mathbb{N} :\ F_{x_{ij}-\xi}(\epsilon) \leq 1 - \lambda\}$ has at most finitely many terms. Since every finite subset has double density zero then ,$\delta_2(\{(i,j) \in \mathbb{N} \times \mathbb{N} : F_{x_{i,j}-\xi}(\epsilon) \leq 1 - \lambda\}) = 0$.

References

[1] Aghajani A.(2006). *Convex sets in probabilistic normed spaces.* Chaos, Solitons & Fractals.

[2] Connor J. (1988). The statistical and strong p-Cesaro convergence of sequences. *Analysis*, **8**, 47-63.

[3] Fridy J.A. (1993). On statistical convergence. *Analysis.*, **118**, 1187-1193

[4] Karakus S. (2007). Statistical convergence on probabilistic normed spaces. *Mathematical Communications*, 12, 11-23.

[5] Karakus S. Demici K. (2007). Statistical convergence of double sequences on probabilistic normed spaces. *International Journal of Mathematics and Mathematical Science*, 2007, 11.

[6] Menger K. (1942). *Statistical metric. Proc Nat Acad Sci, USA*, 28, 535-7.

[7] Pringshein A. (1900). Zur theorie der zweifach unendlichen zahlenfolgen. *Mathematische Annalen*, 53(2), 289-321.

[8] Steinhaus H. (1951). Sur la convergence ordinaire et la covergence asymptotique. *Collog. Math*, 2, 73-74.

[9] Schiweizer B, Sklar A. (1983). Probabilistic metric spaces. *North-Holland,. Amesterdam.*

On $M/(G_1,G_2)/1/G(BS)/V_s$ with optional second service

A. Badamchi zadeh

Department of Statistics , Allameh Tabatabaii University

Abstract: We analyze a single server queue with Poisson input, two heterogeneous services in which the second service is optional and the second phase of service has many cases. The first phase of service is essential for all customers, but with probability γ a tagged customers chose second phase or with probability $1-\gamma$ leave the system. After completion of first phase or second phase of service, the server either goes for a vacation with probability $\theta(0 \leq \theta \leq 1)$, or may continue to serve the next unit with probability $1-\theta$, if any. Otherwise, it remains in the system until a customer arrives. In this paper we derive the steady- state equations, and PGF's of the system is obtained. By using them the mean queue size of the system at the departure epoch is obtained as a classical generalization of Pollaczek-Khinchine formula for $M/G/1$ queueing system. Finally, by using Little's formulas we have obtained other measures of system and some extended systems.

Keywords: $M/G/1$ Queue, Two phase of heterogeneous service, Bernoulli vacation, Mean queue size, Mean response time.

1 Introduction

For the first time the concept of Bernoulli vacation were studied by Keilson and Servi [4,5]. They introduced the concept of modified service time which has a main rule in systems with general service and vacation time. Recently Madan [7],Madan and Choudhury [8] studied the $M/G/1$ queueing system with two phase of heterogeneous services so that the first phase of service followed by second phase of service under a Bernoulli vacation schedule. An $M^X/G/1$ queue system with an additional service channel were analyzed by Choudhury [1]. Furthermore a similar work can be found in [9] . Considerable efforts have been devoted in this model by Servi[10], Takagi[11] and Doshi[2].

In many applications such as hospital services, production systems, bank services ,computer and communication networks; there is two phase of services such that the first phase is essential for all customers, but as

soon as the essential services completed, it may leave the system or may immediately go for the second phase of service. We assume that the second phase of service has two optional cases, where customers may choose one of them with certain probability, however these cases are extended to k cases easily. This approach is explained in section 5 of this paper. Also in this system for overhauling or maintenance of the system, or serving other customers, or fatigue of the server or for other reasons not mentioned here , the server may go to vacation.

In this paper we analyze a single server queue with Piosson input, two phase of heterogeneous service which second phase has optional cases and Bernoulli vacation for server. In section 2 we deal with the mathematical model and definitions. Steady-State conditions and generating functions are discussed in section 3.

Mean queue size and mean response time are computed in section 4 where in section 5 a general case and some special cases are investigated.

2 Mathematical model and definitions

We consider a queueing system such that:
i) Customers arrive at the system one by one in a Poisson stream with mean rate $\lambda > 0$.
ii)The server provides two phases of heterogeneous service in succession. The service discipline is assumed to be on the basis of first come, first served(FCFS). The first phase of service is essential for all customers, but as soon as the essential service with probability $\gamma(0 \leq \gamma \leq 1)$ is completed, a tagged customer leaves the system, or moves for second phase with probability $1 - \gamma$. The second phase has two cases(alternative) where server choose first and second case with probability of p_1 and p_2 respectively such that $p_1 + p_2 = 1$.
The service times for two phases are independent random variables, where we denote them by B_1, B_2. Their Laplace-Stieltjes transform (LST)are $B_1^*(s), B_2^*(s)$ where we assume they have finite moments $E(B_i^l)$ for $l1$ and i=1,2. Also for second phase of service the random variables S_1 and S_2 denote the service time of first and second case, respectively.Their corresponding LST are shown as $S_1^*(s)$ and $S_2^*(s)$ respectively. Also we assume

that the $E(S_i^l)$ is finite for $l \geq 1$ and $i = 1, 2$. In the other word

$$B_2 = p_1 S_1 + p_2 S_2$$

and

$$B_2^*(s) = p_1 S_1^*(s) + p_2 S_2^*(s) \tag{1}$$

iii)With probability of $(1 - \gamma)$, as soon as the first phase of a customer complete or the second phase complete, the server may go for a vacation of random length V with probability $\theta(0 \leq \theta \leq 1)$ or it may continue to serve the next customer, if any, with probability $(1 - \theta)$, otherwise it remains in the system and waits for a new arrival.We denote $V(x), V^*(s)$ and $E(V^l)$ for distribution function (DF),LST and $l'th$ finite moment of V, respectively where $l \geq 1$.
iv)The random variables B_1, B_2, V and also S_1, S_2 are all independent.

DEFINITION 0.1 *The modified service time or the time required by a customer to complete the service cycle is given by*

$$B = \begin{cases} B_0 + V & \textit{with probability } \theta \\ B_0 & \textit{with probability } (1 - \theta) \end{cases} \tag{2}$$

then the LST $B^*(s)$ of B is given by

$$B^*(s) = \theta B_0^*(s) V^*(s) + (1 - \theta) B_0^*(s) \tag{3}$$

where

$$B_0 = \begin{cases} B_1 + B_2 & \textit{with probability } \gamma \\ B_1 & \textit{with probability } (1 - \gamma) \end{cases} \tag{4}$$

and

$$B_0^*(s) = \gamma B_1^*(s) B_2^*(s) + (1 - \gamma) B_1^*(s) \tag{5}$$

and$B_2^*(s)$ is given by (1). Also

$$E(B) = -\frac{dB^*(s)}{ds} \mid_{s=0} = E(B_0) + \theta E(V) \tag{6}$$

where

$$E(B_0) = -\frac{dB_0^*(s)}{ds} \mid_{s=0} = E(B_1) + \gamma E(B_2) \\ = E(B_1) + \gamma[p_1 E(S_1) + p_2 E(S_2)] \tag{7}$$

and

$$E(B^2) = \frac{(-1)^2 d^2 B^*(s)}{ds^2} \mid_{s=0} = E(B_0^2) + 2\theta E(B_0)E(V) + \theta\ E(V^2) \tag{8}$$

where

$$E(B_0^2) = \frac{(-1)^2 d^2 B_0^*(s)}{ds^2} \mid_{s=0} = E(B_1^2) + 2\gamma E(B_1)E(B_2) + \gamma\ E(B_2^2) \tag{9}$$

Further we assume for $i = 1, 2; \quad B_i(0) = 0, \quad B_i(\infty) = 1$ *and* $B_i(x)$ *are continuous at* $x = 0$ *, so that*

$$\mu_i(x)dx = \frac{dB_i(x)}{1 - B_i(x)} \tag{10}$$

are the first order differential equation(hazard rate functions) of B_i *.Also* $V(0) = 0, V(\infty) = 1$ *and* $V(x)$ *is continuous at* $x = 0$ *, so that*

$$\nu(x)dx = \frac{dV(x)}{1 - V(x)} \tag{11}$$

is hazard rate function of V.

We assume $N_Q(t)$ to be the queue size at time 't' and the supplementary variables as follows are defined :

$B_1^0(t)[B_2^0(t)] \equiv$ the elapsed first [second] phase of service at time 't'
$V^0(t) \equiv$ the elapsed vacation time at time 't'

Now let us introduce the following random variables :

$$Y(t) = \begin{cases} 0 & \text{if the server is idle at time 't',} \\ 1[2] & \text{if the server is busy with first[second] phase of service,} \\ & \text{at time 't',} \\ 3 & \text{if the server is on vacation at time 't'.} \end{cases} \tag{12}$$

From this we have a bivariate Markov process $\{N_Q(t), L(t)\}$ where $L(t) = 0$ if $Y(t) = 0$; $L(t) = B_i^0(t)$ if $Y(t) = i$ for $i = 1, 2$ and $L(t) = V^0(t)$ if $Y(t) = 3$ .Now for $i = 1, 2$ the following probabilities defined as

$Q_n(x,t) = Prob[N_Q(t) = n, L(t) = V^0(t); x < V^0(t)x + dx] \quad x > 0, \quad n0$

$P_{i,n}(x,t) = Prob[N_Q(t) = n, L(t) = B_i^0(t); x < B_i^0(t)x + dx] \quad x > 0, \quad n0$ and

$$R_0(t) = Prob[N_Q(t) = 0, L(t) = 0]$$

Now the analysis of the limiting behaviour of this queueing process at a random epoch can be performed with the help of Kolmogorov forward equations, provided the following limits exist and independent of initial state :

$$R_0 = \lim_{t\to\infty} R_0(t)$$
$$P_{i,n}(x)dx = \lim_{t\to\infty} P_{i,n}(x,t)dx \quad i = 1, 2 \quad x > 0, \ n0$$
$$Q_n(x)dx = \lim_{t\longrightarrow\infty} Q_n(x,t)dx \quad x > 0, \quad n > 0$$

Now for $i = 1, 2$ the PGF of this probabilities are defined as follow:

$$P_i(x,z) = \sum_{n=0}^{\infty} z^n P_{i,n}(x) \quad |z|1, \quad x > 0 \tag{13}$$

$$P_i(0,z) = \sum_{n=0}^{\infty} z^n P_{i,n}(0) \quad |z|1 \tag{14}$$

Also

$$Q(x,z) = \sum_{n=0}^{\infty} z^n Q_n(x) \quad |z|1, \quad x > 0 \tag{15}$$

$$Q(0,z) = \sum_{n=0}^{\infty} z^n Q_n(0) \tag{16}$$

3 Steady-state probability generating function

From kolmogorov forward equations , for $i = 1, 2$ the steady-state conditions can be written as follow

$$\frac{d}{dx}P_{i,n}(x) + [\lambda + \mu_i(x)]P_{i,n}(x) = \lambda P_{i,n-1}(x) \quad n \geq 0, \quad x > 0 \tag{17}$$

and

$$\frac{d}{dx}Q_n(x) + [\lambda + \nu(x)]Q_n(x) = \lambda Q_{n-1}(x) \quad n \geq 0, \quad x > 0 \tag{18}$$

also

$$\begin{aligned} \lambda R_0 &= (1-\theta)(1-\gamma)\int_0^{+\infty} \mu_1(x)P_{1,0}(x)dx + (1-\theta) \\ &\quad \times \int_0^{+\infty} \mu_2(x)P_{2,0}(x)dx + \int_0^{+\infty} \nu(x)Q_0(x)dx \end{aligned} \tag{19}$$

we set $P_{1,-1}(x) = 0 \quad , P_{2,-1}(x) = 0$ and $Q_{-1}(x) = 0$ in (17), (18) and (19) .

These set of equations are to be solved under the following boundary conditions at $x = 0$

$$\begin{aligned} P_{1,0}(0) = \lambda R_0 + (1-\theta)(1-\gamma)\int_0^{+\infty} \mu_1(x)P_{1,1}(x)dx + (1-\theta) \\ \times \int_0^{+\infty} \mu_2(x)P_{2,1}(x)dx + \int_0^{+\infty} \nu(x)Q_1(x)dx \end{aligned} \tag{20}$$

and for $n > 0$

$$\begin{aligned} P_{1,n}(0) = (1-\theta)(1-\gamma)\int_0^{+\infty} \mu_1(x)P_{1,n+1}(x)dx + (1-\theta) \\ \int_0^{+\infty} \mu_2(x)P_{2,n+1}(x)dx + \int_0^{+\infty} \nu(x)Q_{n+1}(x)dx \end{aligned} \tag{21}$$

$$P_{2,n}(0) = \gamma \int_0^{+\infty} \mu_1(x)P_{1,n}(x)dx, \quad n0 \tag{22}$$

also

$$Q_n(0) = (1-\gamma)\theta \int_0^{+\infty} \mu_1(x)P_{1,n}(x)dx + \theta \int_0^{+\infty} \mu_2(x)P_{2,n}(x)dx, \quad n0 \tag{23}$$

Finally the normalizing condition is

$$R_0+\sum_{i=1}^{2}\sum_{n=0}^{\infty}\int_{o}^{+\infty}P_{i,n}(x)dx+\sum_{n=0}^{\infty}\int_{0}^{+\infty}Q_n(x)dx=1 \tag{24}$$

LEMMA 0.2 *For $i=1,2$ from (17) we have*

$$P_i(x,z)=P_i(0,z)[1-B_i(x)]e^{-\lambda(1-z)x}\quad x>0 \tag{25}$$

and from (18)

$$Q(x,z)=Q(0,z)[1-V(x)]e^{-\lambda(1-z)x}\quad x>0 \tag{26}$$

PROPOSITION 0.3 *If for i=1,2*

$$B_i^*(\lambda-\lambda z)=\int_{0}^{+\infty}e^{-\lambda(1-z)x}dB_i(x) \tag{27}$$

$$V^*(\lambda-\lambda z)=\int_{0}^{+\infty}e^{-\lambda(1-z)x}dV(x) \tag{28}$$

are the z-transform of B_i and V respectively ,then by multiplying (21) in z^n and summation from $n=1$ to $+\infty$, adding (20) to result and using (19) we have :

$$\begin{aligned}zP_1(0,z)=\lambda R_0(z-1)+(1-\theta)(1-\gamma)P_1(0,z)B_1^*(\lambda-\lambda z)\\+(1-\theta)P_2(0,z)B_2^*(\lambda-\lambda z)+Q(0,z)V^*(\lambda-\lambda z)\end{aligned} \tag{29}$$

PROPOSITION 0.4 *By multiplying (22) in z^n and summation on $n=1$ to $+\infty$ we have:*

$$P_2(0,z)=\gamma P_1(0,z)B_1^*(\lambda-\lambda z) \tag{30}$$

similarly from (23)

$$Q(0,z)=(1-\gamma)\theta P_1(0,z)B_1^*(\lambda-\lambda z)+\theta P_2(0,z)B_2^*(\lambda-\lambda z) \tag{31}$$

In the rest of this section for simplifying we omit $(\lambda-\lambda z)$.

COROLLARY 0.5 *I) By substitute $P_2(0,z)$ and $Q(0,z)$ from (30), (31) in (24) we have*

$$P_1(0,z) = \frac{\lambda R_0(z-1)}{z - \{(1-\gamma)[(1-\theta)+\theta V^*]B_1^* + \gamma[(1-\theta)+\theta V^*]B_1^*B_2^*\}} \quad (32)$$

Since

$$P_1(z) = \int_0^{+\infty} P_1(x,z)dx$$

hence from (25) for $i = 1$, using (32) and integration by part we have

$$P_1(z) = \frac{R_0(1-B_1^*)}{\{(1-\gamma)[(1-\theta)+\theta V^*]B_1^* + \gamma[(1-\theta)+\theta V^*]B_1^*B_2^*\} - z} \quad (33)$$

Similarly from (25) for $i = 2$, (30) and (33) we have

$$\begin{aligned} P_2(z) &= \int_0^{+\infty} P_2(x,z)dx \int_0^{+\infty} P_2(0,z)[1-B_2(x)]e^{-\lambda(1-z)x}dx \\ &= \int_0^{+\infty} \gamma P_1(0,z)B_1^*[1-B_2(x)]e^{-\lambda(1-z)x}dx \\ &= \frac{R_0\gamma B_1^*(1-B_2^*)}{\{(1-\gamma)[(1-\theta)+\theta V^*]B_1^* + \gamma[(1-\theta)+\theta V^*]B_1^*B_2^*\} - z} \end{aligned} \quad (34)$$

Finally from (26),(31) and with the same method we have

$$\begin{aligned} Q(z) &= \int_0^{+\infty} Q(x,z)dx = \\ &\frac{R_0\theta B_1^*[1-\gamma(1-B_2^*)][1-V^*]}{\{(1-\gamma)[(1-\theta)+\theta V^*]B_1^* + \gamma[(1-\theta)+\theta V^*]B_1^*B_2^*\} - z} \end{aligned} \quad (35)$$

REMARK 0.6 *The unknown constant R_0 can be determined by using normalizing condition (24) which is*

$$R_0 + P_1(1) + P_2(1) + Q(1) = 1 \quad (36)$$

from (33),(34) and (35) by using L'Hopital rule we have

$$P_1(1) = \frac{\lambda E(B_1)}{1-\lambda[E(B_1)+\gamma E(B_2)+\theta E(V)]}$$
$$P_2(1) = \frac{\gamma\lambda E(B_2)}{1-\lambda[E(B_1)+\gamma E(B_2)+\theta E(V)]}$$
$$Q(1) = \frac{\lambda\theta E(V)}{1-\lambda[E(B_1)+\gamma E(B_2)+\theta E(V)]}$$

hence by substituting in (36) and simplifying we have $R_0 = 1 - \rho$ *where*

$$\rho = \lambda[E(B_1) + \gamma E(B_2) + \theta E(V)] \tag{37}$$

R_0 *is the steady-state probability that the server is idle but available in the system, hence* $\rho < 1$ *can be the stability condition under which the steady state solution exist.*

COROLLARY 0.7 *From (33),(34) and (35) the PGF of the queue size distribution at a random epoch is*

$$\begin{aligned} P(z) &= P_1(z) + P_2(z) + Q(z) \\ &= (1-\rho)\{\frac{1 - B_1^*(1-\gamma) - \gamma B_1^* B_2^* + [\gamma\theta B_1^* B_2^* + (1-\gamma)\theta B_1^*](1-V^*)}{\{(1-\gamma)[(1-\theta)+\theta V^*]B_1^* + \gamma[(1-\theta)+\theta V^*]B_1^* B_2^*\} - z}\} \end{aligned} \tag{38}$$

and PGF of the queue size distribution at departure epoch(e.g see [3]) is

$$\begin{aligned} P_Q(z) &= R_0 + zP(z) \\ &= (1-\rho)\frac{(z-1)\{(1-\gamma)[(1-\theta)+\theta V^*]B_1^* + \gamma[(1-\theta)+\theta V^*]B_1^* B_2^*\}}{z - \{(1-\gamma)[(1-\theta)+\theta V^*]B_1^* + \gamma[(1-\theta)+\theta V^*]B_1^* B_2^*\}} \end{aligned} \tag{39}$$

if choose

$$G(z) = \{(1-\gamma)[(1-\theta)+\theta V^*]B_1^* + \gamma[(1-\theta)+\theta V^*]B_1^* B_2^*\} \tag{40}$$

then

$$P_Q(z) = (1-\rho)\frac{(z-1)G(z)}{z-G(z)} \tag{41}$$

which has familiar form by comparing with same formula in $M/G/1$ queue.

4 Mean queue size and other measures of system

Let L_Q be the mean number of customers in the queue (i.e mean queue size), then we have

$$L_Q = \frac{dP_Q(z)}{dz} \mid_{z=1} \tag{42}$$

PROPOSITION 0.8 *From (42) , using (39)and (37) we have*

$$L_Q = \rho + \frac{\lambda^2[E(B_1^2) + \gamma E(B_2^2) + 2\gamma E(B_1)E(B_2)]}{2(1-\rho)} \tag{43}$$

$$+\frac{\lambda^2[\theta E(V^2) + 2\theta E(B_1)E(V) + 2\gamma\theta E(B_2)E(V)]}{2(1-\rho)} \tag{44}$$

which by using (6),(7),(8) and (9) a brief form is

$$L_Q = \rho + \frac{\lambda^2 E(B^2)}{2(1-\rho)} \tag{45}$$

The formula (44) is interesting because is like as $M/G/1's$ formula for L_Q ; i.e Pollaczek-Khinchin formula.

Now for computing the mean response time of a test customer in this model , we use the approach of Kleinrock [6].Let $W_Q^*(s)$ be the LST of DF of waiting time of a tagged customer in this model.Then we have

$$W_Q^*(\lambda - \lambda z)B^*(\lambda - \lambda z) = P_Q(z) \tag{46}$$

where B^* is defined in (3).
If W_R denote the time interval from arrival time to the time when a tagged customer leaves the system after of completion service completion; i.e waiting time plus service time, then

$$W_R^*(s) = W_Q^*(s)B^*(s) \tag{47}$$

and mean response time of a tagged customer is

$$E(W_R) = -\frac{dW_R^*(s)}{ds} \mid_{s=0} \tag{48}$$

By substituting from (45) in (46) we have

$$W_R^*(s) = P_Q(1 - \frac{s}{\lambda}) \tag{49}$$

by using (39) and from (47) we have

$$E(W_R) = \frac{1}{\lambda} L_Q \tag{50}$$

Also the average system size is $L = L_Q + \rho$ where ρ is in (37).

5 General and Particular cases

5.1 General case

In this note we assume that the second phase of service has two cases where a tagged customer choose one of them with probability $p_i (i = 1, 2)$.
We can generalized this to k cases that a tagged customer can choose the case i with probability p_i for $i = 1, 2, ..., k$ and $p_1 + p_2 + ...p_k = 1$. In this version only $B_2^*(s)$ is changed .For $i = 1, 2, ..., k$ if S_i denote that the service time of $i'th$ case and $S_i^*(s)$ its LST, then

$$B_2^*(s) = \sum_{i=1}^{k} p_i S_i^*(s) \tag{51}$$

and

$$E(B_2) = \sum_{i=1}^{k} p_i E(S_i) \qquad , E(B_2^2) = \sum_{i=1}^{k} p_i E(S_i^2) \tag{52}$$

5.2 Particular case

I)If $\theta \to 0$; i.e there is no vacation in the system , then from (39) we have

$$P_Q(z) = (1-\rho)\frac{(z-1)\{(1-\gamma)B_1^* + \gamma B_1^* B_2^*\}}{z - \{(1-\gamma)B_1^* + \gamma B_1^* B_2^*\}} \tag{53}$$

where

$$\rho = \lambda[E(B_1) + \gamma E(B_2)] \tag{54}$$

and

$$L_Q = \rho + \frac{\lambda^2[E(B_1^2) + \gamma E(B_2^2) + 2\gamma E(B_1)E(B_2)]}{2(1-\rho)} \tag{55}$$

II)If $\gamma \to 1$; i.e a tagged customer surely accept the second phase of service , then from (39) we have

$$P_Q(z) = (1-\rho)\frac{(z-1)[(1-\theta) + \theta V^*]B_1^* B_2^*}{z - [(1-\theta) + \theta V^*]B_1^* B_2^*} \tag{56}$$

where

$$\rho = \lambda[E(B_1) + E(B_2) + \theta E(V)] \tag{57}$$

and

$$L_Q = \rho + \frac{\lambda^2[E(B_1^2) + E(B_2^2) + 2E(B_1)E(B_2)]}{2(1-\rho)} \tag{58}$$

$$+\frac{\lambda^2[2\theta[E(B_1) + E(B_2)]E(V) + \theta E(V^2)]}{2(1-\rho)} \tag{59}$$

which is the result of [8] if we set $p_1 = 1$ and $p_2 = 0$.
III) If $\gamma \to 0$; i.e there isn't second phase , or we have $M/G/1$ system with vacation , then from (39)

$$P_Q(z) = (1-\rho)\frac{(z-1)[(1-\theta) + \theta V^*]B_1^*}{z - [(1-\theta) + \theta V^*]B_1^*} \tag{60}$$

where

$$\rho = \lambda[E(B_1) + \theta E(V)] \tag{61}$$

and

$$L_Q = \rho + \frac{\lambda^2[E(B_1^2) + 2\theta E(B_1)E(V) + \theta E(V^2)]}{2(1-\rho)} \tag{62}$$

Also if $\theta \to 0$,then we have the classical $M/G/1$ queue and from (53)

$$P_Q(z) = (1-\rho)\frac{(z-1)B_1^*}{z - B_1^*} \tag{63}$$

where $\rho = \lambda E(B_1)$, and

$$L_Q = \rho + \frac{\lambda^2 E(B_1^2)}{2(1-\rho)} \tag{64}$$

which is the Pollaczek-Khinchine familiar formula.

References

[1] Choudhury,G. ,A Bath Arrival Queueing System with an additional Service Channel,Information and Managment Sciences , Vol 14 ,Number 2 ,pp.17-30,2003

[2] Doshi,B.T.,Queueing sysytem with vacation - A survey, Queueing System, vol 1 (1986)29-66

[3] Kashyap,B.R.K,Chaudhry.M.L.,An Introduction to Queueing Theory, Aarkay,Calcutta, India,1988

[4] Keilson,J. ; Servi,L.D., Oscillating random walk models for $G1/G/1$ vacation systems with Bernoulli schedules, Journal of Applied Probability 23 (1986), 790-802.

[5] Keilson,J. ; Servi,L.D., Dynamic of the M/G/1 vacation model, Operation Research ,Vol 35,No4,July-Agust 1987.

[6] Kleinrock,L., Queueing System;Wiley:New York,1975;Vol.1.

[7] Madan,K.C.,On a single server queue with two stage heterogenous service and deterministic server vacation,International Journal of System Sciences 32(2001)837-844

[8] Madan,K.C,Choudhury.G.,A single server queue with two phases of heterogenous service under Bernoulli schedule and a general vcation time,Information and Managment Science, vol 16,Number 2,pp1-16,2005

[9] Madan,K.C,Al-rawi,Al-naser,On $M^X/\binom{G_1}{G_2}/1/G(BS)/V_s$ Vacation Queue with two types of general heterogeneous service , Journal of Applied Mathematics and Decison Science 2005:(2005) 123-135

[10] Servi,L.D,Average delay approxmation of $M/G/1$ cyclic service queues with Bernoulli schedules,IEEE Selected Areas of Communications 4(6)(1986)813-820(Correction in 5(3)(1987)547).

[11] Takagi,H.,Queueing Analysis :A Foundation of Performance Evalution. vol , North Holland, Amesterdam, 1991

State Dependence and Feedback Effects in Individual Poverty Histories

Martin Biewen[‡]

Department of Economics, University of Mainz, Mainz, 55128, GERMANY

Abstract: This paper argues that the strict exogeneity assumption which is usually invoked in econometric models of state dependence is inappropriate in the context of poverty dynamics. The reason is that poverty may have feedback effects to important determinants of contemporaneous poverty status such as employment status and household composition. The paper compares different conventional models of state dependence and proposes a model that incorporates feedback effects.

Keywords: Poverty persistence, Dynamic Binary Response Models, Correlated Random Effects, Initital Conditions, Strict Exogeneity.

1 Introduction

For policy purposes, it is of considerable interest whether there are state dependence effects of poverty, i.e. whether experiencing poverty in one period increases the risk of being poor in future periods ('poverty trap'). In order to rule out spurious state dependence effects (individuals who were poor in past periods because of unfavourable characteristics are more likely to be poor in future periods if the underlying characteristics do not change) state dependence effects are usually investigated using dynamic binary response models with observed and unobserved heterogeneity. However, most of these models are based on the assumption of strict exogeneity, which rules out feedback effects from the dependent variable (here the poverty status) on the explanatory variables. This paper argues that strict exogeneity of explanatory variables is violated in the poverty context, as important variables determining contemporaneous poverty risk, in particular employment status and household composition, are likely to be influenced by past poverty outcomes. Using an idea due to Wooldridge (2000), the paper proposes a model that explicitly allows for feedback effects from past poverty

[‡]biewen@uni-mainz.de

status to some of the variables that determine contemporaneous poverty risk. Based on data from the German Socio-Economic Panel (GSOEP), estimates from this model are contrasted to estimates from conventional dynamic binary response models based on the strict exogeneity assumption (dynamic correlated random-effects probit and fixed-effects logit models) as well as to estimates from a pooled model. The results suggest that there are indeed feedback effects and that failure to take them into account leads to biased estimates of the state dependence effect. Correctly taking into account feedback effects, and controlling for observed and unobserved heterogeneity, the results suggest state dependence effects of over 30 percentage points.

2 Econometric models

2.1 distributions

The first approach to modelling state dependence in individual poverty status considered here is a dynamic random effects probit model (see e.g. Wooldridge (2002)). If the dummy variable y_{it} denotes poverty status of individual $i = 1 \dots N$ in period $t = 1 \dots T$ then

$$y_{it} = 1\{\theta_1 y_{it-1} + \theta_2 z_{it} + c_i + e_{it} \geq 0\} \tag{1}$$

($1\{\cdot\}$ is the indicator function) describes the evolution of poverty conditional on i's poverty status in the previous period, a vector of exogenous variables z_{it} and two unobservables c_i and e_{it}. The state dependent effect is the effect of past poverty y_{it-1} on current poverty y_{it}. The unobservables c_i may be related to the observables and the initial condition y_{i0} through

$$c_i = \alpha_0 + \alpha_1 y_{i0} + \alpha_2 \bar{z}_i + \alpha_3 y_{i0} \bar{z}_i + a_i, \tag{2}$$

where $\bar{z}_i = T^{-1} \sum_{t=1}^{T} z_{it}$ denotes the time-average of the observed variables z_{it} and $a_i \sim \mathcal{N}(0, \sigma_a^2)$. The idiosyncratic errors e_{it} are assumed to be standard normally distributed.

In order to estimate the model by conditional maximum-likelihood methods, one has to make the crucial assumption that

$$P(y_{it} = 1 | z_i, y_{it-1}, y_{it-2}, \dots, y_{i0}, c_i) = P(y_{it} = 1 | z_{it}, y_{it-1}, c_i) \tag{3}$$

(strict exogeneity of z_i, where z_i summarizes the exogenous information $(z_{i1}, \dots, z_{iT})$). This assumption means that conditional on poverty status

in the previous period and conditional on the unobserved individual-specific characteristics c_i, poverty in period t must not be related to the value of the explanatory variables in past or in future periods. This requires in particular that there must not be any feedback from poverty in period t to future values of the explanatory variables. In the given context, this is likely to be unrealistic, as experiencing poverty in one period may possibly influence employment decisions or household composition in future periods. The random effects model can be estimated under the strict exogeneity assumption by using numerical integration methods described in Butler/Moffit (1982). For more details on the random effects model and the models described in the following sections, see Biewen (2008).

2.2 Fixed effects estimation

One drawback of the correlated random effects model of the preceding section is that it assumes a rather specific relationship between the explanatory variables z_i and the unobserved effect c_i. Although this relationship does not necessarily have to be specified as above, i.e. as a linear function of the time averages $\bar{z}_i$ (this is usually done to save degrees of freedom), it could be restrictive in the given context. In order to avoid this restriction, a fixed effects logit approach can be employed (see Honoré/Kyriazidou (2000)).

For the dynamic fixed effects logit model, it is assumed that

$$y_{it} = 1\left\{\theta_1 y_{it-1} + \theta_2 z_{it} + c_i + e_{it} \geq 0\right\} \tag{4}$$

as before, but with

$$\begin{aligned} P\left(y_{it} = 1 | z_i, y_{it-1}, c_i\right) &= P\left(y_{it} = 1 | z_{it}, y_{it-1}, c_i\right) \\ &= \frac{\exp\left(\theta_1 y_{it-1} + \theta_2 z_{it} + c_i\right)}{1 + \exp\left(\theta_1 y_{it-1} + \theta_2 z_{it} + c_i\right)}. \end{aligned} \tag{5}$$

The first equation in (5) is the strict exogeneity assumption again, while the second equation implies that the e_{it}'s follow an i.i.d. logistic distribution, independent of z_i, c_i and y_{i0}. In addition, it is assumed that the initial observation y_{i0} has an arbitrary probability distribution given z_i and c_i

$$P\left(y_{i0} = 1 | z_i, c_i\right) = p_0\left(z_i, c_i\right). \tag{6}$$

Honoré/Kyriazidou (2000) showed how a sophisticated conditioning scheme can be used to estimate this model.

2.3 Pooled estimation

Both the random effects and the fixed effects approach are based on the strict exogeneity assumption, which may be questionable in the given context. A simple but inefficient alternative that avoids the strict exogeneity assumption is a pooled estimator (see e.g Wooldridge (2002)). Let

$$y_{it} = 1\left\{\theta_1 y_{it-1} + \theta_2 z_{it} + c_i + e_{it} \geq 0\right\} \tag{7}$$

with

$$c_i = \alpha_0 + \alpha_1 y_{i0} + \alpha_2 \bar{z}_i + \alpha_3 y_{i0}\bar{z}_i + a_i, \tag{8}$$

and $e_{it} \sim \mathcal{N}(0,1), a_i \sim \mathcal{N}(0,\sigma_a^2)$ as in the random effects model. The pooled approach uses a simple probit model applied to the pooled observations, ignoring the panel structure of the data.

2.4 A Model with feedback effects

The aim of this section is to develop an econometric model which explicitly allows for feedback effects from poverty status to future employment decisions and household composition. Let y_{it} denote individual poverty status as before, and let w_{it} and v_{it} indicate whether individual i is employed and whether he (the empirical analysis will focus on prime-age men only) is living together with other persons. This will be a partner and possibly children in most cases. Then, under assumptions analogous to those made in the case of the random effects probit model, the joint density of $y_{i1},\ldots,y_{iT}, w_{i1},\ldots,w_{iT}, v_{i1},\ldots,v_{iT}$ given exogenous variables z_i, initial

values y_{i0}, w_{i0}, v_{i0} and an individual-specific effect c_i can be written as

$$\begin{aligned}
&f\left(y_{i1},\ldots,y_{iT},w_{i1},\ldots,w_{iT},v_{i1},\ldots,v_{iT}|z_i,y_{i0},w_{i0},v_{i0},c_i,\theta,\gamma,\beta\right)\\
&= \prod_{t=1}^{T} f\left(y_{it}|z_{it},w_{it},v_{it},y_{it-1},w_{it-1},v_{it-1},c_i,\theta\right)\\
&\quad \cdot g\left(w_{it},v_{it}|z_{it},y_{it-1},w_{it-1},v_{it-1},c_i,\gamma,\beta\right)\\
&= \prod_{t=1}^{T} \Phi((2y_{it}-1)(\theta_1 z_{it}+\theta_2 w_{it}+\theta_3 v_{it}\\
&\quad +\theta_4 y_{it-1}+\theta_5 w_{it-1}+\theta_6 v_{it-1}+c_i).\\
&\qquad \cdot\Phi_2\left[(2w_{it}-1)(\gamma_1 z_{it}+\gamma_2 y_{it-1}+\gamma_3 w_{it-1}+\gamma_4 v_{it-1}+\gamma_5 c_i),\right.\\
&\qquad .(2v_{it}-1)(\beta_1 z_{it}+\beta_2 y_{it-1}+\beta_3 w_{it-1}+\beta_4 v_{it-1}+\beta_5 c_i)\\
&\qquad\qquad \left..(2w_{it}-1)(2v_{it}-1)\rho\right], \qquad (9)
\end{aligned}$$

where $\Phi_2[\cdot,\cdot,\rho]$ denotes the cumulative distribution function of a bivariate standard normal distribution with correlation coefficient ρ. The individual-specific effect now also includes the influence of the initial values of employment status and living arrangements, i.e.

$$c_i = \alpha_0+\alpha_1 y_{i0}+\alpha_2 w_{i0}+\alpha_3 v_{i0}+\alpha_4 \bar{z}_i+\alpha_5 y_{i0}\bar{z}_i+\alpha_6 w_{i0}\bar{z}_i+\alpha_7 v_{i0}\bar{z}_i+a_i. \quad (10)$$

The model endogenizes employment status w_{it} and household composition v_{it} in the form of a bivariate probit model upon the outcomes of which poverty status is conditioned. Note that the additional equations for employment status (9) and household composition (9) also include lagged poverty experience, capturing possible feedback effects from poverty to employment status and to whether the individual lives alone. Such feedback effects may reflect detrimental effects of poverty on the morale of the individual, leading to lower employment probabilities, or on the stability of marriages or cohabitative relationships, increasing the probability of a household split.

3 Empirical results

The empirical analysis in this section is based on a balanced sample taken from the German Socio-Economic Panel (GSOEP). All estimates make use

of longitudinal sample weights in order to take account of unequal probability sampling and panel attrition (for a detailed description of the data used here, see Biewen (2008)). Table 1 shows the estimates for the random effects model, the fixed effects model and the pooled model. All three models suggest highly significant state dependence effects. However, the pooled estimate (which is robust to violations of the strict exogeneity assumption) is different from the random effect estimate (which uses the strict exogeneity assumption). Because of the unmodelled error variance, the fixed effects estimates cannot be directly compared to those of the other two methods. However the relative magnitudes of the estimates seem similar to those in random effects model.

Table 2 shows the estimates of the model with feedback. Column 1 of the table displays the coefficients for the poverty equation. The results also suggest a sizeable and significant state dependence effect, the magnitude of which even exceeded that of employment. There are also strong and significant poverty-reducing effects of employment and of living together with others, which are of a similar magnitude as in the other models. The effects for the other variables show very similar patterns as in the previous models. Columns 2 and 3 show the equations for employment status and household composition. The fact that lagged poverty significantly reduces the employment probability of a given period constitutes a clear violation of the strict exogeneity assumption. In addition, the results in the third column show a significant feedback effect of poverty on household composition, providing another violation of the strict exogeneity assumption.

It is interesting to express these effects in differences of percentage points. Table 3 presents the most interesting estimates expressed as average partial effects. According to the random effects estimates, being poor increased the poverty risk in the next period by roughly 10 percentage points, holding everything else constant. By contrast, the pooled model estimated this effect as approximately 28 percentage points. The pooled estimate was very similar to the estimate from the feedback model, which resulted in 31 percentage points. The fact that the pooled model and the model with feedback yielded very similar results and that these results were different from those of the potentially misspecified random effects model provide another indication that the random effects estimates might be biased. Note

that, due to the unmodelled relationship between the explanatory variables and the unobserved individual-specific effect, no direct estimate of the magnitude of state dependence is available for the fixed effects logit model.

4 Conclusion

This paper highlighted the role of the strict exogeneity assumption in estimating dynamic binary response models with unobserved heterogeneity in the context of poverty dynamics. The results show that explanatory variables in a dynamic poverty model are not strictly exogenous and that feedback effects from past poverty status to future values of these explanatory variables should be explicitly modeled.

Tables

Table 1. Single equation dynamic binary response models for individual poverty status (standard errors in parentheses)

	random effects[a]		fixed effects[b]		pooled model[c]	
lagged poverty status	.6946757	(.077607)	1.594907	(.1836328)	1.470528	(.0732569)
employed	-.8350687	(.0785319)	-1.525076	(.3676939)	-.8661508	(.0864569)
living w. others	-.3014123	(.067902)	-.5946901	(.430487)	-.3207881	(.059306)
26 - 35 years	-.0120356	(.2566974)	.3904746	(1.047917)	.024477	(.2773118)
36 - 45 years	-.2133647	(.173677)	-.6725629	(.8102958)	-.1818307	(.1794543)
46 - 55 years	-.1204192	(.1188686)	-1.061619	(.6591849)	-.1022485	(.1212862)
Initial conditions, correlated part of random effect (time averages), and interactions						
poverty status in 2000	1.038701	(.1649636)			.625276	(.1371623)
poverty 2000 × employed	-.482127	(.2123242)			-.3348532	(.1856247)
poverty 2000 × disability	-.3205893	(.2242235)			-.3099992	(.1925815)
employed × disability	.7727144	(.2600542)			.6448195	(.2170425)
university × East Germany	-.4025646	(.1880236)			-.376984	(.1639195)
employed	-.5877584	(.1493492)			-.3992091	(.1420373)
university	-.2863257	(.1228917)			-.2297788	(.1102328)
Abitur or *Lehre*[d]	-.1111345	(.0970825)			-.0855922	(.0830449)
26 - 35 years	.429217	(.2791804)			.338345	(.2963396)
36 - 45 years	.7174684	(.2201027)			.6213021	(.2181563)
46 - 55 years	.5334242	(.164469)			.4731471	(.1589938)
disability	-.4243889	(.1704368)			-.3545088	(.1421246)
non-German nationality	.5023872	(.132991)			.4488087	(.109134)
North Germany	.1557889	(.1091984)			.1124901	(.0959621)
Middle Germany	.2733224	(.1323344)			.2399854	(.1122583)
South Germany	-.0272791	(.1098548)			-.022885	(.0978533)
Berlin Germany	.1464195	(.1581653)			.1261594	(.1428351)
East Germany	.6134964	(.1000315)			.518853	(.0878258)
constant	-.9848526	(.1607587)			-1.096977	(.140579)

Source: GSOEP, 2000 - 2006, balanced sample of 3952 prime-aged men (26 to 65 years), longitud. weighted

[a] Coefficients were rescaled by $(1 + \hat{\sigma}_a^2)^{-1/2}$ to ensure comparability with pooled model.
The estimate of σ_a was $\hat{\sigma}_a = .8098921$ with estimated standard error .0682928.

[b] Honoré/Kyriazidou (2000)

[c] Standard errors account for clustering of observations at individual level

[d] High-school degree and/or vocational training

Table 2. Model of poverty status with endogenous employment status, endogenous household composition, and feedback effects (standard errors in parentheses)

	poverty status		employed		living w. others	
lagged poverty status	1.62686	(.0794945)	-.6662807	(.0948899)	-.5848139	(.1088997)
employed	-1.498639	(.0807484)				
lagged employment status			1.914135	(.1223355)	-.4540126	(.1120635)
living w. others	-.7894862	(.0848741)				
living w. others, lagged			-.5243237	(.1306013)	3.130104	(.1035554)
university degree			1.65258	(.3829242)	.6417651	(.2047935)
Abitur or *Lehre*[a]			.6966695	(.2192367)	.2394706	(.1373028)
26 - 35 years	-.0269727	(.1798619)	.3181698	(.2432962)	-.6049053	(.1685352)
36 - 45 years	-.0590165	(.1405056)	.223066	(.1889031)	-.6094557	(.1428444)
46 - 55 years	.0417365	(.0954722)	.3614941	(.0925828)	-.3502511	(.0957454)
disability			-.3568577	(.1152455)	-.0062215	(.1001874)
non-German nationality			-.5916449	(.2224929)	.2401081	(.1530035)
North Germany			.0118745	(.1802855)	-.01512	(.1436889)
Middle Germany			-.3096945	(.1877033)	-.2293624	(.1519916)
South Germany			.1269503	(.195484)	-.0535384	(.1346134)
Berlin Germany			-.0583541	(.2640176)	-.2982236	(.2029335)
East Germany			-.7315598	(.2580738)	-.4860962	(.1351701)
regional unemployment rate			-.0280906	(.0158297)		
linear time trend					-.0550824	(.0191758)
Initial conditions, correlated part of random effect (time averages), and interactions						
poverty status in 2000	.2705542	(.1078242)				
employed in 2000	.4600827	(.0757733)				
living w. others in 2000	.3912583	(.0769544)				
poverty 2000 × employed 2000	-.2889731	(.0950686)				
employed 2000 × university	-.2087634	(.0805644)				
employed 2000 × disability	.2075472	(.0944432)				
university degree	-.3490871	(.1286945)				
Abitur or *Lehre*[a]	-.1995001	(.0802878)				
26 - 35 years	.4882873	(.1619074)				
36 - 45 years	.4266077	(.1336338)				
46 - 55 years	.2955451	(.081639)				
disability	-.2428803	(.0914465)				
non-German nationality	.3836259	(.1191499)				
North Germany	.0422811	(.099706)				
Middle Germany	.1779878	(.1187321)				
South Germany	-.0475139	(.0929573)				
Berlin Germany	.142626	(.1331886)				
East Germany	.3841361	(.0955914)				
constant	-1.06556	(.1883359)				
σ_a	.259985	(.0324312)				
correlation coefficient ρ			-.25773	(.064024)		
loading factor γ_5			-1.722195	(.4193814)		
loading factor β_5					1.183929	(.2009223)

Source: GSOEP, 2000 - 2006, balanced sample of 3952 prime-aged men (26 to 65 years), longitudinally weighted

[a] High-school degree and/or vocational training

Table 3. State dependence and feedback effects (average partial effects, pooled sample 2001 - 2006)

	poverty status	employed	living w. others
Random effects probit			
lagged poverty status	.09927558[a]		
Fixed effects logit			
lagged poverty status	- [b]		
Pooled probit			
lagged poverty status	.27835979[a]		
Model with feedback effects			
lagged poverty status	.31217871[a]	-.09332394[a]	-.05082249[a]
employed	-.23509029[a]		
lagged employment status		.33005296[a]	-.03193042[a]
living w. others	-.08998[a]		
living w. others, lagged		-.05506463[a]	.80162281[a]

Source: GSOEP, 2000 - 2006, balanced sample of 3952 prime-aged men (26 to 65 years), longitudinally weighted
[a] Underlying coefficient statistically significant at the 5 percent level
[b] Average partial effect cannot be calculated

References

[1] Biewen, M. (2008): Measuring state dependence in individual poverty histories when there is feedback to employment status and household composition. Submitted for publication.

[2] Butler, J.S, Moffit, R. (1982). A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model. *Econometrica*, **50**, 761 - 764.

[3] Wooldridge, J.M. (2000). A framework for estimating dynamic, unobserved effects panel data models with possible feedback to future explanatory variables. *Economics Letters* **68**, 245 - 250.

[4] Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA. MIT-Press.

[5] Honoré, B.E., Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables, *Econometrica* **68**, 839 - 874.

The Bayesian Testing for Weighted Distributions

R. Chinipardaz[§], and S. M. R. Alavi

Department of Statistics, University of Shahid Chamran

Abstract: This paper is concerned with the bayesian testing in weighted distributions. When the observations are subjected to a weight function the testing problem may be changed dramatically. In this paper bayes factor for weighted distributions is obtained for two competition hypotheses when the hypotheses are simple against simple, one sided and simple against two-sided hypotheses. Some common weight functions are examined for calculating the bayes factor in testing the parameter of exponential distribution and compared with unweighted version of the bayes factor. The lower bound of bayes factor when the prior distribution is belonged to a reasonable class of distributions for weighted distributions are derived and are compared with classical bayes factor.

Keywords: Weighted distributions; Bayesian analysis; Significance testing.

1 Introduction

Suppose that the real-valued random variable X is distributed to the density function $f(x \mid \theta)$; $\theta \in \Theta$ where Θ is some finite or infinite interval in R^1.
Now, consider $X = x$ is recorded or entered in the sample with a weight function, $w(x, \beta)$. In this case x is not an observation on random variable X, but on the weighted version of random variable X, X^w having a probability density function, $f^w(x \mid \theta)$ given by

$$f^w(x \mid \theta) = \frac{w(x, \beta) f(x \mid \theta)}{E_\theta[w(X, \beta)]} \tag{1}$$

where $E_\theta[w(X, \beta)] = \int_R w(x, \beta) f(x \mid \theta) dx$ is normalizing factor. The weight function, $w(x, \beta)$, is required to be nonnegative and

$$0 < E_\theta[w(X, \beta)] < \infty$$

for all $\theta \in \Theta$. $f^w(x \mid \theta)$ is said to be weighted density function.

[§]Invited speaker

The weighted distributions have been used as a tool in the selection of appropriate models for data sampled without a proper frame. Such situations are found in many fields. For example in publication biased the higher significance studied have been greater chance of being published (Iyengar and Greenhouse, 1988, Silliman, 1987). In Meta analysis one interests to study of the results of statistical analysis for the propose of drawing general calculasions (see Glass, 1976, Fleiss, 1993). Such other examples can be found in Oil Discovery (West, 1996) and Line Transact Sampling. More examples are also given in Patil (2002).
The concept of weighted distributions was originally introduced by Fisher (1934) to the study of effect of methods of ascertainment upon estimation of frequencies. However, it was Rao (1965, 1985) who presented a unified theory of weighted distributions. Patil and Taillie (1987) calculated the Fisher information for certain exponential family of weighted distributions (see also Bayarri and DeGroot, 1978a, 1978b and Bayarri et al, 1987). In all last four references the Fisher information derived from X^w with the derived from X. The weighted distributions have also been studied in bayesian analysis. For the finite population models West (1994, 1996) considered known or parametric forms of $w(x, \beta)$. The parametric forms of $w(x, \beta)$ have also been investigated and employed for selection models in infinite population models by Bayarri end DeGroot (1987, 1988, 1989, 1990, 1992)
The robustness of posterior probabilities is investigated by Bayarri and Berger (1998). They considered two nonparametric classes of $w(x, \beta)$; the class of bounded functions from above and below by two weighted functions and the class with additional condition that $w(x, \beta)$ be nondecreasing function.
The problem of comparison between two bayes factors in different models also considered by Lorose and Dey (1996, 1998). The comparison was between bayes factors derived from weighted distributions and unweighted distributions.

This study is devoted to bayesian testing problem in weighted distributions. Motivation of the study is that the information provided by bayesian analysis might be different for the weighted distributions with respect to

unweighted distributions. The bayes factor is considered as a criteria for supporting data from one hypothesis from other competing hypotheses. Another criteria which can also be used is the posterior probability of the hypothesis is true. However, bayes factors are usually preferred because they are independent from prior distributions which are based on subjective approach.

This paper includes five sections; in section two the bayesian analysis for weighted random variables are discussed. In section three we derive bayesian testing in weighted random variables for simple against simple hypotheses, one-sided against one-sided hypotheses and point null hypotheses against two-sided hypotheses. The bayes factor in weighted version of X is obtained and compared with bayes factor in unweighted case. Section four is devoted to exponential distribution. The theory based on section three followed for this distribution using some common weighted functions. Finally, in section five lower bound of weighted bayes factor, when the prior distribution is belonged to a class of distributions, is derived and compared with unweighted bayes factor.

2 Bayesian analysis on weighted distributions

Suppose that $f(x\,|\theta)$ is the probability density function of X and $w(x,\beta)$ is a weighted function which is supposed to be independent of θ with $E_\theta[w(X,\beta)] < \infty$. The weighted density function in given in (1). Now, let the prior distribution of θ on Θ is $g(\theta)$. The weighted posterior distribution of θ conditional on x is

$$\pi^w(\theta\,|x) = \frac{f^w(x\,|\theta)g(\theta)}{\int_\Theta f^w(x\,|\theta)g(\theta)d\theta}$$

Replacing $f^w(x,\beta)$ from (1) is

$$\begin{aligned}\pi^w(\theta\,|x) &= \frac{f^w(x\,|\theta)g(\theta)}{\int_\Theta f^w(x\,|\theta)g(\theta)d\theta}\\ &= \frac{w(x,\beta)f(x\,|\theta)g(\theta)}{E_\theta[w(X,\beta)]}\Big/\int_\Theta \frac{w(x,\beta)f(x\,|\theta)}{E_\theta[w(X,\beta)]}g(\theta)d\theta\\ &= \frac{f(x\,|\theta)g(\theta)}{E_\theta[w(X,\beta)]\int_\Theta f(x\,|\theta)E_\theta^{-1}[w(X,\beta)]g(\theta)d\theta}. \qquad (2)\end{aligned}$$

or

$$\pi^w(\theta\,|x) = \frac{f^w(x\,|\theta)g(\theta)}{m^w(x)} \qquad (3)$$

where

$$\begin{aligned}m^w(x) &= \int_\Theta f^w(x|\theta)g(\theta)d\theta \qquad (4)\\ &= \int_\Theta \frac{w(x,\beta)f(x|\theta)}{E_\theta[w(X,\beta)]}g(\theta)d\theta\end{aligned}$$

is the weighted marginal distribution of the data.
As an example consider exponential distribution

$$f(x|\theta) = \theta e^{-\theta x} \quad \theta > 0,\ x \geq 0 \qquad (5)$$

One candidate for prior distribution is

$$\pi_\lambda(\theta) = \lambda e^{-\lambda\theta}, \quad \lambda > 0. \qquad (6)$$

Now, suppose that the weight function is length biased function, $w(x,\beta) = x$,

$$\begin{aligned}\pi^w(\theta|x) &= \frac{\theta e^{-\theta x}\lambda e^{-\lambda\theta}}{E_\theta[X]\int_0^\infty \frac{\theta e^{-\theta x}}{E_\theta[X]}\lambda e^{-\lambda\theta}d\theta} \qquad (7)\\ &= \frac{\theta^2 e^{-\theta(x+\lambda)}}{\int_0^\infty \theta^2 e^{-\theta(x+\lambda)}d\theta}\\ &= \frac{(x+\lambda)^3}{2}\theta^2 e^{-\theta(x+\lambda)} = \Gamma_{(3,x+\lambda)}(\theta).\end{aligned}$$

It is easy to show that the unweighted posterior distribution, $\pi(\theta|x)$ is $\Gamma_{(2,x+\lambda)}(\theta)$. Using the ratio $\frac{\pi^w(\theta|x)}{\pi(\theta|x)} = \left(\frac{x+\lambda}{2}\right)\theta$ shows that posterior probability of θ given the data is larger than unweighted version if $(x+\lambda) > \frac{2}{\theta}$ and is smaller for $(x+\lambda) < \frac{2}{\theta}$.

3 Bayesian Testing in Weighted Distributions

Suppose x is an observed value from a population with probability density function $f(x|\theta)$, $\theta \in \Theta \subseteq R$. We wand to test $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$ where $\Theta_0 \cap \Theta_1 = \emptyset$ and $\Theta_0 \cup \Theta_1 = \Theta$. In general of bayesian testing, the prior distribution, $g(\theta)$, is defined as

$$g(\theta) = \begin{cases} \pi_0 g_0(\theta), & \theta \in \Theta_0 \\ (1-\pi_0) g_1(\theta), & \theta \in \Theta_1 \end{cases}$$

where π_0 and $(1-\pi_0)$ are the prior probabilities of H_0 and H_1 to be true, respectively. $g_0(\theta)$ and $g_1(\theta)$ describe how the prior mass is spread out over the two hypotheses. Now, define the weighted marginal function of Θ_j as

$$\begin{aligned} m_j^w(x) &= \int_{\Theta_j} f^w(x \,|\, \theta) g_j(\theta) d\theta \\ &= \int_{\Theta_j} \frac{w(x,\beta) f(x \,|\, \theta)}{E_\theta[w(X,\beta)]} g_j(\theta) d\theta. \end{aligned}$$

Now, consider the testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$. In this case the prior distribution is

$$g(\theta) = \begin{cases} \pi_0, & \theta = \theta_0 \\ (1-\pi_0), & \theta = \theta_1 \end{cases}$$

and the posterior probability of H_0 being true is given by

$$\begin{aligned} \pi^w(H_0 \,|\, x) &= \frac{\pi_0 f^w(x \,|\, \theta_0)}{\pi_0 f^w(x \,|\, \theta_0) + (1-\pi_0) f^w(x \,|\, \theta_1)} \\ &= \frac{\pi_0 \frac{f(x \,|\, \theta_0)}{E_{\theta_0}[w(X,\beta)]}}{\pi_0 \frac{f(x \,|\, \theta_0)}{E_{\theta_0}[w(X,\beta)]} + (1-\pi_0) \frac{f(x \,|\, \theta_1)}{E_{\theta_1}[w(X,\beta)]}} \\ &= \left[1 + \frac{1-\pi_0}{\pi_0} \frac{f(x \,|\, \theta_1)}{f(x \,|\, \theta_0)} \left(\frac{E_{\theta_0}[(w(X,\beta))]}{E_{\theta_1}[(w(X,\beta))]}\right)\right]^{-1} \end{aligned}$$

and with the similar way

$$\begin{aligned}\pi^w(H_1 \mid x) &= \frac{(1-\pi_0)f^w(x \mid \theta_1)}{\pi_0 f^w(x \mid \theta_0) + (1-\pi_0)f^w(x \mid \theta_1)} \\ &= \left[1 + \frac{\pi_0}{1-\pi_0}\frac{f(x \mid \theta_0)}{f(x \mid \theta_1)}\left(\frac{E_{\theta_1}[w(X,\beta)]}{E_{\theta_0}[w(X,\beta)]}\right)\right]^{-1}\end{aligned}$$

and therefore the weighted Bayes factor in favour of H_0 can be obtained as

$$\begin{aligned}B_{01}^w &= \frac{\pi^w(H_0 \mid x)/\pi^w(H_1 \mid x)}{\pi_0/(1-\pi_0)} = \frac{f(x \mid \theta_0)}{f(x \mid \theta_1)}\left(\frac{E_{\theta_1}[w(X,\beta)]}{E_{\theta_0}[w(X,\beta)]}\right) \\ &= B_{01}\left(\frac{E_{\theta_1}[w(X,\beta)]}{E_{\theta_0}[w(X,\beta)]}\right)\end{aligned}$$

where B_{01} is the unweighted bayes factor in favour of H_0 with respect to H_1.

For Bayesian testing of $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ it is denoted $0 < \pi_0 < 1$ as the probability of $H_0 : \theta = \theta_0$ and $1 - \pi_0$ as the probability of $H_1 : \theta \neq \theta_0$ and it is supposed the mass on H_1 is spread out according to the density $g_1(\theta)$. Following this

$$\begin{aligned}\pi^w(H_0 \mid x) &= \frac{\pi_0 f^w(x \mid \theta_0)}{\pi_0 f^w(x \mid \theta_0) + (1-\pi_0)\int_{\theta\neq\theta_0} f^w(x \mid \theta)g_1(\theta)d\theta} \\ &= \frac{\pi_0 f^w(x \mid \theta_0)}{\pi_0 f^w(x \mid \theta_0) + (1-\pi_0)m_1^w(x)} \\ &= \left[1 + \frac{1-\pi_0}{\pi_0}\frac{E_{\theta_0}[w(X,\beta)]m_1^w(x)}{f(x \mid \theta_0)}\right]^{-1}\end{aligned}$$

where $m_1^w(x) = \int_{\theta\neq\theta_0} \frac{w(x,\beta)f(x \mid \theta)}{E_\theta[w(X,\beta)]} g_1(\theta)d\theta$. $\pi^w(H_1 \mid x)$ can be obtained as

$$\pi^w(H_1 \mid x) = \left[1 + \frac{\pi_0}{1-\pi_0}\frac{f(x \mid \theta_0)}{E_{\theta_0}[w(X,\beta)]m_1^w(x)}\right]^{-1}.$$

Therefore,

$$\begin{aligned} B_{01}^{w}(x) &= \frac{f(x\,|\,\theta_0)}{E_{\theta_0}[w(X,\beta)]\int_{\theta\neq\theta_0}\frac{f(x\,|\,\theta)}{E_{\theta}[w(X,\beta)]}g_1(\theta)d\theta} \\ &= B_{01}\frac{\int_{\theta\neq\theta_0}f(x|\,\theta)g_1(\theta)d\theta}{E_{\theta_0}[w(X,\beta)]\int_{\theta\neq\theta_0}\frac{f(x\,|\,\theta)}{E_{\theta}[w(X,\beta)]}g_1(\theta)d\theta} \\ &= B_{01}\frac{m_1(x)}{E_{\theta_0}[w(X,\beta)]\int_{\theta\neq\theta_0}\frac{f(x\,|\,\theta)}{E_{\theta}[w(X,\beta)]}g_1(\theta)d\theta}. \end{aligned}$$

For testing the hypothesis $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$ the prior distribution is given in

$$g(\theta) = \begin{cases} \pi_0 g_0(\theta), & \theta \leq \theta_0 \\ (1-\pi_0)g_1(\theta), & \theta > \theta_0 \end{cases}$$

Therefore,

$$\begin{aligned} P^{w}(\theta \leq \theta_0|\,x) &= \int_{\theta\leq\theta_0}\pi^{w}(\theta|\,x)d\theta \\ &= \frac{\pi_0\int_{\theta\leq\theta_0}f^{w}(x\,|\,\theta)g_0(\theta)d\theta}{\pi_0\int_{\theta\leq\theta_0}f^{w}(x\,|\,\theta)g_0(\theta)d\theta + (1-\pi_0)\int_{\theta>\theta_0}f^{w}(x\,|\,\theta)g_1(\theta)d\theta} \\ &= \frac{\pi_0 m_0^{w}(x)}{\pi_0 m_0^{w}(x) + (1-\pi_0)m_1^{w}(x)} \end{aligned}$$

and the weighted Bayes factor is given by

$$\begin{aligned} B_{01}^{w}(x) &= \frac{m_0^{w}(x)}{m_1^{w}(x)} = \frac{\int_{\theta\leq\theta_0}f^{w}(x\,|\,\theta)g_0(\theta)d\theta}{\int_{\theta>\theta_0}f^{w}(x\,|\,\theta)g_1(\theta)d\theta} \\ &= \frac{\int_{\theta\leq\theta_0}\frac{f(x\,|\,\theta)}{E_{\theta}[w(X,\beta)]}g_0(\theta)d\theta}{\int_{\theta>\theta_0}\frac{f(x\,|\,\theta)}{E_{\theta}[w(X,\beta)]}g_1(\theta)d\theta} \end{aligned}$$

4 Testing for exponential distribution

4.1 Simple against simple hypothesis

Table 1: The weighted bayes factor for $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ in exponential distribution for various weight function $\pi_0 = \frac{1}{2}$

Weight function	$w(x,\beta)$	Weighted bayes factor	The supporting rate of H_0	$B_{01}^w > B_{01}$
Unweighted	c	$\left(\frac{\theta_0}{\theta_1}\right) e^{(\theta_1-\theta_0)x}$	-	1
Size biased	x	$\left(\frac{\theta_0}{\theta_1}\right) B_{01}$	$\frac{\theta_0}{\theta_1}$	$\theta_0 > \theta_1$
Length biased	x^r	$\left(\frac{\theta_0}{\theta_1}\right)^r B_{01}$	$\left(\frac{\theta_0}{\theta_1}\right)^r$	$\theta_0 > \theta_1$
Exponential	$e^{\tau x}$	$\frac{\theta_0-\tau}{\theta_1-\tau}\left(\frac{\theta_0}{\theta_1}\right)^{r-2} B_{01}$	$\frac{\theta_1(\theta_0-\tau)}{\theta_0(\theta_1-\tau)}$	$\theta_0 < \theta_1$
Set Selection(I)	$X > a$	$e^{a(\theta_0-\theta_1)} B_{01}$	$e^{a(\theta_0-\theta_1)}$	$\theta_0 > \theta_1$
Set Selection(II)	$X < b$	$\frac{1-e^{-b\theta_1}}{1-e^{-b\theta_0}} B_{01}$	$\frac{1-e^{-b\theta_1}}{1-e^{-b\theta_0}}$	$\theta_0 < \theta_1$
Set Selection(III)	$c < X < d$	$\frac{e^{-c\theta_1}-e^{-d\theta_1}}{e^{-c\theta_0}-e^{-d\theta_0}} B_{01}$	$\frac{e^{-c\theta_1}-e^{-d\theta_1}}{e^{-c\theta_0}-e^{-d\theta_0}}$	always

As an example consider the exponential distribution with the parameter θ. To test $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$(say,$\theta_1 < \theta_0$) with size biased weight function, $w(x,\beta) = x^r$,

$$\begin{aligned} B_{01}^w &= B_{01}\left(\frac{E_{\theta_1}[w(X,\beta)]}{E_{\theta_0}[w(X,\beta)]}\right) \\ &= \left(\frac{\theta_0}{\theta_1}\right)^{r+1} e^{(\theta_1-\theta_0)x} = \left(\frac{\theta_0}{\theta_1}\right)^r B_{01} \end{aligned}$$

$H_0(H_1)$ is more supported than $H_1(H_0)$ when x is observed in weighting form in comparison with original form. The rate of supporting of H_0 is $\left(\frac{\theta_0}{\theta_1}\right)^r$. It is depend on the parameter of weight function and the parameters given in H_0 and H_1. Table (1) shows the weighted bayes factor under some common weight functions with $\pi_0 = \frac{1}{2}$ comparing with unweighted bayes factor. The forth column of table shows the supporting rate of H_0 when the weighted version of X is used superseding of X.

4.2 Simple against two-sided hypothesis

Consider $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. If the prior distribution consider to be the improper distribution given by $g(\theta) = 1$ we have the unweighted

Table 2: The weighted bayes factor for $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ in exponential distribution for various weight function $\pi_0 = \frac{1}{2}$ and $g(\theta) = 1$

Weight function	$w(x,\beta)$	Weighted bayes factor	The supporting rate of H_0	$B_{01}^w > B_{01}$ if
Unweighted	c	$\frac{x^2}{\Gamma(2)}\theta_0 e^{-\theta_0 x}$	-	1
Length biased	x	$(x\theta_0)B_{01}$	$x\theta_0$	$x\theta_0 > 1$
Size biased	x^r	$\frac{(x\theta_0)^r}{\Gamma(r+2)}B_{01}$	$\frac{(x\theta_0)^r}{\Gamma(r+2)}$	$x\theta_0 > \Gamma^{-r}(r+2)$
Exponential	$e^{\tau x}$	$(\theta_0-\tau)x^2 e^{-x(\theta_0-\tau)}$	$\left(1-\frac{\tau}{\theta_0}\right)e^{\tau x}$	$x > \frac{1}{\tau}\ln\left(\frac{\theta_0}{\theta_0-\tau}\right)$
Set Selection(a)	$X > a$	$(x-a)^2\theta_0 e^{-\theta_0(x-a)}$	–	–
Set Selection(b)	$X < b$	$\frac{\theta_0 e^{-\theta_0 x}}{(1-e^{-\theta_0 b})\sum_{j=0}^{\infty}\frac{1}{(bj+x)^2}}$	–	–
Set Selection(c)	$c < X < d$	$\frac{\theta_0 e^{-\theta_0 x}}{(e^{-\theta_0 a}-e^{-\theta_0 b})\int_0^{\infty}\frac{\theta e^{-\theta x}}{e^{-\theta a}-e^{-\theta b}}d\theta}$	–	-

Bayes factor is

$$B_{01} = \frac{f(x|\theta_0)}{\int_{\theta\neq\theta_0} f(x|\theta)d\theta} = \frac{\theta_0 e^{-\theta_0 x}}{\int_{\theta\neq\theta_0}\theta e^{-\theta x}d\theta} = \frac{x^2}{\Gamma(2)}\theta_0 e^{-\theta_0 x}.$$

Now consider, as an example, $w(x,\beta) = x^r$

$$\begin{aligned} B_{01}^w(x) &= \frac{\theta_0 e^{-\theta_0 x}}{E_{\theta_0}[X^r]\int_{\theta\neq\theta_0}\frac{\theta e^{-\theta x}}{E_\theta[X^r]}g(\theta)d\theta} \\ &= \frac{\theta_0^{r+1}e^{-\theta_0 x}}{\int_{\theta\neq\theta_0}\theta^{r+1}\theta^{-\theta x}d\theta} = \frac{(x\theta_0)^r}{\Gamma(r+2)}B_{01} \end{aligned}$$

Again we consider six common weight functions. The results of weighted Bayes factor are given in table (2)

5 Lower bound on weighted bayes factor

One important problem in point null hypothesis testing $H_0 : \theta = \theta_0$ against two-sided $H_1 : \theta \neq \theta_0$ is to look at the lower bound of the Bayes factor or to the posterior probability of H_0 be true when the prior distribution is belonged to a reasonable class of distributions. Different classes of prior distributions can be suggested depending the probability density function and

parameter space, Θ (see Berger and Sellke, 1987 and Berger and Delampady, 1987 for some suggestions). In this section the class of all reasonable prior distributions is considered. Define G_A as

$$G_A = \{g(\theta); g(\theta) \text{ is all reasonable distributions}\}.$$

To derive the $\inf_{g\in G_A} B_{01}^w(x)$ for weighted distribution we have

$$\begin{aligned}\inf_{g\in G_A} B_{01}^w &= \inf_{g\in G_A} \frac{f^w(x\,|\,\theta_0)}{m_1^w(x)} = \inf_{g\in G_A} \frac{f^w(x\,|\,\theta_0)}{E_{\theta_0}[w(X,\beta)]\int_\Theta \frac{f(x\,|\,\theta)}{E_\theta[w(X,\beta)]}g(\theta)d\theta} \\ &= \frac{f(x\,|\,\theta_0)}{E_{\theta_0}[w(X,\beta)]\sup_{g\in G_A}\int \frac{f(x\,|\,\theta)}{E_\theta[w(X,\beta)]}g(\theta)d\theta} \qquad (8)\end{aligned}$$

Now suppose that we wand to test $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ in exponential distribution, i.e.

$$f(x|\theta) = \theta e^{-\theta x} \quad \theta > 0, \quad x \geq 0$$

If $w(x,\beta) = x^r$ then

$$\begin{aligned}\inf_{\pi\in G_A} B_{01}^w(x) &= \frac{\theta_0^{r+1}e^{-\theta_0 x}}{\sup_{\theta>0}\theta^{r+1}e^{-\theta x}} \\ &= \Big(\frac{x\theta_0}{r+1}\Big)^{r+1}e^{-\theta_0 x+r+1}\end{aligned}$$

If $w(x,\beta) = e^{\tau x}$ then

$$\begin{aligned}\inf_{g\in G_A} B_{01}^w(x) &= \frac{\theta_0 e^{-\theta_0 x}}{\frac{\theta_0}{\theta_0-\tau}\sup_{\theta>\tau}(\theta-\tau)e^{-\theta x}} \\ &= x(\theta_0-\tau)e^{-\theta_0 x+r+1}\end{aligned}$$

if $w(x,\beta) = I_{[a,\infty)}$ then

$$\begin{aligned}\inf_{g\in G_A} B_{01}^w(x) &= \frac{\theta_0 e^{-\theta_0 x}}{e^{-a\theta_0}\sup_{\theta>a}\left\{\theta e^{-\theta(x-a)}\right\}} \\ &= \frac{\theta_0 e^{-\theta_0(x-a)}}{\sup_{\theta>a} e^{-\theta(x-a)}} = \theta_0(x-a)e^{-\theta_0(x-a)+1}\end{aligned}$$

if $w(x,\beta) = I_{[0,b]}$ then

$$\begin{aligned}\inf_{g\in G_A} B_{01}^{w}(x) &= \frac{\theta_0 e^{-\theta_0 x}}{\left(1-e^{-a\theta_0}\right)\sup_{\theta>0}\left[\theta e^{-\theta x}(1-e^{-\theta b})^{-1}\right]} \\ &= \frac{\theta_0 e^{-\theta_0 x}}{\left(1-e^{-\theta_0 b}\right)}\left[\hat{\theta}e^{-\hat{\theta}x}\left(1-e^{-\hat{\theta}b}\right)^{-1}\right]\end{aligned}$$

where $\hat{\theta}$ is satisfied in

$$\hat{\theta} = x + 1 - (1-e^{-\hat{\theta}b})^{-1}$$

if $w(x,\beta) = I_{[a,b]}(x)$ then

$$\begin{aligned}\inf_{g\in G_A}(H_0\,|\,x) &= \frac{\theta_0 e^{-\theta_0 x}}{\left(e^{-a\theta_0}-e^{-b\theta_0}\right)\sup_{a<\theta<b}\frac{\theta e^{-\theta x}}{e^{-\theta a}-e^{-\theta b}}} \\ &= \frac{\theta_0 e^{-\theta_0 x}}{\left(e^{-a\theta_0}-e^{-b\theta_0}\right)\frac{\hat{\theta}e^{-\hat{\theta}x}}{e^{-\hat{\theta}a}-e^{-\hat{\theta}b}}}\end{aligned}$$

where $\hat{\theta}$ is satisfied in

$$\hat{\theta} = \left[x - \frac{be^{-\hat{\theta}b} - ae^{-\hat{\theta}a}}{e^{-\hat{\theta}b}-e^{-\hat{\theta}a}}\right]^{-1}.$$

All the results can be derived using (3) and considering the fact that for this class

$$\sup_{g\in G_A}\int \frac{f(x\,|\,\theta)}{E_\theta[w(X,\beta)]}g(\theta)d\theta = \frac{\hat{\theta}e^{-\hat{\theta}x}}{E_{\hat{\theta}}[w(X,\beta)]}$$

where $\hat{\theta}(x) = \arg_{\theta\in\Theta}\sup\left\{\frac{f(x|\theta)}{E_\theta[w(X,\beta)]}\right\}$.

References

[1] Bayarri, M. J. and Berger, J. O. (1998). Robust bayesian analysis of selection models, *Annals of Statistics,* 26, 645-659.

[2] Bayarri, M. J. and DeGroot, M. H. (1978a). Bayesian analysis of selection models, *The Statistician,* 36, 137-146.

[3] Bayarri and DeGroot(1987b). Information in selection models, in: Viertl, R. (Ed.): *Probability and Bayesian Statistics,* Plenum Press, New York, 39-51.

[4] Bayarri, M. J. and DeGroot, M. H. (1987). Bayesian analysis of selection models, *The Statistician,* 36, 137-146.

[5] Bayarri, M. J. and DeGroot, M. H. (1988). "A Bayesian view of weighted distributions and selection models," *In Accelerated Life Testing and Expert's Opinions in Reliability*, 70-82, eds. Clarotti, C. A. and Lindley, D. V., Amsterdam, North-Holand.

[6] Bayarri, M. J. and DeGroot, M. H. (1989). "Comparison of experiments with weighted distributions," *In Statistical Data Analysis and Inference*, 185-197, ed. Dodge, Y., Amsterdam, North-Holand.

[7] Bayarri, M. J. and DeGroot, M. H. (1990). "Selection models and selection mechanisms," *In Bayesian and Likelihood Methods in Statistics and Econometrics*, 211-227, eds. Geisser, S., Hodges, J. S., Press, S. J. and Zellner, A., Amsterdam, North-Holand.

[8] Bayarri, M. J. and M. H. DeGroot (1992). "A BAD view of weighted distributions and selection models. *In Bayesian Statistics*, 4, 17-33, eds. Bernardo, J. M., Berger, J. O., David, A. P. and Smith, A. F. M., New York, Oxford University Press.

[9] Bayarri, DeGroot, M. H. and Hoel, P. K. (1987), Truncation, Information, and the coefficient of variation, in: Gleser, L. J., S. J. and Sampson, A. R. (Eds.) *Contributions to Probability and Statistics,* Springer, New York, 412-428.

[10] Berger, J. O. and Delampady, M. (1987). "Testing Precise Hypotheses," *Statistical Science,* **2,** 317:352

[11] Berger, J. O. and Sellke, T. (1987). "Testing a Point Null hypotheses: The Irreconciliability of P-values and Evidence (with discussion)," *Journal of the American Statistical Association,* **82,** 112:139

[12] Fisher, R. A. (1934), The effect of models of ascertainment upon the estimation of frequencies, Annal. Eugen., 6, 13-25.

[13] Fleiss, J. L. (1993), The statistical basis for meta-analysis, *Statistical Methods in Medical Research,* 2, 147-160.

[14] Glass, G. V., (1976), Primary, secodary, and meta-analysis of research, *Educational Researchers,* 5, 3-8.

[15] Iyengar, S. and Greenhouse, J. B. (1988), Selection models and the file-drawer Problem, *Statistical Science,* 3, 109-135.

[16] Larose, D. T. and Dey, D. K. (1996), Weighted distributions viewed in the context of model selection: A Bayesian perspective, *Test*, 5(1), 227-246.

[17] Larose, D. T. and Dey, D. K. (1998), Modelling publication bias using weighted distributions in Bayesian framework, *Computational Statistics and Data Analysis,* 26, 279-302.

[18] Patil, G. P. (2002). Weighted Distributions, In *Encyclopedia of Environmetrics,* 2369-2377, Eds. Shaarawi, A. H. and Piegorsch, W. W., John Wiley, Chichester.

[19] Patil, G. P. and Taillie, C. (1987), Weighted distributions and the effec of weighted functions on Fisher information, Tech. Report Center for Statistical Eology ans Enviroment Statistics, Dept. of Statistics, Pencnnsylvania State Univ.

[20] Rao, C. R. (1965). On Discrete Distributions Arising out of Methods of Ascertainment, Sankhya, 27, 311-324.

[21] Rao, C. R. (1985). Weighted distributions arising out of methods of ascertainment: What population does a sample represent? In: Atkinson, A. C. and Fienberg, S. E. (Eds.), A Celebration of Statistics, Chap. 24, ISI Centenary Volume. Springer-Verlag, New York, pp. 543-569.

[22] Silliman, N. P. (1987), Hierarchical selection models with applications in meta-analysis, *Journal of the American Statistical Association*, 92(439), 926-936.

[23] West, M. (1994). "Discovery sampling and selection models," *In Statistical Decision Theory and Related Topics V*, 221-235, eds. Gupta, S. S. and Berger, J. O., Springer-Verlag, New York.

[24] West, M. (1996). "Inference in successive sampling discovery models," *J. Econometrics*, **75**, 217-238.

A Bayesian approach for analyzing influential factors on unemployment status using a random coefficient transitional binary logistic model with non-monotone missing pattern

S. Eftekhari Mahabadi, and M. Ganjali

Department of Statistics, Shahid Beheshti University

Abstract: A transitional binary logistic model with random coefficients is proposed to model the unemployment statues of household members in two seasons of spring and summer. Data correspond to the labor force survey performed by Statistical Center of Iran in 1385. This model is introduced to take into account two kinds of correlation in the data; one due to the longitudinal nature of the study, that will be considered using a transitional model, and the other due to the assumed correlation between responses of members of the same household which is taken into account by introducing random coefficients into the model. In these data some kind of non-monotone missing pattern occurs that is considered in the proposed model using the breakdown of the joint distribution of the response variables. A Bayesian approach toward estimating model parameters and the Gibbs sampling method to perform parameter estimation and data augmentation is used. Results of using this model are compared with those of three other transitional models, two of which exclude incomplete cases and include the so-called correlation between household members' responses or not and a model for available data without random coefficients. It is shown that the earlier model gains more interpretability and precision due to consideration of all aspects of the collected data.

Keywords:Clustered Longitudinal Binary Data; Gibbs Sampling; Non-Monotone Missing Pattern; Transitional Binary Logistic Model; Unemployment Statues.

1 Introduction

In a panel or longitudinal study, each subject is measured at several occasions. Sometimes there exist clusters of subjects in each period due to some kinds of initial relationship between subjects at each time, for example subjects that are members of the same family may indicate a cluster. In such clustered longitudinal studies, Y_{ijt} indicates the response variable for the jth member ($j = 1, \ldots, n_i$) of the ith ($i = 1, \ldots, K$) cluster at time

t ($t = 1, \ldots, T$). We use n_i for the number of elements in the ith cluster to allow for variation of cluster sizes in the study as it seems reasonable if we assume households as clusters. In these kinds of studies there exist two aspects of correlation that must be taken into account in the model; one is the correlation between responses of a subject in different occasions, and the other is due to clustering that results in correlated responses for subjects in the same cluster at each time period. We use random coefficients to allow for the clustering correlation in the model. Different models can be used to take into account the correlation between responses raised due to the longitudinal nature of the study. One possibility is marginal modeling, a second possibility is random effects modeling. The third approach would be to use Markov (transition) models (see Kaciroti et al., 2006). In such longitudinal studies mentioned above, often some of the subjects do not respond in some occasions which cause for missing responses. There are often two kinds of missing pattern considered for longitudinal studies; one is the monotone pattern that is most common in medical surveys and is caused by subject's dropout before the study is completed, and the other is the non-monotone or general pattern that occurs when some of the subjects withdraw from the study at some occasion and return to the study again at some other occasion. Rubin (1976) and Little and Rubin (2002), made important distinctions between the various types of missing mechanisms for each of the above mentioned patterns such as missing completely at random (MAR), missing at random (MAR) and not missing at random (NMAR). From a likelihood point of view MCAR and MAR are ignorable but NMAR is non-ignorable. In this article a Markov transition model with random coefficients for binary response variables is proposed. Moreover, we consider a general non-monotone missing data pattern with a MAR mechanism. Also we will introduce prior distributions for all model parameters in order to take advantage of using a Bayesian approach.

In the next Section, the attributes of the data belonging to the labor force survey are discussed. In Section 3, the random coefficient transitional binary logistic model is presented. In this section we also give the likelihood function and the posterior distribution of the model parameters. In Section 4 we discuss the appropriate model for labor force data and its computational approach using Gibbs sampling method. We also give

the parameter estimates for the proposed model and a comparison of this model with three other models with different missing and random effect considerations. In Section 5 we give a brief conclusion.

2 Labor force survey data

The data used in this paper are related to the unemployment statues of household members in two seasons of spring and summer in 1385. The information about 224691 people who were present in the study for at least one season is available. Among the sampling units, 9237 people refuse to fill the questionnaire and we omit them from the study because they only form four percent of the whole sample. Therefore there are 215454 subjects remained in the study. We first remove inactive part of the spring sample from the study that results in 130344 economically active people in the study. Again, among these we omit the ones who were active in spring but became inactive in summer that leads to 83144 subjects in the whole sample. At last, we remove 13 people from the resulting sample due to incomplete covariates that leads into a sample of 83131 economically active people that includes 51791 households. Due to the special rotating method of sampling in this survey, the sample is divided into three parts. There are 31153 people in the sample who are present only in the first season, 29890 people who are included in the survey only in the second season and 22088 numbers of them were present in both seasons. Variables used through the analysis process are The binary variable indicating unemployment statues (employed and unemployed) as the response variable and covariates including Gender, Age, Highest educational qualification, Number of people living in the Household, Marital statues and living area. All the above variables are considered as categorical with the exception of age that is a continuous variable.

3 Random coefficient transitional binary logistic model

Let Y_{ijt} denotes binary response variable of a clustered longitudinal study in T periods. Also let X_{ijt} be the set of all covariates in the study those which

can be regarded as time varying. We assume a continuous latent variable U_{ijt} related to the binary response of Y_{ijt} with the following mechanism for $i = 1, ..., K; \quad j = 1, ..., n_i; \quad t = 1, ..., T$:

$$Y_{ijt} = 0 \Leftrightarrow U_{ijt} \leq \alpha_{0t} \tag{1}$$

Where α_{0t} for $t = 1, \ldots, T$ are cut point parameters. Also We consider following transition model with random coefficients for the latent variable:

$$U_{ijt} = -X'_{ijt}\beta_t - \gamma_t y_{ij,t-1} - b_{it} + \varepsilon_{ijt} \tag{2}$$
$$\varepsilon_{ijt} \overset{iid}{\sim} F; b_i = (b_{i1}, b_{i2}, ..., b_{iT}) \overset{iid}{\sim} MVN(0, \Sigma_b); b_{it} \perp \varepsilon_{ijt}$$

In which $y_{ij0} = 0$ for all i and j, and we assume Σ_b as a diagonal matrix with diagonal elements $(\tau_1, ..., \tau_T)$. Using the above model U_{ijt} is independent of $U_{ij,t-1}$ given $Y_{ij,t-1}$ as a consequence of transitional nature of the model. Also random effects b_{it} are introduced to account for the correlation due to clustering. We can consider the binary response of Y_{ijt} as a Bernoulli distributed variable with success probability $\pi_{ijt} = \Pr(Y_{ijt} = 1 | X_{ijt}, y_{ij,t-1}, b_{it})$ which according to the relation between Y_{ijt} and U_{ijt}, given in equation (1) and the proposed model in (2) can be regarded as:

$$\pi_{ijt} = F(\alpha_{0t} + X'_{ijt}\beta_t + \gamma_t y_{ij,t-1} + b_{it}) \tag{3}$$

Applying Different distribution functions, F, lead to different models. In our application, We especially use the logistic distribution that leads into a binary logistic model with success probability of

$$\pi_{ijt} = \frac{e^{\alpha_{0t} + X'_{ij}\beta_t + \gamma y_{ij,t-1} + b_{it}}}{1 + e^{\alpha_{0t} + X'_{ij}\beta_t + \gamma y_{ij,t-1} + b_{it}}}. \tag{4}$$

3.1 Likelihood and posterior function

In order to obtain the likelihood function for the set of model parameters $\Theta = \{\alpha_{0t}, \beta_t, \gamma_t, \tau_t; t = 1, ..., T\}$ using the model proposed in the previous

section, we have:

$$L(\Theta|Y_1,...,Y_T) = \prod_{i=1}^{K}\prod_{t=1}^{T} \times \int_{b_i} \prod_{j=1}^{n_i} f(Y_{ijt}|Y_{ij,t-1}, X_{ijt}, \alpha_{0t}, \beta_t, \gamma_t, b_{it}, \tau_t)\phi_{1,...,T}(b_i)db_i.$$

In which $f(Y_{ijt}|Y_{ij,t-1}, X_{ijt}, \alpha_{0t}, \beta_t, \gamma_t, b_{it}, \tau_t)$ is the Bernoulli density with corresponding success probability π_{ijt} and also $\phi_{1,...,T}(.)$ is the multivariate normal density with mean 0 and covariance matrix Σ_b. To complete the model specification, a diffuse prior distribution for Θ, $p(\theta)$, is assumed where for $t = 1, \ldots, T$, $(\alpha_{0t}, \beta_t, \gamma_t)$ have independent diffuse normal priors with mean 0 and some large variance and τ_t follows a diffuse inverse gamma distribution. The primary inferential quantity of interest is $(\alpha_{0t}, \beta_t, \gamma_t; t = 1, .., T)$. Obviously other parameters $(\tau_t, b_{it}; t = 1, ..., T; i = 1, ..., K)$ are also of interest. Given the complexity of the model, inference based on the complete data needs to be based on simulation techniques. For example Gibbs sampling or Markov Chain Monte Carlo (MCMC) methods can be used to construct inferences based on values drawn from the joint posterior density,

$$P(\Theta, b|Y_1, \ldots, Y_T, X) \propto L(\Theta|Y_1, \ldots, Y_T, X)P(\Theta)$$

When there are missing values in $Y = (Y_1, \ldots, Y_T)$, and missing data mechanism is ignorable (MAR), the Gibbs sampling for the complete data model can be easily modified. We include missing values in Y in the Gibbs sampling steps simply by drawing values from its conditional predictive distribution, given the observed values and the current draw of parameters that will be discussed more in the next sections.

4 Model and results for labor force plan data

In labor force plan data we assume the unemployment statues of household members as the binary response variable of interest. These data only contains two seasons, hence we have $T = 2$. Also we consider X as the matrix of all covariates introduced in Section 2. We consider π_{ijt} as the employment probability conditional on model covariates, previous response

and random effects as defined in equation (3). Similar to model in (4), we assume a transitional binary logistic model with random effects distributed as mentioned in (2) for the response variable. For the complete data we have the likelihood as (5) but considering the missing data problem with a non-monotone pattern which occurred in the labor force plan data, we have to compute the appropriate likelihood function. The missing mechanism in these data based on the design doesn't depend on the missed part of the response variable so that, it has a MAR mechanism. Hence the likelihood function for these data should be viewed in three parts. First part consists of households who were present in both seasons:

$$L^{(1)}(\Theta|Y_{1\text{obs}}, Y_{2\text{obs}}, X) = \prod_{i\in I_1}\int_{b_{i2}}\int_{b_{i1}}\prod_{j=1}^{n_i}[\pi_{ij1}^{y_{ij1}}(1-\pi_{ij1})^{1-y_{ij1}} \times\pi_{ij2}^{y_{ij2}}(1-\pi_{ij2})^{1-y_{ij2}}]\phi_{1,2}(b_{i1},b_{i2})db_{i1}db_{i2}.$$

Where I_1 is the set of all households who were present in both seasons and Θ is the set of all parameters. For the second part of the likelihood, there are some households that were only present in the first season with the corresponding set I_2:

$$L^{(2)}(\Theta_1|Y_{1\text{obs}}, X_1) = \prod_{i\in I_2}\int_{b_{i1}}\prod_{j=1}^{n_i}[\pi_{ij1}^{y_{ij1}}(1-\pi_{ij1})^{1-y_{ij1}}]\varphi_1(b_{i1})db_{i1}.$$

where $\Theta_1 = (\alpha_{01}, \beta_1, \tau_1)$. Here it is obvious that there is no need to consider the remaining vector of parameters $\{\Theta - \Theta_1\}$ due to the transitional nature of the model and that the missing mechanism is MAR. The third part consists of households who only their summer response is observed with all individuals belonging to I_3:

$$L^{(3)}(\Theta|Y_{2\text{obs}}, X) = \prod_{i\in I_3}\prod_{j=1}^{n_i}\sum_{y_{ij1}=0}^{1}\{\int_{b_{i2}}\int_{b_{i1}}[\pi_{ij1}^{y_{ij1}}(1-\pi_{ij1})^{1-y_{ij1}} \times\pi_{ij2}^{y_{ij2}}(1-\pi_{ij2})^{1-y_{ij2}}]\phi_{1,2}(b_{i1},b_{i2})db_{i1}db_{i2}\}.$$

We have used the marginal distribution of the second response by summing over two possible outcomes of the first season to be able to find conditional

probabilities of the second response. From a frequentist point of view, product of all individuals' likelihood might be used to obtain parameter estimates using optimizing functions available in softwares S-Plus or R. However, considering independent diffuse normal priors for $(\alpha_{01}, \alpha_{02}, \beta_1', \beta_2', \gamma)$ with mean 0 and variance 1.0×10^6 and independent diffuse gamma priors for $(1/\tau_1, 1/\tau_2)$ with shape parameter 0.001 and scale parameter 0.001, can be used to base inference on the resulting posterior distribution based on the product of the above three parts of the model likelihood. We will use Gibbs sampling along with the above decomposition to account for the missing problem as will be explained in the next section.

4.1 Computations based on Gibbs sampling

A posterior distribution of Θ is interactable; hence the inferential statistics on the parameters of interest can be constructed based on values drawn from the joint posterior distribution, $P(\Theta, b|Y_1, Y_2, X)$ obtained using Gibbs sampling. WinBUGS software (Spiegelhalter et al. 2003) will be used to implement the draws and derive inferences on parameters of interest. For the complete data inferences, Gibbs sampling (Gelfand and Smith, 1990) involves iteratively drawing from the known conditional distributions. Draws from $P(\beta, \gamma, b, \tau|Y, X)$ where $\beta = (\beta_1, \beta_2)$, $b = (b_1, b_2)$ and $\tau = (\tau_1, \tau_2)$, are generated by Gibss sampling based on the following conditioned distributions:

$(i)[\beta_1, b_1, \tau_1|X, Y_1]$	$(ii)[\beta_2, \gamma, b_2, \tau_2|Y_1, Y_2, X]$
$(i.1)$ $[\beta_1|b_1, \tau_1, X, Y_1]$	$(ii.1)$ $[\beta_2|\gamma, b_2, \tau_2, Y_1, Y_2, X]$
$(i.2)$ $[b_1|\beta_1, \tau_1, X, Y_1]$	$(ii.2)$ $[\gamma|\beta_2, b_2, \tau_2, Y_1, Y_2, X]$
$(i.3)$ $[\tau_1|b_1, \beta_1, X, Y_1]$	$(ii.3)$ $[b_2|\beta_2, \gamma, \tau_2, Y_1, Y_2, X]$
	$(ii.4)$ $[\tau_2|b_2, \beta_2, \gamma, Y_1, Y_2, X]$

In practical problems, however, not all of the conditional distributions are known or have closed form. In such cases, rejection sampling, adaptive rejection sampling, the Metropolis algorithm, or the Metropolis-Hastings algorithm are commonly used for drawing values from the distributions.

For inferences Under ignorable missing data mechanism for the response variables in both seasons (non-monotone pattern) with disjoint

parameter spaces in spring and summer, inferences can be made by drawing values from $P(\beta, b, \tau|Y_{obs}, X)$ which is equivalent to drawing from $P(\beta, b, \tau, Y_{miss}|Y_{obs}, X)$. These draws are obtained using Gibbs sampling based on the data augmentation algorithm (Tanner and Wong, 1987) implemented in the following conditional distributions:

$(i)[Y_{1,miss}|\beta_1, b_1, \tau_1, X]$
$(ii)[\beta_1, b_1, \tau_1|Y_1, X]$
$(i.1)$ $[\beta_1|b_1, \tau_1, Y_1, X]$
$(i.2)$ $[b_1|\beta_1, \tau_1, Y_1, X]$
$(i.3)$ $[\tau_1|b_1, \beta_1, Y_1, X]$

$(i')[Y_{2,miss}|\beta_1, b_1, \tau_1, Y_1, X]$
$(ii')[\beta_2, \gamma, b_2, \tau_2|Y_1, Y_2, X]$
$(ii'.1)$ $[\beta_2|\gamma, b_2, \tau_2, Y_1, Y_2, X]$
$(ii'.2)$ $[\gamma|\beta_2, b_2, \tau_2, Y_1, Y_2, X]$
$(ii'.3)$ $[b_2|\beta_2, \gamma, \tau_2, Y_1, Y_2, X]$
$(ii'.4)$ $[\tau_2|b_2, \beta_2, \gamma, Y_1, Y_2, X]$

The four blocks (i), (ii), (i') and (ii') represent an outer Gibbs sampling, from which draws from $P(\beta, b, \tau, Y_{miss}|Y_{obs}, X)$ are obtained. Because drawing directly from blocks (ii) and (ii') is not feasible, an inner Gibbs sampling, as described for complete cases, was implemented within each block. Blocks (i) and (i') are consisted of the I-step in data augmentation, with missing data generated to create a complete data set.

4.2 Results for labor force plan data

In this section, four different transitional models will be compared. The first two models only consider complete cases (22088 individuals who were present in both seasons) whereas the other two models work with all available cases. Model (I) does not assume any random effect parameters but the second model (Model (II)) includes random effects, due to household clustering. Both models are of the form (3) in Section 3, but excluding b_{it} for Model (I). Model (III) and Model (IV) are assumed to be transitional models for the available cases with non-monotone missing pattern as discussed in the previous section. Model (III) does not include random effect parameters but Model (IV) takes into account the existing correlation within household members by including random coefficients (similar to the model proposed in Section 3). Using Bayesian approach, we have performed the iterative Gibbs sampling procedure in 5000 iterations, ignoring the first 1000 iterations as burn-in, we obtain inferences about the model

parameters using 4000 remained iterations. We use the posterior mean of each parameter as its estimate and the sample standard deviation as the estimated standard deviation of the parameter of interest. Results for the proposed models are presented in Table (1) and Table (2).

		Model (I)		Model (II)	
Par.		Spring	Summer	Spring	Summer
		Est. (S. E.)	Est. (S. E.)	Est. (S. E.)	Est. (S. E.)
Intercept		**1.400** (0.232)	0.722 (0.367)	**1.766** (0.304)	0.583 (0.457)
Age		**0.041** (0.003)	**0.023** (0.004)	**0.051** (0.004)	**0.032** (0.005)
Gender	Female (ref.)	-	-	-	-
	Male	0.002 (0.066)	**0.162** (0.079)	0.039 (0.084)	**0.223** (0.104)
Marital status	Single (ref.)	-	-	-	-
	Married	**1.021** (0.068)	**0.924** (0.084)	**1.228** (0.094)	**1.193** (0.114)
	Widowed	0.615 (0.369)	**1.710** (0.577)	0.684 (0.419)	**2.327** (0.712)
	Divorced	-0.215 (0.282)	0.261 (0.357)	-0.179 (0.357)	0.393 (0.466)
Educational degree	Other (ref.)	-	-	-	-
	illiterate	-0.252 (0.244)	**-1.010** (0.388)	-0.388 (0.281)	**-1.163** (0.478)
	Under diploma	-0.295 (0.222)	**-1.241** (0.361)	-0.293 (0.261)	**-1.340** (0.451)
	Diploma or upper	**-0.699** (0.224)	**-1.574** (0.363)	**-0.709** (0.263)	**-1.700** (0.460)
	Bachelor	**-0.874** (0.232)	**-1.573** (0.369)	**-0.948** (0.276)	**-1.677** (0.468)
	Master or PhD	0.918 (0.488)	-0.795 (0.519)	**1.168** (0.549)	-0.744 (0.654)
Family size	At least 4(ref.)	-	-	-	-
	1 member	0.613 (0.414)	0.426 (0.468)	0.695 (0.502)	0.612 (0.627)
	2 members	0.053 (0.114)	-0.104 (0.131)	0.116 (0.143)	-0.086 (0.172)
	3 members	0.143 (0.075)	-0.031 (0.086)	0.188 (0.094)	-0.040 (0.111)
Living Area	Rural (ref.)	-	-	-	-
	Urban	**-0.739** (0.059)	**-0.623** (0.069)	**-0.901** (0.083)	**-0.810** (0.097)
Effect of spring response		-	**2.740** (0.062)	-	**3.756** (0.153)
Variance of random effect		-	-	**2.203** (0.079)	**2.994** (0.053)
DIC		11402.900	8570.300	10670.100	7752.730

Table (1): Results for complete data; (parameter estimations highlighted in **Bold** are significant at 0.05 level of error and DIC is Deviance Information criterion)

According to the results of using complete cases given in Table (1), in both seasons, employment odds of the active part of population increases as the age grows for both models (I) and (II) with an stronger effect in Model (II), fixing all other model covariates. In spring, men's odds of employment is 1.002 times women's in Model (I) and it increases to 1.039 for Model (II) and this odds ratio is 1.176 for Model (I) in summer and 1.249 for Model (II). In Model (I), Employment odds of married people are 2.775 times than that of singles in spring and this odds ratio increases to 2.519 in summer. Model (II) indicates a higher employment odds ratio of 3.414 and 3.296 for married people respectively in spring and summer. People with master degree or PhD have odds of employment 2.504 times than that of illiterate in spring according to Model (I) which increases to 3.215 in Model (II). This odds ratio decreases to 0.417 in Model (I) and 0.475 in Model (II) for summer. In spring, employment odds of households having 1 member is 1.846 times than that of having at least 4 members and this odds ratio reduces to 1.531 for summer based on Model (I) and these odds ratios are 2.003 and 1.84 for model (II). It is noteworthy that employment odds of people belonging to rural households is 2.096 times than that of people in urban households in spring and 1.864 times in summer for Model (I) and these odds ratios increase in Model (II). Also the summer employment odds of people who were employed in spring is 15.487 times than that of others that were unemployed in spring based on Model (I) and this odds ratio increases to 42.777 in Model (II). The overall comparison of Model (I) and Model (II) indicates that the parameter effects are more stronger when the random coefficients are included in the model and that the variance of both random effects in summer and spring are highly significant in Model (II). Hence, for complete data, consideration of the existing correlation between household members by random effects has made the effects more obvious.

Table (2) summarizes the results for models fitted to all available cases with the assumption of MAR mechanism. The magnitude of influence of parameters are nearly the same as Model (I) and Model (II) but these two models gain more precision due to consideration of available observations and the use of data augmentation procedure to consider the special missing pattern of the data. In comparison with the complete data models, theses two model leads to significant effects for all covariates except for some of

the levels in multilevel covariates. It is also apparent that the significance of the effect of all covariates except for some of their levels are obtained in comparison with models for complete data. Also the effect of spring outcome on the employment statues in summer has been decreased in comparison with Model (I) and Model (II) due to data augmentation. In Model (IV) the parameter effects are more stronger than Model (III).

According to the deviance information criterion (DIC; Spiegelhalter et al., 2002) calculated for all 4 models (see Table (1) and (2)), consideration of random effect parameters has reduced DIC in both complete and available data models. Also, comparing Table (1) results with its corresponding results in Table (2) illustrates that using data augmentation method toward available data analysis and including random effects as in Model (IV) have reduced variance of parameter estimates and increased their significant level due to consideration of the potential correlation between household members, hence yielding more precision for analyzing these data. Hence, we would suggest Model (IV) as the best for the employment probability prediction and inference in these data.

5 Conclusion

We use a transition logistic model with random coefficients for longitudinal binary response with MAR non-monotone missing pattern. The model is so flexible to be used with different distributions for the measurement error of the latent variable model. This model provides the ability to model not only monotone missing (dropout) but also non-monotone missing data for two period longitudinal studies. Gibbs sampling is used to obtain Bayesian parameter estimation and data augmentation. For labor force data (Statistical Center of Iran, 1385) we find a random effect transitional model as the best available. We obtain that the response variable in summer is dependent on the spring response and that the correlation due to being a member of the same household is strongly significant. For further work the model can be extended to be used for longitudinal ordinal responses with non-ignorable missing mechanism.

References

[1] Gelfand, A. E., and Smith, A. F. M. (1990). Sampling-Based Approaches to Calculate Marginal Densities. *Journal of the American Statistical Association*, **85**, 398-409.

[2] Kaciroti, N.A., Raghunathan, T.E., Schork, M.A., Clark, N.M., Gong, M. (2006). A Bayesian Approach for Clustered Longitudinal Ordinal Outcome With Nonignorable Missing Data: Evaluation of an Asthma Education Program. *Journal of the American Statistical Association*. **474**, 435-446.

[3] Little, R.J. A, Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. New York: John Wily.

[4] McCullagh, P. (1980). Regression models for ordinal data (with discussion). *J. Roy. Statist. Soc. B*, **42**, 109-142.

[5] Rubin, D. B. (1976). Inference and missing data. *Biometrika*, **63**, 581-592.

[6] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of Royal Statistical Society* B, **64**, 583-639.

[7] Spiegelhalter, D. J., Thomas, A., Best, N. G., and Lunn, D. (2003). Win- BUGS User Manual, Version 1.4, MRC Biostatistics Unit, available at www.mrc-bsu.cam.ac.uk/bugs.

[8] Tanner, M., and Wong, W. H. (1987). The Calculation of Posterior Distribution by Data Augmentation (with discussion). *Journal of the American Statistical Association*, **82**, 528-550.

Par.		Model (III)		Model (IV)	
		Spring	Summer	Spring	Summer
		Est.(S. E)	Est.(S. E)	Est.(S. E)	Est.(S. E)
Intercept		**1.135** (0.018)	**0.851** (0.224)	**1.340** (0.161)	**1.029** (0.258)
Age		**0.038** (0.0003)	**0.026** (0.002)	**0.052** (0.003)	**0.036** (0.003)
Gender	Female (ref.)	-	-	-	-
	Male	**0.197** (0.013)	**0.427** (0.038)	**0.284** (0.052)	**0.604** (0.056)
Marital statues	Single (ref.)	-	-	-	-
	Married	**0.954** (0.015)	**1.086** (0.048)	**1.217** (0.057)	**1.355** (0.067)
	Widowed	**0.832** (0.208)	**0.684** (0.217)	**0.995** (0.279)	**0.909** (0.267)
	Divorced	**-0.466** (0.132)	0.169 (0.206)	-0.338 (0.229)	0.271 (0.269)
Educational degree	Other (ref.)	-	-	-	-
	illiterate	0.056 (0.041)	-0.166 (0.223)	-0.057 (0.157)	-0.289 (0.240)
	Under diploma	**-0.172** (0.009)	**-0.556** (0.210)	-0.149 (0.140)	**-0.604** (0.227)
	Diploma or upper	**-0.627** (0.014)	**-1.017** (0.210)	**-0.616** (0.142)	**-1.154** (0.227)
	Bachelor	**-0.702** (0.047)	**-1.055** (0.214)	**-0.686** (0.149)	**-1.184** (0.231)
	Master or PhD	**0.765** (0.126)	-0.054 (0.296)	**0.906** (0.305)	0.062 (0.348)
Family size	At least 4 (ref.)	-	-	-	-
	1 member	0.062 (0.101)	**0.885** (0.266)	0.224 (0.269)	**1.172** (0.346)
	2 members	**0.160** (0.046)	-0.048 (0.074)	0.079 (0.089)	0.008 (0.099)
	3 members	**0.168** (0.018)	0.003 (0.049)	**0.154** (0.059)	0.034 (0.066)
Living Area	Rural (ref.)	-	-	-	-
	Urban	**-0.767** (0.028)	**-0.757** (0.041)	**-0.914** (0.053)	**-0.938** (0.059)
Effect of spring response		-	**1.167** (0.044)	-	**1.550** (0.071)
Variance of random effect		-	-	**2.544** (0.024)	**2.878** (0.061)
DIC		29938.200	25920.900	27659.100	23640.000

Table (2): Results for the available data (assuming MAR); (parameter estimations highlighted in **Bold** are significant at 0.05 level of error and DIC is Deviance Information criterion)

The asymptotic properties of some discrete distributions generated by standard skewed stable laws

Davood Farbod, and Karen V. Gasparian

Department of Mathematics, Yerevan State University, Yerevan, 375025, ARMENIA

Abstract: There are some well-known frequency distributions in bioinformatics having properties like stable densities. It is of interest to construct such frequency distributions. In the present article the large-sample distributions of the Maximum Likelihood Estimates (M.L.E) of the index and skewness parameters for some discrete distributions generated by *skewed* stable densities are studied. It is shown the existence, strong consistency, asymptotic normality and asymptotic efficiency of that.

Keywords:Asymptotic properties; M.L.E; Regularity conditions; Stable Laws.

1 Introduction

A basic subject of any statistical inference in bioinformatics is characterization of the distributions of object frequencies for a population so-called frequency distributions. Several common statistical facts on frequency distributions have been discovered (see Jeong et al. (2000), Kuznetsov (2003) and Rzhetsky and Gomez (2001)). From the mathematical point of view these are: skewness; regular variation at infinity; continuity by parameters; unimodality, etc. For more details on this, we refer to Astola and Danielian (2006).

In bioinformatics some frequency distributions are widely used. But the variety of such systems requires to generate new ones that satisfy the empirical facts above. Taking into consideration statistical facts, new frequency distributions so-called *Stable Laws* are suggested. The *Stable Laws* are a rich class of probability distributions that allow skewness, heavy tails and have many intriguing mathematical properties.

Any stable distribution requires four parameters to describe: an index exponent $\alpha \in (0, 2]$, a skewness parameter $\beta \in [-1, 1]$, a scale parameter $\gamma > 0$ and a location parameter $\delta \in \Re$, (see, for example, Nolan (2007) and Zolotarev (1986)). There are several equivalent definitions of stable distributions in Nolan (2007). One of them is as follows:

Definition 1.1 A random variable X is said to have a *stable* distribution if for X_1 and X_2 independent copies of X and any positive numbers A and B, there exists positive number C and real number D such that

$$AX_1 + BX_2 \stackrel{d}{=} CX + D,$$

where $C^\alpha = A^\alpha + B^\alpha$, for a unique $\alpha \in (0, 2]$. (The symbol $\stackrel{d}{=}$ means equality in distribution).

In this article we suppose that

$$\Theta = \{\theta = (\alpha, \beta) : \ 0 < \varepsilon \leq \alpha < 1 \ or \ 1 < \alpha < 2, \ 0 < \beta < 1\}, \qquad (\varepsilon \ is \ some \ small \ constant) \qquad (1)$$

and D is an arbitrary open subset of Θ whose closure $\overline{D}$ is also contained in Θ.

Now, we construct the following discrete distribution

$$p(x, \theta) = c_\theta^{-1} s(x, \theta) \quad x = 0, 1, 2, ... \quad with \quad c_\theta = \sum_{y=0}^{\infty} s(y; \theta), \qquad (2)$$

where $s(x; \theta)$ is the density of *Standard Stable Laws* which has the following well-known forms (see Astola and Danielian (2006), Feller (1971) and Zolotarev (1986)):

For $0 < \alpha < 1$ and $x > 0$

$$s(x; \theta) = \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha + 1)}{k!} (-1)^{k-1} x^{-\alpha k - 1} \sin[\frac{\pi k \alpha (1 + \beta)}{2}]. \qquad (3)$$

It is necessary to notice that (3), for $1 < \alpha < 2$, is as an asymptotic expansion when $x \longrightarrow +\infty$. For more details on this, see, for example, Zolotarev (1986).

For $x = 0$ and $0 < \alpha \leq 2$, $\alpha \neq 1$,

$$s(0;\theta) = \frac{1}{\pi}\Gamma(1+\frac{1}{\alpha})\cos(\frac{\pi}{2}\beta\frac{K(\alpha)}{\alpha}), \tag{4}$$

where $K(\alpha) = \alpha - 1 + sign(1-\alpha)$.

2 The asymptotic properties of the M.L.E

Let $X^n = (X_1, X_2, ..., X_n)$ be a finite sample from $p(x;\theta)$, $\theta \in \Theta$, and $x^n = (x_1, x_2, ..., x_n)$ be the realization of X^n, and suppose $\widehat{\theta}_n$ is the M.L.E of θ. Our purpose is to demonstrate the strong consistency, asymptotic normality and asymptotic efficiency of the M.L.E of the parameter θ. To prove this statement, in accordance with the following theorem (see Dumouchel (1973), Borovkov (1998) and Lehmann (1983)) it is sufficient to show that the family $p(x,\theta)$, $\theta \in \Theta$, satisfies the regularity conditions 1-5.

Theorem 1 *Let Θ be as in (1) and $\theta_0 \in \overline{D}$ be the true value of θ. Under satisfying the regularity conditions 1-5, the M.L.E $\widehat{\theta}_n$ is:*

(i) with probability 1, the unique solution of the likelihood equation $\frac{\partial L(x^n;\theta)}{\partial\theta} = 0$ in the region $|\widehat{\theta}_n - \theta_0| < \eta$, where η is some positive number independent of α_0, $L(x^n;\theta) = \prod_{i=1}^n p(x_i;\theta)$;

(ii) strongly consistent, asymptotically normal and asymptotically efficient, i.e. as $n \longrightarrow \infty$ then $\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I^{-1}(\theta_0))$, where $I(\theta_0)$ is a Fisher's information matrix.

For proving of this theorem, we refer the reader to Borovkov (1998) and Lehmann (1983).

3 Statement and proving of the regularity conditions

Condition 1. *The function $p(x;\theta)$ is continuous of θ for $\theta \in \Theta$, and has continuous partial derivatives of first and second order with respect to θ for $\theta \in \overline{D}$.*

Condition 2. *For all $\theta \in \Theta$ and for all $\theta_0 \in \overline{D}$ $(\theta_0 \neq \theta)$, the condition $\sum_{x=0}^{\infty} |p(x;\theta) - p(x;\theta_0)| > 0$ is met.*

The proofs of Conditions 1 and 2 are obvious.

Condition 3. *For all $\theta_0 \in \overline{D}$,*

$$E_{\theta_0}[\sup_{\theta \in \Theta - \overline{D}} \ln \frac{L(X^n;\theta)}{L(X^n;\theta_0)}] < \infty.$$

Proof: Assume that, without loss of generality, $x_1 = 0$ and $x_i \neq 0$ $(i = 2, 3, ...)$. We have

$$L(x^n;\theta) = \frac{s(0;\theta)}{c_\theta} \times \prod_{i=2}^{n} \frac{s(x_i;\theta)}{c_\theta}.$$

Let us consider, for $1 < \alpha < 2$, $\beta = 2 - \alpha$. For all $(\alpha, \beta) \in \{\Theta - \overline{D}\}$, $L(x^n;\theta)$ is bounded uniformly except for the cases $(\alpha, \beta) \longrightarrow (2, 0)$ or $(\alpha, \beta) \longrightarrow (1, 1)$.

If $(\alpha, \beta) \longrightarrow (2, 0)$, then

$$L(x^n;\theta) \longrightarrow 0.$$

If $(\alpha, \beta) \longrightarrow (1, 1)$, then for n fixed it can be shown that $L(x^n;\theta)$ is finite, which completes the proof of Condition 3.

Condition 4. *Let $A(x;\theta) = A_{ij}(x;\theta)$, $i, j = 1, 2$, be the matrix of second order derivatives $A_{ij}(x;\theta) = \frac{\partial^2 \ln p(x;\theta)}{\partial\theta_i \partial\theta_j}$, $\theta_1 = \alpha$, $\theta_2 = \beta$. Then for $\theta \in \overline{D}$, each element of $|A(x;\theta)|$ is majorized by a function $B(x)$, for which*

$$\sum_{x=0}^{\infty} B(x)p(x;\theta) < \infty. \tag{5}$$

Proof: Since $\overline{D}$ is a compact set and $A(x,\theta)$ is continuous with respect to θ, for x fixed and $\theta \in \overline{D}$, the elements of matrix $|A(x;\theta)|$ are restricted by a function $B(x)$, which is itself bounded in any closed interval including x. It remains to investigate the behavior of $B(x)$ as $x \longrightarrow +\infty$.
Let us assume

$$p_\theta(x;\theta) = \frac{\partial}{\partial\theta}(\frac{s(x;\theta)}{c_\theta}), \quad p_{\theta\theta}(x;\theta) = \frac{\partial^2}{\partial\theta^2}(\frac{s(x;\theta)}{c_\theta}) \tag{6}$$

Moreover, we have

$$\begin{aligned} s_\alpha(x;\theta) &= \tfrac{1}{\pi}\textstyle\sum_{k=1}^{\infty} \frac{\Gamma'(k\alpha+1)}{(k-1)!}(-1)^{k-1}x^{-\alpha k-1}\sin(\frac{\pi k\alpha(1+\beta)}{2}) \\ &+ \tfrac{1}{\pi}\textstyle\sum_{k=1}^{\infty} \frac{\Gamma(k\alpha+1)}{(k-1)!}(-1)^{k-1}x^{-\alpha k-1}[\frac{\pi(1+\beta)}{2}\cos(\frac{\pi k\alpha(1+\beta)}{2}) \\ &- (\ln x)\sin(\tfrac{\pi k\alpha(1+\beta)}{2})], \end{aligned} \tag{7}$$

$$\begin{aligned} s_{\alpha\alpha}(x;\theta) &= \tfrac{1}{\pi}\textstyle\sum_{k=1}^{\infty} \frac{\Gamma''(k\alpha+1)}{(k-1)!}k(-1)^{k-1}x^{-\alpha k-1}\sin(\frac{\pi k\alpha(1+\beta)}{2}) \\ &+\tfrac{2}{\pi}\textstyle\sum_{k=1}^{\infty} \frac{\Gamma'(k\alpha+1)}{(k-1)!}k(-1)^{k-1}x^{-\alpha k-1}[\frac{\pi(1+\beta)}{2}\cos(\frac{\pi k\alpha(1+\beta)}{2})] \\ &-(\ln x)\sin(\tfrac{\pi k\alpha(1+\beta)}{2}) + \tfrac{1}{\pi}\textstyle\sum_{k=1}^{\infty} \frac{\Gamma(k\alpha+1)}{(k-1)!}k(-1)^{k-1}x^{-\alpha k-1}[\\ &(\ln^2 x - \tfrac{\pi^2(1+\beta)^2}{4})\sin(\tfrac{\pi k\alpha(1+\beta)}{2}) - \pi(1+\beta)(\ln x)\cos(\tfrac{\pi k\alpha(1+\beta)}{2})], \end{aligned} \tag{8}$$

and

$$s_\beta(x;\theta) = \frac{\alpha}{2}\sum_{k=1}^{\infty}\frac{\Gamma(k\alpha+1)}{(k-1)!}(-1)^{k-1}x^{-\alpha k-1}\cos(\frac{\pi k\alpha(1+\beta)}{2}), \tag{9}$$

$$s_{\beta\beta}(x;\theta) = \frac{\pi\alpha^2}{4}\sum_{k=1}^{\infty}\frac{\Gamma(k\alpha+1)}{(k-1)!}k(-1)^{k}x^{-\alpha k-1}\sin(\frac{\pi k\alpha(1+\beta)}{2}). \tag{10}$$

Now, using (6)-(10) and doing some relatively tedious calculations we obtain $B(x) = O(ln^2 x),\ x \longrightarrow +\infty$.

Due to the value of $B(x)$ as $x \longrightarrow +\infty$, it is easy to see that (5) is met. The condition 4 is proved.

Condition 5. *For each $\theta \in \overline{D}$, the Fisher's information matrix $I(\theta) = \|I_{ij}(\theta)\|$, $I_{ij}(\theta) = E_\theta[\frac{\partial \ln p(X;\theta)}{\partial \theta_i}.\frac{\partial \ln p(X;\theta)}{\partial \theta_j}]$ is nonsingular and is continuous with respect to θ.*

Proof: We have

$$I(\theta) = \begin{pmatrix} I_{11}(\theta) & I_{12}(\theta) \\ I_{21}(\theta) & I_{22}(\theta) \end{pmatrix}.$$

For proving that $I(\theta)$ is nonsingular it suffices to show that

$$[cov(\frac{\partial \ln p(X_1;\theta)}{\partial \alpha}.\frac{\partial \ln p(X_1;\theta)}{\partial \beta})]^2 < var(\frac{\partial \ln p(X_1;\theta)}{\partial \alpha})var(\frac{\partial \ln p(X_1;\theta)}{\partial \beta}).$$

This is equivalent to saying that

$$\frac{\partial \ln p(X_1;\theta)}{\partial \alpha} \neq b\frac{\partial \ln p(X_1;\theta)}{\partial \beta} \qquad (b \text{ is some positive constant}). \qquad (11)$$

The proof of (11) with the help of (7) and (9) is not difficult.

In addition, it is readily seen that the continuity condition of $I(\theta)$ holds. The Condition 5 is established.

Now, the following theorem based on theorem 2.1 is proved:

Theorem 2 *When sampling from $p(x,\theta)$, the M.L.E $\widehat{\theta}_n$ for $\theta \in \Theta$ based on the first n observations is strongly consistent, asymptotically normal and asymptotically efficient if θ_0 (the true value of θ) is in the interior of the parameter space Θ.*

Acknowledgment. The authors are grateful to Professor Eduard Danielian for useful discussions.

References

[1] Astola, J. and Danielian, E. (2006). *Frequency Distributions in Biomolecular Systems and Growing Networks*. TICSP Series No. 31, Tampere, Finland.

[2] Astola, J., Gasparian, K.V. and Danielian, E. (2007). The Maximum Likelihood Estimators for Distributions with Moderate Growth, *Proceedings of CSIT Conference*, Sept. 24-28, Yerevan, Armenia, 91-94.

[3] Borovkov, A.A. (1998). *Mathematical Statistics*, Gordon and Breach Science Publishers, Translated from the Russian into English.

[4] DuMouchel, W.H. (1973). On the Asymptotic Normality of the Maximum Likelihood Estimate when Sampling from a Stable Distribution, *Ann. Statist.*, **1**, 948-957.

[5] Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, 2nd Edition, Vol. 2, John Wiley and Sons.

[6] Jeong, H., Tombor,B., Albert, R., Ottval, Z.N. and Barabasi, A.L. (2000). The Large-Scale Organization of Metabolic Networks, *Nature*, **407**(6804), 651-654.

[7] Kuznetsov, V.A. (2003). Family of Skewed Distributions associated with the Gene Expression and Proteome Evolution, *Signal Process*, **83**(4), 889-910.

[8] Lehmann, E.L. (1983). *Theory of Point Estimation*, John Wiley and Sons.

[9] Matsui, M. and Takemura, A. (2004). Some Improvments in Numerical Evaluation of Symmetric Stable Density and its Derivatives , *Technical Report* , CIRJE-F-Univ. of Tokyo, http://arxiv.org/abs/math/0408321.

[10] Nolan, J.P. (2007). *Stable Distributions - Models for Heavy Tailed Data*, Boston: Brikhauser, In progress, chapter 1 online at academic 2, American.edu/ $\sim$ Jpnolan.

[11] Rzhetsky, A. and Gomez, S.M. (2001). Birth of Scale-Free Molecular Networks and the Number of Distinct DNA and Protein Domains per Genome, *Bioinformatics*, **17** (10), 988-996.

[12] Zolotarev, V.M. (1986). *One-Dimensional Stable Distributions*, Volume 65 of Translations of Mathematical Monographs, American Mathematical Society, Translation from the original 1983 Russian edition.

Step-Stress Accelerated Life Test Planning for Life Data Analysis

Nasser Fard, Chenhua Li

Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA 02115

Abstract: This paper presents time-censored (type I) and failure-censored (type II) step-stress accelerated life test (SSALT) for exponential and Weibull distributed time to failure data. Model for SSALT data analysis and the cumulative exposure model (CE model) are presented. For the exponentially distributed SSALT data, maximum likelihood estimation method (MLE) is used for parameter estimation. The transformed least squares approach for Type II exponentially distributed SSALT data analysis is discussed. We also propose the analysis of Type I censored Weibull data.

Keywords: Accelerated life test; step-stress; time-censored; failure censored; maximum likelihood estimation.

1 Introduction

Reliability and maintainability are important components in engineering design. Their growth has been motivated by several factors, which include the increased complexity and sophistication of systems, public awareness and insistence on product quality. Reliability is defined as the probability that a given component or system will perform its required function without failure for a given period of time, when used under stated operating conditions. In other words, reliability is the probability that a given system will perform as anticipated. Accurate prediction and control of reliability plays an important role in preventive maintenance scheduling, warranty conditions and periods. Manufacturing systems that can economically design and market products that meet their customers' reliability expectations have a strong competitive advantage in today's marketplace.

An integrated product test program may consist of several types of tests, each having different objectives, such as functional or operational tests, environmental stress testing, reliability qualification tests, reliability life testing, safety testing, and reliability growth testing. The primary

objective of reliability life testing is to obtain information concerning failures in order to quantify reliability, to determine whether reliability and safety goals are being met, and to improve product reliability. One of the most important reliability testing is accelerated life testing. Accelerated life testing (ALT) consists of a variety of test methods for shortening the life of products or hastening the degradation of their performance by accelerating failures of highly reliable products. Accelerated life testing is often performed to quickly obtain information on the life distribution or product reliability under normal operating conditions. Specimens are tested at higher-than-operating levels of stress (e.g., temperature, voltage, pressure, vibration, and cycling rate) to induce early failure. Failure information is then extrapolated to normal operating conditions, based on an assumed life-stress relationship. Such testing could save much time and money.

In the traditional life data analysis, times-to-failure data (of a product, system or component) obtained under normal operating conditions are analyzed in order to quantify the life characteristics of the product, system or component, and to make predictions about products performance. In many situations, and for many reasons, it may be difficult or even impossible to obtain such a life data. The long life times of today's products, the small time period between design and release and the challenges in testing products that are used continuously under normal operating conditions are among the difficulties. Therefore, in order to observe products failures for analyzing their failure modes and understanding their life characteristics in a short time, accelerated life tests (ALT) are developed. Through ALT, stress levels which cause product failure are increased and life data for the product under accelerated stress conditions are captured. ALT approach is suitable for products or component that are used on a continuous time basis, such as tires, toasters, heaters, light bulbs.

In general, the accelerated life testing can be divided into two categories: qualitative accelerated testing such as: HALT, HASS, torture tests, shake and bake tests, and quantitative measures in accelerated life testing. Quantitative accelerated life testing (QALT) consists of tests designed to quantify the life characteristics of the product, component or system under normal use conditions and thereby provide reliability information. Reliability information includes the prediction of mean life and reliability of the

product under normal use conditions, and projected returns and warranty costs. QALT can be classified as constant stress, step stress, continuously increasing stress, and etc. These types of loads are classified according to the time dependency of the stress variables.

In step-stress accelerated life test (SSALT), the stress applied to the test product is increased in a specified discrete sequence. Test units are initially placed on a specified low stress, and then at a pre-determined time, the stress is changed to a higher level. Simple step-stress tests, which use 2 test-levels, have been widely studied.

In recent years, the studies of step-stress accelerated life testing design and analysis have attracted many interests and efforts. Nelson (1980) described step-stress accelerated life testing and proposed the cumulative exposure model (CE model) for data analysis. The model assumes that the remaining life of specimens depends only on the current cumulative fraction failed and current stress — regardless of how the fraction is accumulated. Miller and Nelson (1983) first presented optimum simple SSALT plans for the case where test units have exponentially distributed life and all units run to failure. Bai, Kim and Lee (1989) extended the results of Miller & Nelson (1983) to the case where a prescribed censoring time is involved. Bai and Kim (1993) proposed an optimum simple SSALT for the Weibull distribution under Type I censoring. Khamis and Higgins (1996) considered quadratic stress-life relationship and derived the optimum 3-step SSALT for the exponentially distributed time-censored data. Yeo and Tang (1999), Tang (2003) derived an optimal simple SSALT, where not only the optimum hold time under low stress but also the optimum low stress level is determined by taking into consideration the target acceleration factor. Xiong (1998), Xiong and Milliken (1999) explored an optimum plan for simple failure-step SSALT, when the stress change time is an order statistic from the exponential lifetime under the low-stress model. Teng and Yeo (2002) proposed transformed least squares approach for the data analysis of failure-censored step-stress accelerated life tests with exponentially distributed failure time. Khamis and Higgins (1998) proposed a time transformation of the exponential CE model for analyzing Weibull SSALT data. Alhadeed and Yang (2002) and Fard and Li (2004) obtained the optimal design for the simple SSALT using the Khamis-Higgins model (K-H model)

with Weibull distribution assumption for complete data and censored data. Li and Fard (2007) presented the optimal SSALT model considering different stress change times for two stress variables and censored Weibull failure data.

In this paper, we discuss the data analysis for both time-censored and failure-censored SSALT with exponential data or Weibull data. In the following section we present the steps for analysis of QALT data: choosing the appropriate life distribution and selecting the proper life-stress models. Section 3 shows the basic model for SSALT data analysis, the cumulative exposure model (CE model). Figures and mathematical expressions are both used to illustrate the CE model. Section 4 presents the analysis for Type I censored exponentially distributed SSALT data, where the procedure is given and MLE method is used for parameter estimation. In section 5, we discuss the transformed least squares approach for Type II censored exponentially distributed SSALT data analysis. In section 6, we proposed the analysis for Type I censored Weibull data. K-H model is assumed and the MLE method is used.

2 Life Distribution and Life-Stress Models

Analysis of QALT data consists of an underlying life distribution that describes the product at different stress levels and a life-stress relationship (or model) that quantifies the manner in which the life istribution changes across different stress levels.

The first step in performing a QALT data analysis is to choose an appropriate life distribution. The assumed underlying life distribution can be any life distribution. The most commonly used life distributions are exponential, Weibull, and lognormal probability density functions. The exponential distribution has been widely used in the past because of its simplicity, which may not be appropriate in many applications. The Weibull and lognormal distributions, which require more involved calculations, are more appropriate in most cases.

After selecting an appropriate underlying distribution, the second step is to select a life-stress model. Figure 1 shows the relationship between applied stress and the life. The curve shows that at higher stress level

less time is required to for a failure occurrence and as the stress is reduced product life becomes longer. This relationship between stress and life provides an effective means for accelerating the test. Commonly used stress-life relationships are Arrhenius Relationship, Eyring Relationship, Inverse Power Law Relationship, and Temperature-Humidity Relationship, Temperature-Non thermal Relationship, General Log-Linear Relationship, and Proportional Hazards Model.

Figure 1: Relationship between Stress and Life

Table 1 lists some commonly used life-stress relationships.

Table 1: Commonly Used Life-Stress Relationship

Relationship	Model
Arrhenius Relationship	$\ln(L(V)) = \ln(C) + \frac{B}{V}$
Eyring Relationship	$L(V) = \frac{1}{V} e^{-\left(A - \frac{B}{V}\right)}$
Inverse Power Law Relationship	$\ln(L) = -\ln(K) - n\ln(V)$
Temperature-Humidity Relationship	$\ln(L(V,U)) = \ln(A) + \frac{\phi}{V} + \frac{b}{U}$
Temperature-Non thermal Relationship	$\ln(L(U,V)) = \ln(C) - n\ln(U) + \frac{B}{V}$
General Log-Linear Relationship	$\ln(L) = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_n X_n$
Proportional Hazards Model	$\lambda(t;X) = \lambda_0(t) \cdot e^{\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m}$

Where L represents a quantifiable life measure, V, U, Xi 's are the stress variables, λ is the failure rate, and $A, B, C, K, n, \phi, b, \alpha_i$ are unknown parameters.

3 SSALT Data Analysis Model

In a SSALT, the data are usually obtained from the step-stress cumulative distribution function (Cdf). However, we are usually interested in the life distribution under constant stress. To analyze the data from SSALT, a model is needed to relate the distribution under step-stress to the distribu-

tion under constant stress. Nelson (1980) defined one such model, called cumulative exposure model (CE model). The model assumes that the remaining life of specimens depend only on the current cumulative fraction failed and current stress — regardless of how the fraction accumulated. Moreover, if held at the current stress, survivors will fail according to the Cdf of stress, but starting at the previously accumulated fraction failed.

Figure 2 describes the CE model, with a three step SSALT (a), and the three Cdf's for the constant stresses *V1*, *V2*, *V3* (b). It shows that the test units first follow the Cdf for *V1* up to the first hold time *t1*. When the stress increased from *V1* to *V2*, the unfailed units continue along the Cdf for *V2*, starting at the accumulated fraction failed. Similar derivation can be obtained when the stress increases from *V2* to *V3*. Figure 1-c shows the Cdf of the step-stress ALT, which includes the segments of the Cdf's for the constant stresses.

Figure 2: Cumulative exposure model

The CE model for k-step SSALT can be expressed as follows:

$$F_0(t) = \begin{cases} F_1(t) & 0 \le t \le t_1 \\ F_2(t - t_1 + s_1) & t_1 \le t \le t_2 \\ F_3(t - t_2 + s_2) & t_2 \le t \le t_3 \\ \vdots & \vdots \\ F_k(t - t_{k-1} + s_{k-1}) & t_{k-1} \le t \le t_k \end{cases} \tag{1}$$

where $s_0 = t_0 = 0$; and $s_i (i > 0)$ is the solution of:

$$F_{i+1}(s_i) = F_i(t_i - t_{i-1} + s_{i-1}), \quad for \quad i = 1, \ldots, k-1.$$

4 Analysis for Type I Censored Exponential Data

Test Procedure: 1) n units are initially placed on lower stress *S1*, and run until timeτ(also called hold time), when the stress is increased to *S2* and the test is continued until all units fail or until a predetermined censoring time T, whichever comes first.

2) ni failures are observed at time $tij, j = 1, 2, \ldots n_i$ while testing at stress *Si*, $i = 1, 2$, and nc units are censored.

Basic Assumptions:

1) For any level of stress, the life of test units is exponentially distributed.

2) The mean life θ_i of a test unit at stress S_i is a log-linear function of stress. That is,

$$\log(\theta_i) = \alpha_0 + \alpha_1 S_i \tag{2}$$

where α_0 and α_1 (< 0) are unknown parameters depending on the nature of the product and the method of test.

3. A cumulative exposure model holds.

From the assumptions of CE model and exponentially distributed life, the Cdf of a test unit under simple SSALT is:

$$G(t) = 1 - \begin{cases} \exp\left(-\frac{t}{\theta_1}\right) & \text{for } 0 \le t \le \tau \\ \exp\left(-\frac{t-\tau}{\theta_2} - \frac{\tau}{\theta_1}\right) & \text{for } \tau \le t \le \infty \end{cases} \tag{3}$$

Thus the likelihood function from observations *tij*, $i = 1, 2$,$j = 1, 2, \ldots n_i$, is presented as follows:

$$\begin{aligned} L(\theta_1, \theta_2) &= \prod_{j=1}^{n_1} \left[\frac{1}{\theta_1} \cdot \exp\left(-\frac{t_{1,j}}{\theta_1}\right)\right] \\ &\cdot \prod_{j=1}^{n_2} \left[\frac{1}{\theta_2} \cdot \exp\left(-\frac{t_{2,j}-\tau}{\theta_2} - \frac{\tau}{\theta_1}\right)\right] \cdot \prod_{j=1}^{n_c} \exp\left(-\frac{\tau}{\theta_1} - \frac{T-\tau}{\theta_2}\right) \end{aligned} \tag{4}$$

where $n_c = n - n_1 - n_2$. The log likelihood function is then:

$$\log L(\theta_1, \theta_2) = -\left(n_1 \log\theta_1 + n_2 \log\theta_2 + \frac{U_1}{\theta_1} + \frac{U_2}{\theta_2}\right) \tag{5}$$

where $U_1 = \sum_{j=1}^{n_1} t_{1,j} + (n_2 + n_c)\cdot\tau$ and $U_2 = \sum_{j=1}^{n_2} (t_{2,j} - \tau) + n_c (T - \tau)$.

Then the maximum likelihood estimator is obtained by differentiating the log likelihood function. The result is: $\hat{\theta}_1 = \frac{U_1}{n_1}$, $\hat{\theta}_2 = \frac{U_2}{n_2}$ and $\log\hat{\theta}_0 = \frac{\log\hat{\theta}_1 - x_1 \cdot \log\hat{\theta}_2}{1-x_1} = \frac{\log\left(\frac{U_1}{n_1} - x_1 \cdot \log\left(\frac{U_2}{n_2}\right)\right)}{1-x_1}$, where $x_i = \frac{S_i - S_0}{S_2 - S_0}$ for $i = 0, 1, 2..$

5 Analysis for Type II Censored Exponential Data

Here we consider an m-step failure censored SSALT. All n test units are initially placed at lowest stress $x1$ and run until $n1$ failure occurs. At t_{1,n_1}, the stress is changed to $x2$. The test is continued, and the stress changes at times t_{i,n_i}. The test is terminated after nm units have failed at stress level xm. Based on the assumptions of exponentially distributed failure time and the log-linear life-stress relationship, Teng and Yeo (2002) proposed a transformed least squares approach (TLS) for the data analysis of failure-censored SSALT. This approach is based on two lemmas. Proofs of those two lemmas can be found in probability and statistics book.

Lemma 1: Let $t_{(1)}, t_{(2)}, \ldots, t_{(n)}$ be order statistics derived from a sample of n exponential observations with mean θ. Let $t_{(0)} = 0$; then for $i = 1, \ldots, n$, $Z_i = (n - i + 1) \cdot \left(t_{(i)} - t_{(i-1)}\right)$ are s-independent exponential random variable with mean θ.

Lemma 2: Let Z have an exponential distribution with mean θ. Then $\log(Z)$ has an extreme value distribution with location parameter $\log(\theta)$ and scale parameter 1. The mean and variance of $\log(Z)$ are: $\log(\theta) - 0.5772$ and 1.283^2, respectively.

The observed failure time $t_{i,j}$ from the failure-censored SSALT are ordered statistics from n exponential observations with mean θ. Let $t_{1,0} = 0, t_{i,0} = t_{i-1,n_{i-1}}$ for $i = 2, \ldots, m$. Then the transformed time difference between two consecutive failures $Z_{i,j} = \left(n - \left(\sum_{k=1}^{i-1} n_k + j\right) + 1\right) \cdot (t_{i,j} - t_{i,j-1})$ are statistically independent and exponentially distributed with mean θ_i by Lemma 1. Then the $\log(Z_{i,j})$ are therefore extreme value random variable with mean $\log(\theta_i) - 0.5772$ and variance 1.283^2.

If define a random variable $W_{i,j} = \log(Z_{i,j}) + 0.5772$, the regression model becomes $W_{i,j} = \alpha_0 + \alpha_1 \cdot x_i + \varepsilon_{i,j}$, where $\varepsilon_{i,j}$ are independent extreme value random variable with mean 0 and variance 1.283^2.

By least squares estimation, the TLSE of α_0, α_1 satisfy

$$\begin{bmatrix} n & \sum_i n_i \cdot x_i \\ \sum_i n_i \cdot x_i & \sum_i n_i \cdot x_i^2 \end{bmatrix} \times \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} \sum_{i,j} W_{i,j} \\ \sum_{i,j} x_i \cdot W_{i,j} \end{bmatrix} \tag{6}$$

Hence $\begin{bmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \end{bmatrix} = \begin{bmatrix} n & \sum_i n_i \cdot x_i \\ \sum_i n_i \cdot x_i & \sum_i n_i \cdot x_i^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum_{i,j} W_{i,j} \\ \sum_{i,j} x_i \cdot W_{i,j} \end{bmatrix}$

And $\log\left(\hat{\theta}_0\right) = \hat{\alpha}_0 + \hat{\alpha}_1 \cdot x_0$ is the TLS estimator of the log of mean life under normal operating condition.

6 Analysis for Type I Censored Weibull Data

The assumption of exponentially distributed life times has limited the application of the above results. Although Weibull distribution is more flexible, the computation complexity of the SSALT model with Weibull distribution has hindered the further research and application. To overcome such difficulty, Khamis and Higgins (1998) proposed a time transformation of the exponential CE model for analyzing Weibull SSALT data. The K-H model for a simple SSALT is:

$$G(t) = 1 - \begin{cases} \exp\left(-\frac{t^\delta}{\theta_1}\right) & 0 \le t \le \tau \\ \exp\left(-\frac{t^\delta - \tau^\delta}{\theta_2} - \frac{\tau^\delta}{\theta_1}\right) & \tau \le t \le \infty \end{cases} \tag{7}$$

where δ is the shape parameter, and θ_i is the scale parameter.

This time-transformation enables the reliability engineer to use known results for multiple-step, multiple-stress models that have been developed for the exponential step-stress model when δ is known. Alhadeed and Yang (2002) considered the case where δ is unknown, and proposed the MLE and the optimal test design using the K-H model, based on the assumption of complete data. Here, we consider Type I censored Weibull data, and present the parameter estimate using MLE method.
Let $y = \log(t)$, $\sigma = 1/\delta$, $u_i = \log(\theta_i) = \beta_0 + \beta_1 S_i$, and $C_i = (\log(\tau) - u_i)/\sigma$, for $i = 0, 1, 2$. Then (6) becomes:

$$G(y) = \begin{cases} 1 - \exp\left(-\exp\left(\frac{y-u_1}{\sigma}\right)\right), & -\infty < y < \log(\tau) \\ 1 - \exp\left(-\exp\left(\frac{y-u_2}{\sigma}\right) + \exp(C_2) - \exp(C1)\right), & \log(\tau) \le y < \infty \end{cases} \tag{8}$$

The log likelihood function for the time censored Weibull data Y_{ij}, $i = 1, 2$,

$j = 1, 2, \ldots, n_i$ is:

$$\log\left[L\left(u_1, u_2, \sigma\right)\right] = -\left(n_1 + n_2\right)\log\left(\sigma\right) + \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\frac{y_{ij} - u_i}{\sigma} - \exp\left(\frac{y_{ij} - u_i}{\sigma}\right)\right) + \left(n_2 + n_c\right)\left(\exp\left(C_2\right) - \exp\left(C_1\right)\right) - n_c\exp\left(C_T\right) \tag{9}$$

where $C_T = \frac{\log(T) - u_2}{\sigma}$. The MLE for u_1, u_2, σ can be obtained by differentiating the log-likelihood function.

$$\frac{\partial\log\left[L\left(u_1, u_2, \sigma\right)\right]}{\partial u_1} = -\frac{1}{\sigma}\left(n_1 - \sum_{j=1}^{n_1}\exp\left(\frac{y_{1j} - u_1}{\sigma}\right) - \left(n_2 + n_c\right)\exp\left(C_1\right)\right) = 0;$$

$$\frac{\partial\log\left[L\left(u_1, u_2, \sigma\right)\right]}{\partial u_2} = -\frac{1}{\sigma}\left(n_2 - \sum_{j=1}^{n_2}\exp\left(\frac{y_{2j} - u_2}{\sigma}\right) + \left(n_2 + n_c\right)\exp\left(C_2\right) - n_c\exp\left(C_T\right)\right) = 0; \tag{10}$$

$$\frac{\partial\log\left[L\left(u_1, u_2, \sigma\right)\right]}{\partial\sigma} = -\frac{1}{\sigma}\left[\begin{array}{l}\left(n_1 + n_2\right) + \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\frac{y_{ij} - u_i}{\sigma}\left(1 - \exp\left(\frac{y_{ij} - u_i}{\sigma}\right)\right)\right) + \\ \left(n_2 + n_c\right)\left(C_2\exp\left(C_2\right) - C_1\exp\left(C_1\right)\right) - n_cC_T\exp\left(C_T\right)\end{array}\right] = 0;$$

The MLE of log of the life with R reliability of the Weibull distribution at normal operating condition is:

$$\hat{Y}_R\left(S_0\right) = \log\left(\hat{\theta}_0\right) + U_R\cdot\hat{\sigma} = \frac{\hat{u}_1 - x\hat{u}_2}{1 - x} + U_R\cdot\hat{\sigma}$$

where $x = \frac{S_1 - S_0}{S_2 - S_0}$, and $U_R = \log\left(-\log\left(R\right)\right)$.

References

[1] A.A. Alhadeed, and S.S. Yang, 2002, "Optimal Simple Step-Stress Plan for Khamis-Higgins Model," IEEE Transactions on Reliability, 51(2), 212-215.

[2] D.S. Bai, M.S. Kim, and S.H. Lee, 1989, "Optimum Simple Step-Stress Accelerated Life Tests with censoring," IEEE Transactions on Reliability, 38(5), 528-532.

[3] D.S. Bai, and M.S. Kim, 1993, "Optimum Simple Step-Stress Accelerated Life Tests for Weibull Distribution and Type-I Censoring," Naval Research Logistics, 40, 193-210.

[4] N. Balakrishnan and A.P. Basu, *The Exponential Distribution: Theory, Methods and Applications*: Gordon & Breach, 1995.

[5] N. Fard, and C. Li, 2004, "Optimum Plan for Step-Stress Accelerated Life Tests", *Proc. of the 2004 placePlaceTypeInstitute of PlaceNameIndustrial Engineering Research Conference*, May 2004 (CD-ROM).

[6] I.H. Khamis, and J.J. Higgins, 1996, "Optimum 3-Step Step-Stress Tests," IEEE Transactions on Reliability, 45(2), 341-345.

[7] I.H. Khamis, and J.J. Higgins, 1998, "A New Model for Step-Stress Testing," IEEE Transactions on Reliability, 47(2), 131-134.

[8] C. Li, and placeN. Fard, 2007, "Optimum Bivariate Step-Stress Accelerated Life Test for Censored Data," *IEEE Transactions on Reliability*, 56(1), 77-84.

[9] R. Miller, and W. Nelson, 1983, "Optimum Simple Step-Stress Plans for Accelerated Life Testing," IEEE Transactions on Reliability, R-32(1), 59-65.

[10] W. Nelson, 1980, "Accelerated Life Testing — Step-Stress Models and Data Analysis," IEEE Transactions on Reliability, R-29(2), 103-108.

[11] W. Nelson, 1990. *Accelerated Testing — Statistical Models, Test Plans, and Data Analyses*. placeStateNew York: Wiley.

[12] S.L. Teng, and K.P. Yeo, 2002, "A Least-Squares Approach to Analyzing Life-Stress Relationship in Step-Stress Accelerated Life Tests". *IEEE Transactions on Reliability*, Vol. 51, No. 2, pp. 177-182.

[13] L.C. Tang, 2003, "Chapter 24: Multiple-Steps Step-Stress Accelerated Life Test," appears in Handbook of Reliability Engineering, Hoang Pham (eds.), Springer Verlag, New York, 441-456.

[14] C. Xiong, 1998, "Inferences On a Simple Step-Stress Model with Type-II Censored Exponential Data," IEEE Transactions on Reliability, 47(2), 142-146.

[15] C. Xiong, and G.A. Milliken, 1999, "Step-Stress Life-Testing with Random Stress-Change Times for Exponential Data," IEEE Transactions on Reliability, 48(2), 141-148.

[16] K.P. Yeo, and L.C. Tang, 1999, "Planning Step-Stress Life-Test with a Target Acceleration-Factor," IEEE Transactions on Reliability, 48(1), 61-67.

A Monte Carlo Method In Application To Estimate Extinction Probability of Bisexual Galton-Watson Process

R. Farnoosh, and Z. Zarabi Zadeh

Department of Applied Mathematics, University of Science and Technology,
Islamic Azad University, Karaj Branch

Abstract: In this paper a bisexual Galton-Watson branching process is introduced. For certain class of processes $\{Z_n\}$ extinction probability is studied when initially population size (Z_0) has a different value. Monte Carlo method is purposed to calculate the extinction probability .

Keywords: Bisexual Galton-Watson process, extinction probability, simulation, Monte Carlo method.

1 Introduction

The bisexual Galton-Watson branching process initially introduced by Daley [3]: Successive generations n=0,1,2,... of an evolutionary process consist of F_n females and M_n males forming $Z_n = L(F_n, M_n)$ mating units, where $L(x, y)$ is a non-negative integer-valued function which is non-decreasing in both $x, y \in N$. These mating units reproduce the next generation independently through the same offspring probability distribution (p_n) in each generation. Each offspring is female with probability $\alpha(0 < \alpha < 1)$. Daley [3] argues that $\{Z_n\}$ is markov chain with the non-negative integers as states. He also obtained a necessary and sufficient condition to extinction with probability one. Hull [5], Bruss [2] and Alsmeyer and Rosler [1] obtain extinction probability for some class of this process.

Usually the calculus of extinction probability is so difficult. In this paper Monte Carlo method is purposed to calculate the extinction probability. For certain class of this processes extinction probability is studied when initially population size (Z_0) has a different value. In section 2 the bisexual Galton-Watson branching process is introduced. In section 3 the extinction probability is estimated and for certain class these processes are simulated.

2 The bisexual Galton-Watson branching process

In this section the bisexual Galton-Watson branching process is introduced.

A bisexual Galton-Watson branching process can be described as follows:

Let $\{(F_{n,i}, M_{n,i}); n = 0, 1, 2, ..., i = 1, 2, ...\}$ be a family of integer-value, independent and identically distributed bivariate random variables, the mating function $L : Z^+ \times Z^+ \to Z^+$ is a function with following characteristics, monotonic non-decreasing in each argument and integer valued for integer-valued arguments and such that $L(x, y) \leq xy$. We define processes $\{Z_n\}_n$ and $\{(F_n, M_n)\}_n$ by the iterative relation:

$$Z_0 = N \geq 1, \;\; (F_{n+1}, M_{n+1}) = \sum_{i=1}^{Z_n} (F_{n,i}, M_{n,i}), \;\; Z_{n+1} = L(F_{n+1}, M_{n+1}),$$

for $n = 0, 1, 2, ...$ and N is a positive integer, with the empty sum the result is $(0, 0)$.

In this model, $F_{n,i}, M_{n,i}$ represent the number of females and males produced by *ith* mating unit in the *nth* generation, respectively. F_n and M_n are the number of females and males in the *nth* generation, respectively, which form, $Z_n = L(F_n, M_n)$ mating unit. These mating units reproduce the next generation independently through the same offspring probability distribution in each generation. Daley [4] argues that $\{Z_n\}, \{(F_n, M_n)\}$ are markov chain with the non-negative integers as states. The state has stationary one step transition probabilities and 0 and $(0, 0)$ states are absorbing. Let $T_{n,i} = F_{n,i} + M_{n,i}$ for $i = 1, 2, ..., Z_n$, the term $T_{n,i}$ denotes the total number of offspring produced by the *ith* mating unit in the *nth* generation. $\{T_{n,i}\}_{n,i}$ are integer-valued, independent and identically distributed random variables, each offspring is female with probability $\alpha (0 < \alpha < 1)$ or male with probability $(1 - \alpha)$.

The bisexual Galton-Watson branching process are identified by the following four parameters:

1) The number of mating units in the initial generation, (Z_0).
2) The mating function (L).
3) The offspring probability low $(\{p_n\}_{n=0}^{\infty})$.

4) The probability that any individual offspring will be female (α).

A branching process is said superadditive when its mating function L is superadditive *i.e.* for any x_1, x_2, y_1, y_2 in Z^+:

$$L(x_1 + x_2, y_1 + y_2) \geq L(x_1, y_1) + L(x_2, y_2).$$

The extinction probability define as a following:

$$q_j = \lim_{n\to\infty} P(Z_n = 0 \mid Z_0 = j)$$

Let $r_k = k^{-1}E(Z_{n+1}|Z_n = k), k = 1, 2, ...$ (mean growth rates) Dalley [4] shows:

$$r = \lim_{k\to\infty} r_k = \sup_{k>0} r_k$$

exists and moreover

$$q_j = 1 \qquad , j = 1, 2, ... \quad \Leftrightarrow \quad r \leq 1.$$

3 Estimation of extinction probability

In this section an estimator for extinction probability is introduced and for certain class these processes are simulated and extinctions probability are estimated.

Define random variable $(X_{n,j})$, when $Z_0 = j$ as a following:

$$X_{n,j} = I_{\{Z_n=0\}}.$$

$X_{n,j}$ has the Bernolli$(1, q_j)$ distribution and an estimator for the extinction probability is given by:

$$\hat{q}_j = \frac{\sum_{i=1}^{n} X_{i,j}}{n}.$$

This estimator is a UMVU estimator (Lehmann [6]).

Here Monte Carlo simulation of this process is purposed to estimate q_j.

To estimate q_j we should identified four parameters of this process then in each generation for each mating unit we simulate the number of offspring (m). Since number of female in each mating unit has the Bernolli(m, α) we can simulate number of female and male in each mating unit $(F_{n,i}, M_{n,i})$ and finally we can obtain number of female and male in nth generation (F_n, M_n). Then we use mating function (L) and obtain number of mating unit in nth generation (Z_{n+1}).

Example 1 A bisexual Galton-Watson branching process with four following parameters is considered:

$$Z_0 = j \ , \ P(T_{n,i} = 1) = 1 \ , \ L(x,y) = \begin{cases} x+y+1 & x \geq 1, y \geq 1, \\ 0 & x = 0, y = 0. \end{cases}$$

Also each offspring is female with probability α. This implies $Z_{n+1} = 0$ or $Z_{n+1} = Z_n + 1$ moreover, $Z_n > 0$ iff $X_n \geq 1, Y_n \geq 1$. Then we can obtain:

$$\begin{aligned} q_j &= \lim_{n\to\infty} P(Z_n = 0 \mid Z_0 = j) \\ &= 1 - \lim_{n\to\infty} P(Z_n > 0 \mid Z_0 = j) \\ &= 1 - \lim_{n\to\infty} P(Z_n > 0 \mid Z_{n-1} > 0) P(Z_{n-1} > 0 \mid Z_0 = j) \\ &= 1 - \lim_{n\to\infty} \prod_{i=1}^{n} P(Z_i > 0 \mid Z_{i-1} > 0) \\ &= 1 - \lim_{n\to\infty} \prod_{i=1}^{n} (1 - P(X_i = 0 \ \ or \ \ Y_i = 0 \mid Z_{i-1} = Z_0 + i - 1)). \end{aligned}$$

When $\alpha = 0.5$ we have

$$q_j = 1 - \lim_{n\to\infty} \prod_{i=1}^{n} (1 - (0.5)^{Z_0+i-2}).$$

For different value of Z_0 we calculate q_j, also we simulate $\hat{q}_j$ and compare q_j and $\hat{q}_j$ in table 1.

The results for 300000 simulations are reasonable.

$Z_0 = j$	$\hat{q}_j$	$\lvert\hat{q}_j - q_j\rvert$	$Z_0 = j$	$\hat{q}_j$	$\lvert\hat{q}_j - q_j\rvert$
3	0.422217	7.2381×10^{-5}	9	7.7187×10^{-3}	7.3428×10^{-5}
4	0.229687	2.1091×10^{-4}	10	3.775×10^{-3}	1.2617×10^{-4}
5	0.119884	4.0152×10^{-5}	11	1.7906×10^{-3}	1.6123×10^{-4}
6	0.060212	9.9700×10^{-4}	12	8.5×10^{-4}	1.2624×10^{-4}
7	0.030926	2.4156×10^{-4}	13	4.7187×10^{-4}	1.6327×10^{-5}
8	0.015906	3.6245×10^{-4}	14	0	2.4412×10^{-4}

Table 1: Comparison of the q_j and $\hat{q}_j$ (300000 simulations).

Example 2 A bisexual Galton-Watson branching process with four following parameters is considered:

$$Z_0 = j \ , \ P(T_{n,i} = 1) = 1 \ , \ L(x,y) = \begin{cases} x+y+1 & x \geq 2, y \geq 2, \\ 0 & x < 2, y < 2. \end{cases}$$

Also each offspring is female with probability α. This implies $Z_{n+1} = 0$ or $Z_{n+1} = Z_n + 1$ moreover, $Z_n > 0$ iff $X_n \geq 2, Y_n \geq 2$. Then we can obtain:

$$\begin{aligned} q_j &= \lim_{n\to\infty} P(Z_n = 0 \mid Z_0 = j) \\ &= 1 - \lim_{n\to\infty} \prod_{i=1}^{n} (1 - P(X_i \leq 1 \ \textit{or} \ Y_i \leq 1 \mid Z_{i-1} = Z_0 + i - 1)). \end{aligned}$$

When $\alpha = 0.5$ we have:

$$q_j = 1 - \lim_{n\to\infty} \prod_{i=1}^{n} (1 - (0.5)^{Z_0+i-2}(Z_0 + j)).$$

For different value of Z_0 we calculate q_j, also we simulate $\hat{q}_j$ and compare q_j and $\hat{q}_j$ in table 2.
The results for 300000 simulations are reasonable.

References

[1] Alsmeyer, G., Rosler, U. (1996). The bisexual Galton-Watson process with promiscuous mating: extinetion probabilities in the supercritical case. *Ann. Appl. Prob.*, **6**, 922-939.

$Z_0 = j$	$\hat{q}_j$	$\lvert\hat{q}_j - q_j\rvert$	$Z_0 = j$	$\hat{q}_j$	$\lvert\hat{q}_j - q_j\rvert$
5	0.634547	1.363×10^{-3}	12	0.013463	1.439×10^{-4}
6	0.416633	8.232×10^{-4}	13	7.13×10^{-3}	1.756×10^{-4}
7	0.253247	1.097×10^{-3}	14	4.08×10^{-3}	1.790×10^{-4}
8	0.146023	1.799×10^{-3}	15	1.99×10^{-3}	8.371×10^{-5}
9	0.083460	8.879×10^{-5}	16	1.11×10^{-3}	1.178×10^{-5}
10	0.045453	6.569×10^{-4}	17	5.53×10^{-4}	2.672×10^{-5}
11	0.025263	9.683×10^{-5}	18	3.12×10^{-4}	6.858×10^{-5}

Table 2: Comparison of the q_j and $\hat{q}_j$ (300000 simulations).

[2] Bruss, T. (1984). A not on a extinction criteria for bisexual Galton -Watson processes. *J. Appl. prob.*, **21**, 915-919.

[3] Daley, D. (1968). Extinction conditions for bisexual Golton-Watson branching processes. *Z.wahvsheinlinchkeit sth*, **9**, 315-322.

[4] Daley, D. J., Hull, M., Taylor, M. (1986). Bisexual Galton-Watson branching processes with superadditive mating function. *J. Appl. Prob.* , **23**, 585-600.

[5] Hull, M. (1982). A necessary condition for extinction in those bisexual Galton-Watson branching processes governed by superadditive mating function. *J. Appl. prob.*, **19**, 847-850.

[6] Lehmann, E.L. (1983). *Theory of Point Estimation*. John Wiley.

Third order asymptotic behavior of Rao score test for inhomogeneous Poisson processes

Khosrow Fazli

Department of Statistics, University of Kurdistan

Abstract: In this work based on a realization of an inhomogeneous Poisson process whose intensity function depends on a real unknown parameter, we consider a simple null hypothesis against a composite one sided alternative. Under certain regularity conditions we obtain the *power loss* of the score test which measures its performance with respect to the Neyman-Pearson test. We present also the second order approximation of the power of the score test under the close (contiguous) alternatives which permits the numerical calculations. This improves the first order classical representation of the power. The results are applied to several important models.

Keywords: Hypotheses testing, inhomogeneous Poisson process, power loss, second order efficiency.

1 Introduction

The spatial Poisson process is widely used in many fields. The diverse choice of intensity function makes it a suitable model for many real phenomena. Estimation theory for spatial Poisson process was extensively developed (see [3] and [10] and the references therein). For the problem of goodness of fit testing with the simple basic hypothesis and a nonparametric alternative see [9]. One can refer to [4] and [2] where a simple Poissonian null hypothesis is tested against a large classes of alternatives of the type of stationary point processes. The present work is devoted to a parametric hypotheses testing problem in which we study the third order asymptotic behavior of the score test when the parameter is one dimensional. For a multidimensional parameter, the second order asymptotic properties of the Rao score test is studied in [7].

Let $X^{(n)}$ be an inhomogeneous Poisson process observed on some increasing subsets $\mathbb{A}_n,\ n = 1, 2, \ldots$ of d dimensional Euclidian space $\mathbb{R}^d$ with intensity function $S(\vartheta, x), x \in \mathbb{A}_n$ depending on one-dimensional parameter $\vartheta \in \Theta$. In this work $X^{(n)}$ represents an inhomogeneous Poisson process or a

realization of the process. Let $\left\{ {}^{(n)}_{\vartheta},\, \vartheta \in \Theta \right\}$ denote the parametric family of distributions of the random element $X^{(n)}$. Based on $X^{(n)}$ we want to test the hypotheses,

$$\begin{aligned} \mathcal{H}_0 &: \ \vartheta = \vartheta_0 \\ \mathcal{H}_1 &: \ \vartheta > \vartheta_0, \end{aligned}$$

where ϑ_0 is a given value in the parameter space Θ. Let us fix some $\alpha \in (0,1)$ and define the class $\mathcal{K}'_\alpha$ of (sequence of) tests of asymptotic level $1-\alpha$ (size α), i.e.,

$$\mathcal{K}'_\alpha = \left\{ \phi_n : \quad \lim_{n\to\infty} {}_{\vartheta_0}\phi_n \left(X^{(n)} \right) = \alpha \right\}.$$

It is well known that if $n \to \infty$ for any given value ϑ of the alternative the power of any reasonable (consistent) test tends to 1 (see [1]). In order to compare different tests we use the Pitman's approach (see [11]) where instead of a fixed alternative ϑ we consider a sequence of so-called *local alternatives* (*close* or *contiguous alternatives*) which converges to ϑ_0 with a certain rate and hence it is difficult to distinguish between the null hypothesis and alternative. More precisely, let $\{\varphi_n\}$ be a sequence of nonnegative numbers which converges to zero with such a rate that the likelihood ratio

$$\begin{aligned} Z_n(u) =& \frac{\mathrm{d}^{(n)}_{\vartheta_u}}{\mathrm{d}^{(n)}_{\vartheta_0}} \left(X^{(n)} \right) = \\ &= \exp \left\{ \int_{\mathbb{A}_n} \ln \frac{S(\vartheta_u, x)}{S(\vartheta_0, x)} \, X^{(n)}(\mathrm{d}x) - \int_{\mathbb{A}_n} \left[S(\vartheta_u, x) - S(\vartheta_0, x) \right] \mathrm{d}x \right\}. \end{aligned}$$

has a nondegenerate limit for any $u > 0$ with $\vartheta_u = \vartheta_0 + \varphi_n u \in \Theta$. Here for the function $f(\cdot)$ defined on $\mathbb{A}_n$ the stochastic integral w.r.t. the Poisson process $X^{(n)}$ is defined by

$$\int_{\mathbb{A}_n} f(x) \, X^{(n)}(\mathrm{d}x) = \sum_{x_i \in \mathbb{A}_n} f(x_i),$$

where $\{x_i\}$ are the *events* (random points) of the Poisson process. By the Neyman-Pearson lemma the most powerful test for $\mathcal{H}_0 : \ \vartheta = \vartheta_0$ against

the local alternative $\mathcal{H}_u:\ \vartheta=\vartheta_0+\varphi_n u$ with $u>0$, is given by

$$\tilde{\phi}_n\left(X^{(n)}\right)=\begin{cases}1, & \text{if } \Lambda_n(u)>b_n(u)\\ 0, & \text{if } \Lambda_n(u)<b_n(u)\end{cases}$$

where $\Lambda_n(u)=\ln Z_n(u)$ and the constant $b_n(u)$ together with the contribution of the randomized part provide the size ${}_{\vartheta_0}\tilde{\phi}_n\left(X^{(n)}\right)=\alpha$. The power of $\tilde{\phi}_n$ as a function of u is called the *envelope power function* . For any fixed n it is the supremum of the power at the local alternative $\vartheta_u=\vartheta_0+\varphi_n u$ over all the tests at level $1-\alpha$. Notice that $\tilde{\phi}_n$ is not a test for the main hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, because it depends on the parameter u. Let us introduce the score statistic,

$$\Delta_n\left(\vartheta_0\right)=\varphi_n\int_{\mathbb{A}_n}\frac{\dot{S}\left(\vartheta_0,x\right)}{S\left(\vartheta_0,x\right)}\,\pi^{(n)}\left(\mathrm{d}x\right),\quad \varphi_n^{-2}=I_n(\vartheta_0)=\int_{\mathbb{A}_n}\frac{\dot{S}\left(\vartheta_0,x\right)^2}{S\left(\vartheta_0,x\right)}\,\mathrm{d}x,$$

where the normalizing factor φ_n is the inverse square root of the *Fisher information* $I_n(\vartheta_0)$ at the point ϑ_0. Above $\pi^{(n)}\left(\mathrm{d}x\right)=X^{(n)}\left(\mathrm{d}x\right)-S\left(\vartheta_0,x\right)\mathrm{d}x$ is the centered Poisson process and $\dot{S}\left(\vartheta,x\right)$ denotes the derivative of $S\left(\vartheta,x\right)$ with respect to ϑ. Based on $\Delta_n\left(\vartheta_0\right)$ we introduce the Rao score test

$$\bar{\phi}_n\left(X^{(n)}\right)=\begin{cases}1, & \text{if } \Delta_n\left(\vartheta_0\right)>z_\alpha\\ 0, & \text{if } \Delta_n\left(\vartheta_0\right)\le z_\alpha\end{cases}$$

where z_α is $1-\alpha$ quantile of standard Gaussian law, *i.e.*, $\{\zeta>z_\alpha\}=\alpha$ and $\zeta\sim\mathcal{N}(0,1)$. It is well known that if the family $\left\{{}^{(n)}_{\vartheta},\ \vartheta\in\Theta\right\}$ of distributions is *locally asymptotically normal* (LAN) at the point ϑ_0, then the test $\bar{\phi}_n\in\mathcal{K}'_\alpha$ is *locally asymptotically uniformly most powerful* (LAUMP) (or *first order efficient*), *i.e.*, for any $K>0$

$$\sup_{0\le u\le K}{}_{\vartheta_u}\tilde{\phi}_n\left(X^{(n)}\right)-{}_{\vartheta_u}\bar{\phi}_n\left(X^{(n)}\right)=o(1)$$

as $n\to\infty$ (see [14]). Moreover the power function of $\bar{\phi}_n$ at ϑ_u admits the representation

$${}_{\vartheta_u}\bar{\phi}_n\left(X^{(n)}\right)=\{\zeta>z_\alpha-u\}+o\left(1\right)\tag{1}$$

for any $u > 0$, where $\zeta \sim \mathcal{N}(0,1)$. For n large, hence the power of $\bar{\phi}_n$ approximates the envelope power function up to order $o(1)$. In this work we obtain a refinement of (1) which is given in (4). For a family $\left\{ {}^{(n)}_{\vartheta},\, \vartheta \in \Theta \right\}$ of distributions related to a Poisson process with intensity functions $\{S(\vartheta,\cdot),\, \vartheta \in \Theta\}$, the conditions of LAN when ϑ is a multidimensional parameter are obtained by Yu. A. Kutoyants, [10]. The first order efficiency of $\bar{\phi}_n$ follows from the LAN representation which implies in turn the asymptotic normalities of $\Delta_n(\vartheta_0)$ and $\Lambda_n(u)$ under both $\mathcal{H}_0$ and $\mathcal{H}_u$. Therefore the refinement of the central limit theorem by taking into account one term after the Gaussian term improves the situation. This can be done by the help of the Edgeworth type expansion of the distribution function of the stochastic integral $\Delta_n(\vartheta_0)$ and $\Lambda_n(u)$ under $\mathcal{H}_0$ and $\mathcal{H}_u$. Under certain regularity conditions related to second order asymptotic properties of the family $\left\{ {}^{(n)}_{\vartheta},\, \vartheta \in \Theta \right\}$, we can construct a *second order efficient* test, *i.e.*, a test ϕ_n^* such that for any $K > 0$,

$$\sup_{0 \le u \le K} {}_{\vartheta_u}\tilde{\phi}_n\left(X^{(n)}\right) - {}_{\vartheta_u}\,\phi_n^*\left(X^{(n)}\right) = O(\varepsilon_n^2), \tag{2}$$

for some sequence $\varepsilon_n \to 0$ (see [6]). Furthermore the probability of the first type error of ϕ_n^* is given by ${}_{\vartheta_0}\phi_n^*\left(X^{(n)}\right) = \alpha + O(\varepsilon_n^2)$. From (2) it follows that the power function of ϕ_n^* (generally a second order efficient test) approximates the envelope power function up to order $O(\varepsilon_n^2)$ and hence it works as good as $\tilde{\phi}_n$ up to this order. Second order efficiency of ϕ_n^* is related to the fact that the first and second terms in the Edgeworth expansions of the distribution functions of $\Delta_n(\vartheta_0)$ and $\Lambda_n(u)$ under the local alternative are equal up to order $O(\varepsilon_n^2)$. Hence to measure the performance of a second order efficient test (here ϕ_n^*), it is natural to consider the *power loss* of ϕ_n^* with respect to the most powerful test $\tilde{\phi}_n$, which is defined by

$$r(u) = \lim_{n\to\infty} \varepsilon_n^{-2}\left({}_{\vartheta_u}\tilde{\phi}_n\left(X^{(n)}\right) - {}_{\vartheta_u}\,\phi_n^*\left(X^{(n)}\right)\right), \tag{3}$$

for $u > 0$. To achieve this the three terms in the Edgeworth expansions of the distribution functions need to be taken into account. The main object of this work is to obtain the power loss of the score test ϕ_n^* related to an inhomogeneous Poisson process and to give the explicit representation of

$r(u)$. For the power loss results for the tests based on L, R and U statistics in the i.i.d. case, see [1], chapter 3.
The importance of Edgeworth expansions and asymptotic expansions of estimators and test statistics are well known in statistics ([8], [12], [13]). They provide better approximations for a moderate volume of observations than those obtained by the central limit theorem. On the other hand, to compare first order efficient statistical procedures, it is necessary to consider higher order terms. As under certain regularity conditions first order efficiency implies second order efficiency ([13]), then in order to measure the performance of a statistical procedure (an estimator or a statistical test) third order properties of the statistical problem are needed.

2 Edgeworth Type Expansion

The main tool used in this work is based on the Edgeworth type expansion. Hence in this section we present the conditions under which the distribution function $F_n(y) =_{\vartheta}^{(n)} \{I_n(f) < y\}$, of the stochastic integral $I_n(f) = \int_{\mathbb{A}_n} f_n(x)\ \pi^{(n)}(\mathrm{d}x)$, admits an Edgeworth type expansion for two terms after the Gaussian term, where $\pi^{(n)}(\mathrm{d}x) = X^{(n)}(\mathrm{d}x) - S(\vartheta, x)\,\mathrm{d}x$ is the centered Poisson process. We suppose that $\int_{\mathbb{A}_n} f_n(x)^2\ S(\vartheta, x)\,\mathrm{d}x = 1$. The expansion is obtained under the following two conditions:

$\mathcal{B}_1$. *There exists a sequence of real numbers* $\varepsilon_n \to 0$, *as* $n \to \infty$ *and constants* $C_r > 0, r = 3, 4, 5$, *such that*

$$\int_{\mathbb{A}_n} f_n(x)^r\ S(\vartheta, x)\,\mathrm{d}x \le C_r\ \varepsilon_n^{r-2}.$$

$\mathcal{B}_2$. *There exist constants* $\gamma \ge 5/2$ *and* $c_0 > 0$ *satisfying the inequality* $\frac{C_3}{3!}c_0 + \frac{C_4}{4!}c_0^2 + \frac{C_5}{5!}c_0^3 - \frac{1}{2} < 0$ *such that for all large* n *we have*

$$\inf_{\frac{c_0\varepsilon_n^{-1}}{2} < t < \frac{\varepsilon_n^{-2}}{2}} \int_{\mathbb{A}_n} \sin^2\left(t f_n(x)\right)\ S(\vartheta, x)\,\mathrm{d}x \ge \gamma \ln \varepsilon_n^{-1}.$$

Let us introduce the cumulants

$$\gamma_{r,n} = \int_{\mathbb{A}_n} f_n(x)^r\ S(\vartheta, x)\,\mathrm{d}x, \qquad r = 3, 4$$

and the Hermit polynomials:

$$H_2(y) = y^2 - 1, \quad H_3(y) = y^3 - 3y, \quad H_5(y) = y^5 - 10y^3 + 15y.$$

THEOREM 0.9 *Let the conditions $\mathcal{B}_1$, $\mathcal{B}_2$ be fulfilled, then uniformly in $y \in \mathbb{R}$*

$$F_n(y) = \mathcal{N}(y) - \frac{\gamma_{3,n}}{3!} H_2(y)\, n(y) - \frac{\gamma_{4,n}}{4!} H_3(y)\, n(y) - \frac{\gamma_{3,n}^2}{72} H_5(y)\, n(y) + O(\varepsilon_n^3),$$

for all n large. Here $\mathcal{N}(y)$ and $n(y)$ denote the distribution and density functions of the standard Gaussian law, respectively.

Note that $\gamma_{r,n} = O(\varepsilon_n^{r-2})$, $r = 3, 4$ by $\mathcal{B}_1$.
Proof: See [5], Page 36. The proof is a slight modification of a general theorem given by Kutoyants, where the expansion is obtained by the powers of ε_n up to order ε_n^k, $k = 1, 2, ...$ (see ([10]), page 131).

COROLLARY 0.10 *Let $0 < \alpha < 1$ be given and the conditions $\mathcal{B}_1 - \mathcal{B}_2$ be fulfilled. Then the equation $F_n(y) = 1 - \alpha + O(\varepsilon_n^3)$ has a solution $y = c_{n,\alpha}$,*

$$c_{n,\alpha} = z_\alpha + \frac{\gamma_{3,n}}{6} H_2(z_\alpha) + \frac{\gamma_{4,n}}{24} H_3(z_\alpha) + \frac{\gamma_{3,n}^2}{72} H_5(z_\alpha),$$

where z_α is $1-\alpha$ quantile of standard Gaussian law, i.e., $\{\zeta > z_\alpha\} = \alpha$ and $\zeta \sim \mathcal{N}(0, 1)$.

See [5], page 40 for proof.

3 Second Order Efficiency and Power Loss

Let $\beta(u, \phi_n)$ denote the power of a test ϕ_n at the local alternative $\vartheta_u = \vartheta_0 + \varphi_n u$, *i.e.*,

$$\beta(u, \phi_n) =_{\vartheta_u} \phi_n(X^{(n)}).$$

Below $S^{(j)}(\vartheta, x)$ denotes the jth derivative of $S(\vartheta, x)$ with respect to ϑ. We write $\dot{S}(\vartheta, x)$ and $\ddot{S}(\vartheta, x)$ for the first and second derivatives, respectively.

Under certain regularity conditions slightly weaker than $\mathcal{D}_1 - \mathcal{D}_3$ presented below, the score test

$$\phi_n^*\left(X^{(n)}\right) = \begin{cases} 1, & \text{if } \Delta_n(\vartheta_0) > c_n \\ 0, & \text{if } \Delta_n(\vartheta_0) \le c_n, \end{cases}$$

based on

$$\Delta_n(\vartheta_0) = \varphi_n \int_{\mathbb{A}_n} \frac{\dot{S}(\vartheta_0, x)}{S(\vartheta_0, x)} \pi^{(n)}(\mathrm{d}x)$$

with

$$c_n = z_\alpha - \frac{\gamma_{3,n}}{6}\left(1 - z_\alpha^2\right), \quad \gamma_{3,n} = \varphi_n^3 \int_{\mathbb{A}_n} \frac{\dot{S}(\vartheta_0, x)^3}{S(\vartheta_0, x)^2}\, \mathrm{d}x$$

is *second order efficient*, *i.e.*, for any $K > 0$ it satisfies

$$\sup_{0 \le u \le K} \beta(u, \phi_n^*) - \beta(u, \tilde{\phi}_n) = O(\varepsilon_n^2),$$

where the sequence $\varepsilon_n \to 0$ as $n \to \infty$. For proof see [6], Theorem 6.
Now we obtain the explicit representation of the power of ϕ_n^* (up to order $O(\varepsilon_n^2)$) under the local alternative $\vartheta_u = \vartheta_0 + \varphi_n u$, for $u > 0$. By Edgeworth type expansion under the local alternative the power of ϕ_n^* under the local alternative admits the following representation

$$\beta(u, \phi_n^*) = \mathcal{N}\left(\frac{m_n(u) - c_n}{\eta_n}\right) - \frac{\gamma_{3,n}(u)}{6}\left(1 - (u - z_\alpha)^2\right) n(u - z_\alpha) + O(\varepsilon_n^2)$$

where

$$m_n(u) =_{\vartheta_u} \Delta_n(\vartheta_0) = \varphi_n \int_{\mathbb{A}_n} \frac{\dot{S}(\vartheta_0, x)}{S(\vartheta_0, x)} \left(S(\vartheta_u, x) - S(\vartheta_0, x)\right)\, \mathrm{d}x,$$

$$\eta_n^2 =_{\vartheta_u} \left(\Delta_n(\vartheta_0) - m_n(u)\right)^2 = \varphi_n^2 \int_{\mathbb{A}_n} \frac{\dot{S}(\vartheta_0, x)^2}{S(\vartheta_0, x)^2} S(\vartheta_u, x)\, \mathrm{d}x,$$

$$\gamma_{3,n}(u) = \frac{\varphi_n^3}{\eta_n^3} \int_{\mathbb{A}_n} \frac{\dot{S}(\vartheta_0, x)^3}{S(\vartheta_0, x)^3} S(\vartheta_u, x)\, \mathrm{d}x.$$

Now using the Taylor expansion

$$S(\vartheta_u, x) = S(\vartheta_0, x) + \varphi_n u \dot{S}(\vartheta_0, x) + \frac{\varphi_n^2 u^2}{2} \ddot{S}(\vartheta_0, x) + \frac{\varphi_n^3 u^3}{3!} S^{(3)}(\vartheta_n, x),$$

we obtain the following representation

$$\beta\,(u,\phi_n^*) = \mathcal{N}\,(u-z_\alpha) + Q_n(u)\,n(u-z_\alpha) + O(\varepsilon_n^2), \tag{4}$$

for any $u>0$, where the polynomial (in u)

$$Q_n(u) = \frac{u(z_\alpha-2u)}{6}\,\gamma_{3,n} + \frac{\varphi_n^3\,u^2}{2}\int_{\mathbb{A}_n}\frac{\dot S(\vartheta_0,x)\,\ddot S(\vartheta_0,x)}{S(\vartheta_0,x)}\,\mathrm{d}x,$$

is of order $O(\varepsilon_n)$. The equation (4), refines the first order representation (1). Notice also that the the second order efficiency of ϕ_n^* implies that we have the same representation for the power of $\tilde\phi_n$. As the power function of a second order efficient test agrees with that of the most powerful test up to order $O(\varepsilon_n^2)$, hence it is natural to consider the *power loss* of ϕ_n^*, which is defined for any $u>0$ by

$$r(u) = \lim_{n\to\infty}\varepsilon_n^{-2}\left(\beta(u,\tilde\phi_n) - \beta\,(u,\phi_n^*)\right).$$

We consider the following conditions:

$\mathcal{D}_1$. The intensity function $S(\vartheta,x)$ is four times differentiable with respect to ϑ in a right neighborhood of ϑ_0.

$\mathcal{D}_2$. The conditions $\mathcal{B}_1$ and $\mathcal{B}_2$ are satisfied for the stochastic integrals $\Delta_n(\vartheta_0)$ and $\Lambda_n(u)$ under $\mathcal{H}_0$ and $\mathcal{H}_u$ with some sequence $\varepsilon_n\to 0$,

$\mathcal{D}_3$. There exists some functions $f_j(x), j=0,...,4,\ x\in\mathbb{A}_n$ not depending on ϑ such that $S(\vartheta,x)\ge f_0(x), S^{(j)}(\vartheta,x)\le f_j(x), j=0,...,4$ for all $x\in\mathbb{A}_n$ and all ϑ in a right neighborhood of ϑ_0. We suppose also that

$$\varphi_n^k\int_{\mathbb{A}_n}\frac{f_1(x)^k}{f_0(x)^{k-1}}\,\mathrm{d}x = O(\varepsilon_n^{k-2}),\quad k=2,3,4$$
$$\varphi_n^{2j}\int_{\mathbb{A}_n}\frac{f_j(x)^2}{f_0(x)}\,\mathrm{d}x = O(\varepsilon_n^{2j-2}),\quad j=2,3,4.$$

Example 1. Let $X^{(n)}$ be a realization of a Poisson process on the set $\mathbb{A}_n=[\,0,n]$ with positive intensity function $S(\vartheta,x)=\vartheta S(x)+\lambda$ (amplitude parameter) or $S(\vartheta,x)=S(\vartheta+x)+\lambda$ (phase parameter), where $S(\cdot)$ is a four

times differentiable periodic function and $\lambda > 0$ (dark current) is a known constant. In both cases the condition $\mathcal{D}_3$ is satisfied with $\varphi_n \sim C\, n^{-1/2}$ for some $C > 0$ and $\varepsilon_n = n^{-1/2}$. For the frequency modulation model $S(\vartheta, x) = S(\vartheta\, x) + \lambda$ we have $\varphi_n \sim C\, n^{-3/2}$ and $\varepsilon_n = n^{-1/2}$.

Let us denote as J_n the quantity:

$$J_n = \varphi_n^4 \int_{\mathbb{A}_n} \frac{\left(\dot{S}^2 - S\ddot{S}\right)^2}{S^3} \mathrm{d}x - \left(\varphi_n^3 \int_{\mathbb{A}_n} \frac{\dot{S}\left(\dot{S}^2 - S\ddot{S}\right)}{S^3} \mathrm{d}x \right)^2, \qquad (5)$$

where we used the abbreviations $S = S\left(\vartheta_0, x\right), \dot{S} = \dot{S}\left(\vartheta_0, x\right), \ddot{S} = \ddot{S}\left(\vartheta_0, x\right)$. Note that $J_n = O(\varepsilon_n^2)$, by $\mathcal{D}_3$. We have the following theorem:

THEOREM 0.11 *Let the conditions $\mathcal{D}_1 - \mathcal{D}_3$ be fulfilled. Then the power loss of ϕ_n^* with respect to the most powerful test $\tilde{\phi}_n$ is equal to*

$$r(u) = \frac{u^3\, n(u - z_\alpha)}{8} \lim_{n\to\infty} \left(\varepsilon_n^{-2}\, J_n\right),$$

for any $u > 0$.

Proof: We sketch the main points of the proof and omit the details. By $\mathcal{D}_2$ we can write the following third order expansions:

$$\begin{aligned}
\beta\left(u, \phi_n^*\right) \quad &= \mathcal{N}(a_n) + \tfrac{\gamma_{3,n}(u)}{3!} H_2(a_n)\, n(a_n) - \tfrac{\gamma_{4,n}(u)}{4!} H_3(a_n)\, n(a_n) - \\
&\quad - \tfrac{\gamma_{3,n}^2(u)}{72} H_5(a_n)\, n(a_n) + O(\varepsilon_n^3) \\
\beta(u, \tilde{\phi}_n) \quad &= \mathcal{N}(A_n) + \tfrac{\gamma'_{3,n}(u)}{3!} H_2(A_n)\, n(A_n) - \tfrac{\gamma'_{4,n}(u)}{4!} H_3(A_n)\, n(A_n) - \\
&\quad - \tfrac{\gamma'_{3,n}(u)^2}{72} H_5(A_n)\, n(A_n) + O(\varepsilon_n^3),
\end{aligned}$$

where $a_n = \eta_n^{-1}\left(m_n(u) - c_n\right)$, $A_n = \sigma_n(u)^{-1}\left(\mu_n(u) - b_n(u)\right)$. Here $\mu_n(u) =_{\vartheta_u} \Lambda_n(u)$, $\sigma_n^2(u) =_{\vartheta_u} \left(\Lambda_n(u) - \mu_n(u)\right)^2$ and

$$\begin{aligned}
\gamma_{r,n}(u) &= \frac{\varphi_n^r}{\eta_n^r} \int_{\mathbb{A}_n} \frac{\dot{S}\left(\vartheta_0, x\right)^r}{S\left(\vartheta_0, x\right)^r} S\left(\vartheta_u, x\right)\, \mathrm{d}x, \qquad r = 3, 4 \\
\gamma'_{r,n}(u) &= \frac{1}{\sigma_n(u)^r} \int_{\mathbb{A}_n} \left(\ln \frac{S\left(\vartheta_u, x\right)}{S\left(\vartheta_0, x\right)} \right)^r S\left(\vartheta_u, x\right) \mathrm{d}x, \qquad r = 3, 4.
\end{aligned}$$

For simplicity we introduce the notations:

$$I(r_0, r_1, r_2) = \int_{\mathbb{A}_n} \frac{\dot{S}\left(\vartheta_0, x\right)^{r_1} \ddot{S}\left(\vartheta_0, x\right)^{r_2}}{S\left(\vartheta_0, x\right)^{r_0}} \mathrm{d}x$$

for nonnegative integers r_0, r_1, r_2. Using the Taylor expansions of $S\left(\vartheta_u, x\right)$ and $\ln \frac{S(\vartheta_u, x)}{S(\vartheta_0, x)}$, we get

$$\begin{gathered} A_n - a_n = \tfrac{u^3}{8}\, \varphi_n^4\, I(1,0,2) - \tfrac{2u^3 - 2u^2 z_\alpha - u(1-z_\alpha^2)}{4}\, \varphi_n^4\, I(2,2,1) + \\ \tfrac{9u^3 - 12u^2 z_\alpha - 6u(1-z_\alpha^2)}{24}\, \varphi_n^4\, I(3,4,0) + \tfrac{9u^3 - 6u^2 z_\alpha + 2u(1-z_\alpha^2)}{24}\, \varphi_n^6\, I^2(2,3,0) \\ + \tfrac{6u^3 - 6u^2 z_\alpha - 3u(1-z_\alpha^2)}{12}\, \varphi_n^6\, I(2,3,0)\, I(1,1,1) - \tfrac{u^3}{8}\, \varphi_n^6\, I(1,1,1)^2 + O(\varepsilon_n^3) \end{gathered}$$

and

$$\begin{gathered} \gamma'_{3,n}\left(u\right) - \gamma_{3,n}\left(u\right) = \tfrac{3u}{2}\, \varphi_n^4\, I(2,2,1) + \tfrac{3u}{2}\, \varphi_n^6\, I(2,3,0)^2 - \tfrac{3u}{2}\, \varphi_n^4\, I(3,4,0) - \\ - \tfrac{3u}{2}\, \varphi_n^6\, I(2,3,0)\, I(1,1,1) + O(\varepsilon_n^3). \end{gathered}$$

On the other hand one can show that

$$\begin{aligned} \gamma_{4,n}(u) &= \gamma_{4,n} + O(\varepsilon_n^3), \quad \gamma_{3,n}(u)^2 = \gamma_{3,n}^2 + O(\varepsilon_n^3) \\ \gamma'_{4,n}(u) &= \gamma_{4,n} + O(\varepsilon_n^3), \quad \gamma'_{3,n}(u)^2 = \gamma_{3,n}^2 + O(\varepsilon_n^3). \end{aligned}$$

Combining these results we obtain for any $u > 0$

$$\beta(u, \tilde{\phi}_n) - \beta\left(u, \phi_n^*\right) = \frac{u^3\, n(\Delta)}{8}\, S_n + O(\varepsilon_n^3)$$

where $\Delta = u - z_\alpha$ and

$$S_n \;=\; \varphi_n^4\left(\, I(1,0,2) - 2\, I(2,2,1) + I(3,4,0)\right) - \varphi_n^6\, \left(I(2,3,0) - I(1,1,1)\right)^2.$$

By substituting $I(r_1, r_2, r_3)$ it can be shown that S_n is equal to J_n, see (2). Now we have

$$\varepsilon_n^{-2} \left(\beta(u, \tilde{\phi}_n) - \beta\left(u, \phi_n^*\right)\right) = \frac{u^3\, n(\Delta)}{8}\, \varepsilon_n^{-2}\, J_n + O(\varepsilon_n),$$

which completes the proof of the theorem.

Example 2. (Frequency parameter) In this example we consider a strongly inhomogeneous case with a nonclassical rate $n^{-3/2}$ (instead of $n^{-1/2}$ in the i.i.d. case). Suppose that we observe a realization $X^{(n)}$ of a Poisson process on the set $\mathbb{A}_n = [\,0, n]$, $n = 1, 2, \cdots$ with periodic intensity function

$$S(\vartheta, x) = e^{\sin(\vartheta\, x)}, \qquad \vartheta > 0.$$

We consider the hypothesis $\mathcal{H}_0 : \vartheta = \vartheta_0$ against the one sided alternative $\mathcal{H}_1 : \vartheta > \vartheta_0$. The conditions $\mathcal{D}_1 - \mathcal{D}_3$ are satisfied (see [6], pp. 205-207) with $\varepsilon_n = n^{-1/2}$ and $\varphi_n \sim C\, n^{-3/2}$ where

$$C^{-2} = \frac{1}{3\,\tau} \int_0^\tau \cos^2(\vartheta_0\, x)\, e^{\sin(\vartheta_0\, x)}\, \mathrm{d}x, \qquad \tau = \frac{2\pi}{\vartheta_0}.$$

Now one can show easily that

$$\lim_{n\to\infty} \left(\varepsilon_n^{-2}\, J_n\right) = \frac{C^4}{5\,\tau} \int_0^\tau \sin^2(\vartheta_0\, x)\, e^{\sin(\vartheta_0\, x)}\, \mathrm{d}x$$

and hence the power loss is equal to

$$r(u) = \frac{9\,\tau\, u^3\, n(u - z_\alpha)}{40} \frac{\int_0^\tau \sin^2(\vartheta_0\, x)\, e^{\sin(\vartheta_0\, x)}\, \mathrm{d}x}{\left(\int_0^\tau \cos^2(\vartheta_0\, x)\, e^{\sin(\vartheta_0\, x)}\, \mathrm{d}x\right)^2},$$

for any $u > 0$.

Acknowledgment. The author wishes to express his gratitude to Yu. A. Kutoyants for his comments and recommendations which helped the author in improving the presentation of the results in this paper.

References

[1] Bening, V.E. (2000). *Asymptotic Theory of Testing Statistical Hypotheses.* VSP, Utrecht.

[2] Dachian, S., Kutoyants, Yu.A. (2006) *Hypotheses testing: Poisson versus self-exciting*, Scandinavian J. of Statistics, **33**, 391-408.

[3] Daley, D.J., Vere-Jones, D. (1988). *An introduction to the theory of point processes.* Springer, New York.

[4] Davies, R. (1977). *Testing the hypothesis that a point process is Poisson*, Adv. in Appl. Probab., **9**, 724-746.

[5] Fazli, Kh. (2006). *Tests d'hypothèses asymptotiquement optimaux pour les processus de Poisson non homogènes*, Thèse de doctorat de l'Université du Maine .

[6] Fazli, Kh. (2007). *Second order efficient test for inhomogeneous Poisson Processes*, Statistical Inference for Stochastic Processes. **10**, 181-208.

[7] Fazli, Kh. (2008). *Hypotheses Testing for a Multidimensional Parameter of Inhomogeneous Poisson Processes*, To appear in Statistics.

[8] Ghosh, J.K. (1994). *Higher Order Asymptotics*, NSF-CBMS Regional Conference Series in Probability and Statistics, Volume 4 .

[9] Ingster Yu. I., Kutoyants Yu. A. (2007) *Nonparametric hypothesis testing for intensity of the Poisson process*, Mathematical Methods of Statistics, **16**, 217-245.

[10] Kutoyants, Yu.A. (1998). *Statistical Inference for Spatial Poisson Processes*. Springer, New York.

[11] Noether, G.E. (1955). *On a Theorem of Pitman*. Ann. Math.Statist. **26**, 64-68.

[12] Pfanzagl, J. (1980). *Asypmtotic Expansions in Parametric Staticstical Theory*, Developments in Statistics. VOL.**3**, Academic Press. Inc.

[13] Pfanzagl, J. (1985). *Asypmtotic Expansions for General Staticstical Models*, Springer, New york.

[14] Roussas, G.G. (1972). *Contiguity of Probability Measure*. Cambridge University Press, Cambridge.

Detection of Outliers and Leverage points and Robust Regression Analysis of Iranian Urban Household Income and Expenditure Data

M. Ganjali[¶], and S. Eftekhari Mahabadi

Department of Statistics, Shahid Beheshti University

Abstract: In application of regression models some steps are very essential to be considered for analyzing large data sets. These include (a) examination of the overall shape of the data for the need of a transformation on the response of interest (b) checking on the important features, including symmetry and departures from the assumption of the homogeneity for the response's variance (c) investigating or detecting observations that have abnormal distances from other values, outliers or strange observations, these often contain valuable information about the process under investigation or the data gathering and recording process. After all of these, before considering the possible elimination of outliers, one should try to understand why they appeared and then try to use approaches that weigh the effect of these observations down in the analysis. Robust regression is an important tool for analyzing data that are contaminated with outliers. It can be used to detect outliers and to provide robust results in the presence of outliers. This paper provide an overview of robust regression methods to illustrate their application toward fitting regression models. We shall use the most commonly robust regression techniques including M estimation, LTS estimation, S estimation and MM estimation for the urban household income and expenditure data collected, by Statistical Center of Iran (including 14175 households) in 1385. For these data for the first time some robust regression models will be compared using normal curvature of the likelihood displacement.

Keywords: Breakdown point; Outlier; Influential point; Likelihood Displacement; Robust Regression; Household Income and Expenditure.

1 Introduction

Regression analysis is an important statistical tool that is routinely applied in most sciences. Mostly some desired assumptions are made on the regression structure to ease the model formulation and computation. Some of the

[¶]Invited speaker

commonly proposed assumptions are normality of the response variable and homogeneity of the response's variance. It is often the case that we need to use some kind of response transformation to gain normality assumption before modelling the data (Box and Cox, 1964). Also the variance homogeneity should be checked out before going through analysis. One useful approach might be looking at the variation of response in the different levels of model covariates (Cook and Weisberg, 1983).

Out of many possible regression techniques for fitting, the least squares (LS) method has been generally adopted because of tradition and ease of computation. However, there is presently a widespread awareness of the dangers posed by the occurrence of outliers (Rousseeuw and Leroy, 1987). Outliers occur very frequently in real data, and they often go unnoticed because now a days much data is processed by computers, without careful monitoring. Not only can the response variable be outlying, but also the explanatory part can be, leading to so-called leverage points. Both of these, outliers and leverage points may totally spoil an ordinary LS analysis. Often, such influential points remain hidden to the user, because they do not always show up in the usual LS residual plots.

To be protected against outliers and leverage points, there are two useful procedures, one the regression diagnostics (see Cook, 1977 and 1979; Blesley et al., 1980) and the other robust regression (see Huber, 1973 and 1981; Rousseeuw, 1984; Rousseeuw and Yohai, 1984). They both have the same goal but they proceed in the opposite order. In regression diagnostics context, one first fit a regression model then search for potential outliers to refit the model with the resulting good data while in robust regression, a model appropriate for the majority of the data is fitted and then those observations with large residuals from the robust model are detected as outliers.

We can also consider local influence yet another procedure to access sensitivity of the model fitting through different model assumptions which might be conducted through omitting or down weighing suspicious points to see if they are outlier or influential using influence graph. In this paper a new approach for using normal curvature of the likelihood displacement as a comparison tool between different robust regressions, will be presented.

In the next Section we will briefly describe Iranian Urban Household Income and Expenditure data. As the first step toward regression analysis we will check for the appropriate transformation and the homogeneity assumption of the response variable of interest. In Section 3, different methods for robust regression analysis and also the local influence approach for model diagnostics are given. In section 4 robust regression results for our data and their comparison with the LS estimates are discussed. A comparison of some robust regression methods using normal curvature is also presented. In Section 5 we give a brief conclusion.

2 Iranian Urban Household Income and Expenditure Data

The data consist of Iranian urban household Income and Expenditure survey which is conducted by Statistical Center of Iran in 1385 as a survey based on complex random sampling with a high sample size that includes a great amount of information about households. These data consist of 14175 urban households in the sample. The response variable of interest in the present application is the household total expenditure which has a mean of 64815383 Rials and standard error of 58278416. Since we wish to fit a regression model for this response variable, its wide range indicates a necessary investigation for the appropriate Box-Cox transformation. Figure (1) displays the profile log likelihood against different values of λ (the appropriate power for the transformation). It is apparent that the log likelihood as a function of λ is maximized around $\lambda = 0$ which is also included in the 95% confidence interval. Hence, the logarithmic transformation which corresponds to $\lambda = 0$ in Box-Cox transformation seems to be adequate. So, the logarithm of total household expenditure as the regression response variable Y (mean=17.705 and Standard deviation= 0.744) would be used. The set of covariates considered in this study are some characteristics of the head of household including Age as a continuous variable (mean=46.115, Standard deviation=14.597), Gender, Literacy status, Employment status, Marital status and some attributes of the household such as Home possession and Family size. Explanatory categorical variables, their levels and percentage of each level of each variable are presented in Table (1). Box

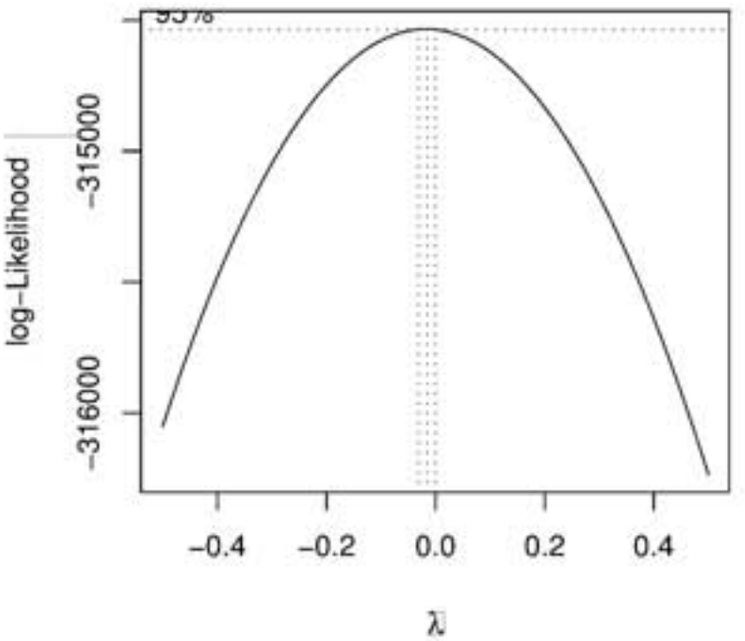


Figure 1: profile log likelihood against different values of λ.

Family size	Perc.	Home possession	Perc.	Marital status	Perc.	Employment status	Perc.	Gender	Perc.	Literacy	Perc.
1	0.040	Other	0.093	Single	0.016	Other	0.241	Male	0.907	Literate	0.803
2	0.134										
3	0.212	Owner	0.675	Married	0.888	Employed	0.737				
4	0.259							Female	0.092	Illiterate	0.196
5	0.176	Rented	0.231	Widowed or divorced	0.091	Unemployed	0.022				
≥6	0.178										

Table (1): Explanatory categorical variables, their levels and percentage of each level of each variable.

plots of Y in different levels of each categorical covariates do not show any evidence against variance homogeneity, hence, hereafter we will consider a regression model with the assumption of homogeneity for response variance.

3 Robust Regression and Local Influence

The classical linear regression considers the following model

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi} + e_i, \quad i = 1, \ldots, n$$

where n is the sample size, y_i is the response variable and $x_{1i}, \ldots, x_{pi}$ are

called explanatory or independent variables for the i th individual. Also e_i is the error term which is assumed to be normally distributed with zero mean and some unknown variance σ^2.

4 Robust Regression

The most popular regression estimates date back to Gauss and Legendre (see Plackett, 1972 and Stigler, 1981 for some historical discussions) which corresponds to minimizing sum of squared residuals r_i^2 as

$$\min_{\hat{\beta}} \sum_{i=1}^{n} r_i^2, \quad where \;\; r_i = y_i - \hat{y}_i$$

hence yielding the ordinary LS estimates. The reason for popularity of this method is the easy and fast computation process that can be performed using some matrix algebra even without use of computers (as it was around 1800). This procedure seems optimal when the error distribution is assumed as normal (see the citation in Huber, 1972 and Le Cam, 1986). Later, some people began to realize that real data usually do not satisfy the classical assumptions which may affect the regression results substantially (see, e.g, Student, 1927; Pearson, 1931; Box, 1953 and Tukey, 1960). Two kinds of strange points can be occurred which have a great influence on LS regression analysis, one that is strange in the y direction (outlier) and the other which is strange in the x direction (leverage). Diagnostic measures are certain quantities computed from the data with the purpose of revealing influential points, after which these outliers can be removed or corrected, followed by an LS analysis on the remaining cases (see Atkinson, 1987).

The other approach to face outliers is robust regression, which tries to obtain estimators that are not highly affected by outliers which make use of estimators with high breakdown point (for definition vide, Hodges, 1976; Hample, 1971; Donoho and Huber, 1983). For the LS method the breakdown point is equal to $\frac{1}{n}$, which means that one outlier is sufficient to disturb the behavior of T.

Edgeworth (1887) introduced L_1 regression which minimizes sum of absolute residuals. The next step in this direction was the use of M

estimators (Huber, 1973 and 1981) which is based on replacing the squared residuals by another less rapidly increasing function $\rho(r_i)$ which is symmetric with a unique minimum at zero, yielding

$$\min_{\hat{\beta}} \sum_{i=1}^{n} \rho(r_i).$$

Differentiating this expression with respect to β and using standrized residuals reduces the above expression to the solution of

$$\sum_{i=1}^{n} \psi(r_i/\hat{\sigma})x_i = 0$$

where $\hat{\sigma}$ must be estimated simultaneously. Different ψ functions can be used in this method such as Huber-type ψ function (Huber, 1964) defined as

$$\psi(t) = \begin{cases} t & if \quad |t| < b \\ bsgn(t) & if \quad |t| \geq b \end{cases}$$

where b is a constant that according to the above definition, the standardized residuals more than that in absolute value will be considered as b in the summation (to protect against outliers). Hample (1974) defined a function that protects the fit even more strongly against outlying observations which is called a three redescending M estimator. Although M estimation is not robust with respect to leverage points (the break down value is 0), it is still used extensively in analyzing data for which it can be assumed that the contamination is mainly in the response direction.

Least Trimmed Squares (LTS) estimation is a high breakdown value method introduced by Rousseeuw (1984), given by

$$\min_{\hat{\beta}} \sum_{i=1}^{h} (r^2)_{i:n}$$

where $(r^2)_{1:n} \leq \ldots \leq (r^2)_{n:n}$ are the ordered squared residuals and h is defined in the range $\frac{n}{2} + 1 \leq h \leq \frac{3n+p+1}{4}$. This formula is very similar to LS, the only difference being that the largest squared residuals are not used

in the summation, thereby allowing the fit to stay away from the outliers. The breakdown point for LTS estimates attain $\frac{n-h}{n}$ and for $h = [\frac{n}{2}] + [\frac{p+1}{2}]$, this method reaches the maximal possible value for the breakdown point $(\frac{[\frac{n-p}{2}+1]}{n})$. While the LTS criterion is easily described, the mechanics of fitting the LTS estimator are complicated (see, for example, Rousseeuw and Leroy, 1987) and there is no simple formula for the standard errors of coefficients.

S estimation is a high breakdown value method introduced by Rousseeuw and Yohai (1984). It has a higher statistical efficiency than LTS estimation with the same breakdown value. The S estimator is defined by minimization of the dispersion of the residuals:

$$\min_{\hat{\beta}} S(r_1(\beta), \ldots, r_n(\beta))$$

with final scale estimate $\hat{\sigma} = S(r_1(\hat{\beta}), \ldots, r_n(\hat{\beta}))$ where the dispersion $S(r_1(\beta), \ldots, r_2(\beta))$ is the solution of

$$\frac{1}{n}\sum_{i=1}^{n}\rho(\frac{r_i}{s}) = K \tag{1}$$

where K is often put equal to $E_\phi[\rho]$, where ϕ is the standard normal density. The function ρ must satisfy the following conditions:

(1) ρ is symmetric and continuously differentiable, and $\rho(0) = 0$.

(2) There exist $c > 0$ such that ρ is strictly increasing on $[0, c]$ and constant on $[c, \infty)$.

If there happen to be more than one solution to (1), then put $s(r_1, \ldots, r_n)$ equal to the supremum of the set of solutions; this means $S(r_1, \ldots, r_n) = sup\{s; 1/n\sum\rho(\frac{r_i}{s}) = K\}$. If there exist no solution to (2), then put $S(r_1, \ldots, r_n) = 0$. This estimator is called an S estimator because it is derived from a scale statistic in an implicit way. Because of condition (2), $\psi(t) = \rho'(t)$ will always be zero for certain values of t on, so ψ is redescending. one possibility for ψ is Tukey's biweight function defined as

$$\psi(t) = \begin{cases} t(1-(t/c)^2)^2 & if \quad |t| < c \\ 0 & if \quad |t| \geq c \end{cases}$$

or the hyperbolic tangent. It can be shown that the breakdown point of S estimators is 50% when having an extra condition on the ρ function, $\frac{k}{\rho(c)} = \frac{1}{2}$ which can be easily fulfilled. In the case of biweight function, it is achieved by taking $c = 1.547$ (see Rousseeuw and Leroy, 1987).

Yohai (1985) introduced a new improvement toward higher efficiency for high breakdown estimators like LTS which combines high breakdown value estimation and M estimation. He called this new class MM estimators. Yohai's estimators are defined in three stages. In the first stage, a high breakdown estimate β^* is calculated, such as LTS. Then, an M estimate of scale s_n with 50% breakdown is computed on the residuals $r_i(\beta^*)$ from the robust fit. Finally, the MM estimator $\hat{\beta}$ is defined as any solution of

$$\sum_{i=1}^{n} \psi(r_i(\beta)/s_n)x_i = 0$$

which satisfies $S(\beta) \leq S(\beta^*)$, where $S(\beta) = \sum_{i=1}^{n} \rho(r_i(\beta)/s_n)$, The function ρ must satisfy the same conditions stated above for S estimators. This implies that $\psi = \rho'$ has to be properly redescending. some possibilities are three part redescenders, Tukey's biweight, or the hyperbolic tangent. The trick is that this ρ may be quite different from that of scale estimator s_n of the second stage, because the first and the second stage must achieve the high breakdown point whereas the third stage is allowed to aim for a high efficiency. Indeed, Yohai showed that MM estimators inherit the 50% breakdown point of the first stage and that they also possess the exact fit property. Moreover, he proved that MM estimators are highly efficient when the errors are normally distributed. It has both the high breakdown property and a higher statistical efficiency than S estimation.

4.1 Local Influence

Statistical conclusions can be viewed as the end result of synthesis of the relevant information provided by the observed data and the prior information provided by the model which is usually a plausible, but necessarily imprecise, description of the actual process that generated the data. The discussions presented by Cook (1986) are based on the informal notion that important conclusions should not depend critically on the hypothe-

sized model or unusual aspects of the data. If a minor modification of an approximate description seriously influences key results of an analysis, there is surely cause for concern. On the other hand, if such modifications are found to be unimportant, the sample is robust with respect to the induced perturbations and our ignorance of the precise model will do no harm (Barnard, 1980). So, an assessment of the influence of minor perturbations of the model is important (vide, Cook, 1986).

Generally, one introduces perturbations into the model through the $q \times 1$ vector ω which is restricted to some open subset Ω of R^q. Let $L(\theta|\omega)$ denote the log-likelihood corresponding to the perturbed model for a given ω in Ω. For a given set of observed data, where θ is a $p\times 1$ vector of unknown parameters, we assume that there is an ω_0 in Ω such that $L(\theta) = L(\theta|\omega_0)$ for all θ. Finally, Let $\hat{\theta}$ and $\hat{\theta}_\omega$ denote the maximum likelihood estimators under $L(\theta)$ and $L(\theta|\omega)$. To assess the influence of varying ω throughout Ω, the likelihood displacement defined as:

$$LD(\omega) = 2[L(\hat{\theta}) - L(\hat{\theta}_\omega)],$$

is considered. An obvious way to see if perturbations of the model influence key results of the analysis is to compare the results derived from the original and perturbed models using an influence graph which is a geometric surface formed by the values of the $(q + 1) \times 1$ vector

$$\alpha(\omega) = (\omega, LD(\omega))'$$

as ω varies through Ω. To characterize the behaviour of an influence graph around ω_0, geometric normal curvature is used. Some direction $l_{q\times 1} \in R^q$ $(||l|| = 1)$ is chosen to see the normal curvature C_l at ω_0 in the direction of l. The expression for C_l will reduce to

$$C_l = 2|l^T \Delta^T (L'')^{-1} \Delta l|$$

where Δ is a $p \times q$ matrix with elements

$$\Delta_{ij} = \frac{\partial^2 L(\theta|\omega)}{\partial\theta_i \partial\omega_j}|_{\{\theta=\hat{\theta},\ \omega=\omega_0\}}$$

and $-L''$ is the observed information matrix for the postulated model ($\omega = \omega_0$). There are several ways in which C_l might be used to study $\alpha(\omega)$ in practice, two possible options are $C_{max} = max_l C_l$ and $C_{min} = min_l C_l$ which correspond to maximum and minimum absolute eigenvalues of $\Delta^T (L'')^{-1} \Delta$ (Cook, 1986).

5 Results for Expenditure Data

Results of using LS, M, LTS, S, MM methods toward model fitting for data described in Section 2, are presented in Table (2) (computations are performed in software R.2.6.1). The M regression used in obtaining these results is based on the Huber function with $b = 1.345$. For the LTS, $h = [\frac{n}{2}] + [\frac{p+1}{2}]$ which attains 50% breakdown value as stated in Section 3.1. Tukey's biweight function with $c = 4.685$ is used in the S regression. The presented MM regression uses a specific set of options which ensures that the estimator has a high breakdown point. The initial set of coefficients and the final scale are selected by an S estimator based on Tukey's biweight function with $c = 1.547$ this gives (for $n >> p$) breakdown point 0.5 and The final estimator is an M-estimator with Tukey's biweight (c=4.685) and fixed scale that will also inherit this breakdown point since $4.685 > 1.547$.

The parameter estimations indicate the significant effect of all model covariates but head of household's Gender in LS, M and MM regression. The parameter estimates are generally increased in the corresponding magnitude in all robust methods comparing with LS results.

M and MM regression use weighted observations corresponding to the ψ function used in their method for data analysis (strange observations will be given lower weights). Figure (2) displays the corresponding weights in M and MM regression for all observations. The minimum weight for M regression is 0.272 and around 82% of observations are assigned weight 1. The minimum weight in MM method is zero which is for observation number 10087 and also there are 47 observations which are weighted less than 0.27. As Figure (2) shows, M regression assigns weights of more than 0.8 to about 90 percent of observations while MM regression assigns it for about 86 percent of the observations. This makes clear the higher breakdown point achieved by using MM regression due to lower weights

Par.		LS	M	LTS	S	MM
		Est. (S. E.)	Est. (S. E.)	Est.	Est.	Est. (S. E.)
Intercept		16.673(0.054)	16.656(0.055)	16.141	15.797	16.654(0.055)
Age		0.008(0.001)	0.008(0.001)	0.010	0.049	0.008(0.001)
Gender	Male*	--	--	--	--	--
	Female	-0.007(0.035)	0.021(0.035)	-0.046	-0.281	0.029(0.035)
literacy status	Literate*	-	--	--	--	--
	Illiterate	-0.609(0.016)	-0.612(0.016)	-0.920	-0.607	-0.612(0.016)
Marital status	Single*	--	--	--	--	--
	Married	-0.111(0.042)	-0.118(0.043)	0.041	-0.446	-0.121(0.043)
	Widowed or Divorced	-0.143(0.051)	-0.166(0.052)	0.389	-0.173	-0.178(0.052)
Employment status	Other*	--	--	--	--	--
	Employed	0.140(0.017)	0.137(0.017)	0.119	-0.288	0.139(0.017)
	Unemployed	-0.161(0.041)	-0.169(0.042)	0.625	-1.315	-0.174(0.042)
Home possession	Other*	--	--	--	--	--
	Owner	0.240(0.021)	0.242(0.021)	-0.157	-0.980	0.238(0.021)
	Rented	0.069(0.022)	0.072(0.022)	-0.993	-0.146	0.069(0.022)
Family size	1 member*	--	--	--	--	--
	2 members	0.412(0.035)	0.415(0.036)	2.000	0.943	0.417(0.036)
	3 members	0.572(0.035)	0.568(0.036)	1.443	1.127	0.572(0.036)
	4 members	0.707(0.035)	0.707(0.036)	0.779	1.142	0.710(0.036)
	5 members	0.722(0.036)	0.723(0.037)	1.619	1.573	0.728(0.037)
	At least 6 members	0.789(0.036)	0.798(0.037)	1.729	1.954	0.804(0.037)
Scale parameter		0.664	0.649	0.884	0.648	0.648

Table (2): Results of using LS, M, LTS, S, MM regressions for Iranian urban household income and expenditure data (* : baseline level).

assigned.

Use of diagnostic measures such as Cook's D_i defined as $D_i = ||\hat{Y} - \hat{Y}_{(i)}||^2/p\sigma^2$ by Cook (1977) might not reflect completely the nature of influential cases since it only considers zero weight to access influential points ($\hat{Y}_{(i)}$ are fitted values based on the data without case i). To investigate this specific concern, Cook (1986) proposed the following slightly more general version of D_i, $D_i(\omega) = ||\hat{Y} - \hat{Y}_\omega||^2/p\sigma^2$, where $\hat{Y}_\omega$ is the vector of fitted values obtained when the ith case has weight ω and the remaining cases have weight 1 (notice that as $\omega \to 0$, $D_i\omega \to D_i(\omega)$). However it might be the case that some nearly similar influential points (based on D_i) have different behavior for the other weights. To view this in our data, we have plotted the influence graph of two influential cases one with $D_{13221} = 1.276$ and the other with $D_{10087} = 1.033$ which have nearly the same influence according

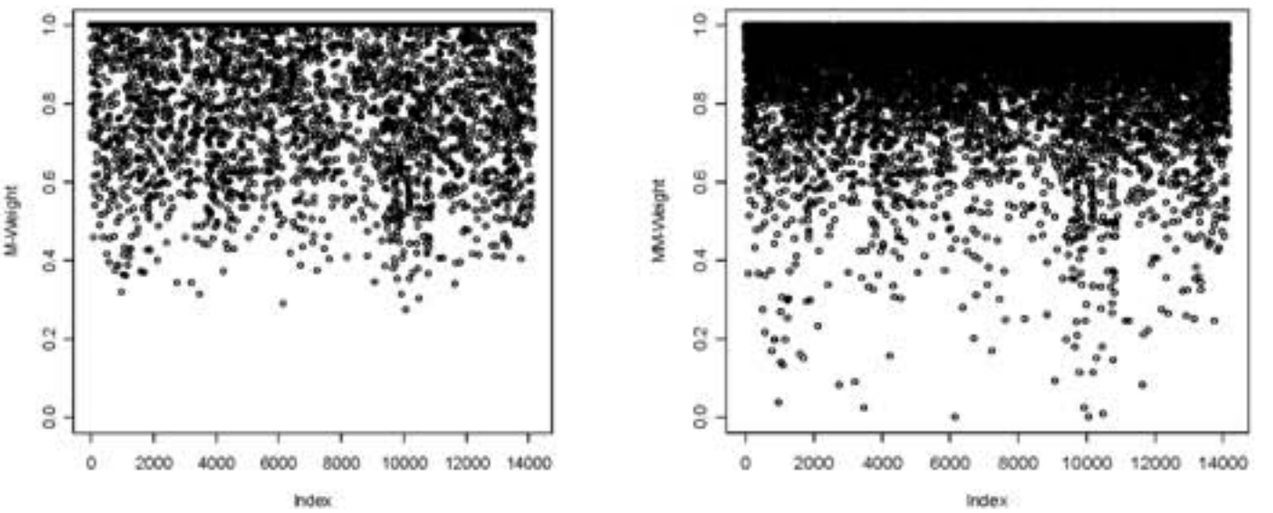


Figure 2: M and MM regression weights to each observation.

to D_i measure. Figure (3) displays the influence graph ($p \times D_i(\omega)$ vs. ω) of these two points which indicates a higher curvature at $\omega = 1$ for 13221-*th* observation which is indicated by the solid line ($C_l = 0.139$) than that of 10087-*th* observation ($C_l = 0.044$) represented by a dashed line. Also the graph shows that similarity at $\omega = 0$ which corresponds to D_i does not show the entire nature of influence. It appears that the observation number 13221 is less influential for $\omega > 1$ while it is more influential for $\omega < 1$.

Also, we apply the local influence concept to compare the two robust M and MM regression methods which are based on weighing the observed cases according to their structure. To see their differences, we can calculate the normal curvature of a model perturbed by the weights ($W_{n\times 1}$) attached by M and MM regression through the corresponding direction of maximum curvature that each considers. Calculations show that for M regression $C_l = 0.019$ whereas $C_l = 0.013$ for MM regression, which indicates less curvature for the weights corresponding to MM regression, hence it could be regarded as a more robust procedure (it is obvious that both curvatures are small due to robustness).

6 Conclusion

In this paper a linear regression model for Urban household expenditure was presented. Appropriate Box-Cox transformation for the response of interest was found to be the logarithmic transformation. Four different

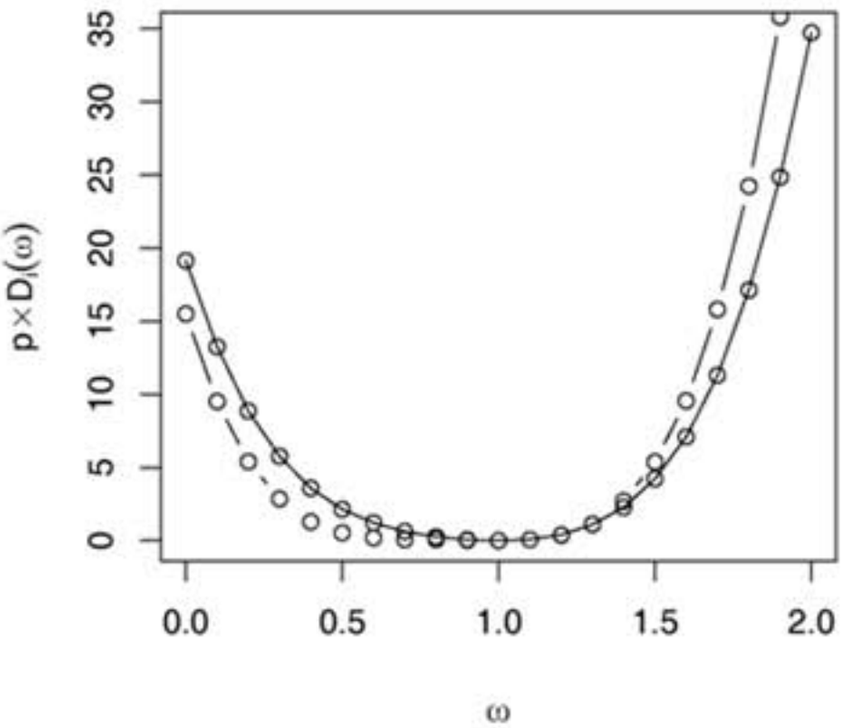


Figure 3: Influence graph of $p \times D_i(\omega)$ vs. ω for observations number 13221 (solid line) and 10087 (dashed line).

robust regressions (M, LTS, S and MM) were applied on these data and their results are compared with those of ordinary LS on these data. Normal curvature was used to compare the parameters of some robust regression methods.

Acknowledgement

We are grateful to the Statistical Research Center of Iran for supporting the present research.

References

[1] Atkinson, A. C. (1987). *Plots, Transformations and regression.* Oxford University Press.

[2] Barnard, G. A. (1980). Discussion of professor Boxs paper. *Journal of the Royal Statistical Society. A*, **143**, 404-406.

[3] Blesley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression Diagnostics.* New York: Wiley.

[4] Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika*, **40**, 318-335.

[5] Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, **B 26**, 211-243.

[6] Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. B*, **48**, 133-169.

[7] Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, **19**, 15-18.

[8] Cook, R. D. (1979). Influential observations in linear regression. *J. Amer. Statist. Ass.*, **74**, 169-174.

[9] Cook, R. D. and Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*, **70**, 1-10.

[10] Donoho, D. L. and Huber, P. J. (1983). The notion of breakdown point, *Festschrift for Erich Lehmann.*

[11] Edgeworth, F. Y. (19887). On observations relating to several quantities. *Hermathena*, **6**, 279-285.

[12] Hampel, F. R. (1971). A general qualitative definition of robustness. *J. Ann. Math. Stat.*, **42**, 1887-1896.

[13] Hampel, F. R. (1974). The influence curve andits role in robust estimation. *J. Ann. Stat. Assoc.*, **69**, 383-393.

[14] Hodges, J. L., Jr. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. *Proc. Fifth Berkeley Symp. Math. Stat. Probab.*, **1**, 163-168.

[15] Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Stat.*, **35**, 73-101.

[16] Huber, P. J. (1972). Robust statistics: A review. *Ann. Math. Stat.*, **43**, 1041-1067.

[17] Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Stat.*, **1**, 799-821.

[18] Huber, P. J. (1981). *Robust Statistics.* Jhon Wiley & Sons: New York.

[19] Le Cam, L. (1986). The central limit theorem around 1935. *Stat. Sci.*, **1**, 78-96.

[20] Pearson, E. S. (1931). The analysis of variance in cases of non-normal variation. *BiometriKa*, **23**, 114-133.

[21] Plackett, R. L. (1972). Studies in the history of probability and statistics XXIX: The discovery of the method of least squares. *Biometrika*, **59**, 239-251.

[22] Rousseeuw, P. J. (1984). Least median of squares regression. *J. Am. Stat. Assoc.*, **79**, 871-880.

[23] Rousseeuw, R. J. and A. M. Leroy. (1987). Robust Regression and Outlier Detection. New York: Wiley.

[24] Rousseeuw, P. J., and Yohai, V. (1984). Robust Regression by means of S estimators. *Robust and Nonlinear Time series Analysis*, Lecture Notes in Statistics, *26*, Springer Verlag, New York, 256-272.

[25] Stigler, S. M. (1981). Gauss and the invention of least squares. *Ann. Stat.*, **9**, 465-474.

[26] Student (1927). Errors of routin analysis. *Biometrika*, **19**, 151-164.

[27] Tukey, J. W. (1960). *A survey of sampling from contaminated distributions, in contributions to probability and statistics*. Olkin, Stanford University Press, Stanford, CA.

[28] Yohai, V. J. (1985). High breakdown-point and high efficiency robust estimates for regression. *Ann. Stat.*, **2**, 642-656.

Testing Independence: A Bayesian Predictive Approach

Amin Ghalamfarsa Mostofi

Department of Statistics, Shiraz University

Abstract: The purpose of this article is to find a Bayesian approach for testing the independence of two continuous random variables. The joint distribution function is assumed to be a mixture of Farlie-Gumbel-Morgenstern distributions. We specify appropriate prior distributions for the unknown parameters. Our test procedure is based on a logarithmic score utility function. We reject the hypothesis for which the corresponding posterior predictive density has the smaller utility. An example is given to illustrate the approach.

Keywords: Dirichlet process; Farlie-Gumbel-Morgenstern distribution; Gibbs sampler; Logarithmic score utility function; Posterior predictive density.

1 Introduction

We are concerned with testing the independence of two continuous random variables versus the hypothesis of dependence. There are a number of nonparametric tests of independence. For example, Hoeffding (1948) and Blum et al. (1961) proposed some of these tests based on functionals of the discrepancy between the empirical joint distribution function and the product of its associated marginals. Several authors applied kernel type density estimates to set up the tests of independence. (See Rosenblatt (1975), Robinson (1991), Rosenblatt and Whalen (1992) and Ahmad and Li (1997).) Among the authors, Zheng (1997) first defined a functional of the discrepancy between the conditional and marginal densities and then proposed a test procedure based on its kernel type estimate.

In this article, we take a Bayesian predictive approach. In Section 2, we consider a Bayesian semiparametric model in which the joint distribution function is written as a scale mixture of Farlie-Gumbel-

Morgenstern (FGM) distributions. (The FGM family of bivariate distribution functions is defined as follows:

$$K(x, y|\alpha) = K_1(x)K_2(y)\{1 + \alpha(1 - K_1(x))(1 - K_2(y))\}, \qquad (1)$$

where $K_1(.)$ and $K_2(.)$ are univariate distribution functions and α is an association parameter. It can be shown that $K_1(.)$ and $K_2(.)$ are the marginal distribution functions of $K(.,.|\alpha)$ for any α. In fact, the FGM formula represents a method for constructing a bivariate distribution function with specified marginals. See Morgenstern (1956), Gumbel (1960), Farlie (1960), and Johnson and Kotz (1975, 1977).) The mixing distribution is assumed to have a Dirichlet process prior. An appropriate prior distribution is also assigned to the association parameter. We describe how to use a Gibbs sampler method to generate a sample from the posterior distribution of this parameter. In section 3, we present our test procedure. The hypothesis of independence and the alternative correspond to specific Bayesian models. So we are concerned with selecting a suitable model. Predictive approaches to model selection have been discussed by several authors. (See Geisser and Eddy (1979), Gutierrez-Pena and Walker (2001), San Martini and Spezzaferri (1984), Gelfand et al. (1992), Gelfand (1995), Gelfand and Ghosh (1998), and Laud and Ibrahim (1995).) Using a logarithmic score utility function, we measure the utility of the posterior predictive density from each of these models. We naturally prefer the model for which the corresponding posterior predictive density has the larger utility. We also extend the method of Gelfand and Kottas (2002) to approximate the predictive densities. Section 4 gives a numerical example of this approach.

2 A Bayesian semiparametric model

Suppose that the random variables X and Y have the joint distribution

function

$$F(x,y|\alpha,G_1,G_2)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}K_1(x|\theta)K_2(y|\eta)\{1+\alpha(1-K_1(x|\theta))\times(1-K_2(y|\eta))\}G_1(d\theta)G_2(d\eta), \quad (2)$$

where $G_1(\,.)$ and $G_2(\,.)$ are unknown distribution functions, $0<\alpha<1$ is unknown association parameter and $K_1(.|\theta)$ and $K_2(.|\eta)$ are two parametric families of distributions, respectively, indexed by θ and η. It can be shown that the joint probability density function of X and Y is as follows:

$$f(x,y|\alpha,G_1,G_2)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}k_1(x|\theta)k_2(y|\eta)\{1+\alpha(1-2K_1(x|\theta))\times(1-2K_2(y|\eta))\}G_1(d\theta)G_2(d\eta),$$

where $k_1(.|\theta)$ and $k_2(.|\eta)$ are the probability density functions, respectively, associated with $K_1(.|\theta)$ and $K_2(.|\eta)$. Then, the marginal probability density functions of X and Y, respectively, are given by

$$f(x|G_1)=\int_{-\infty}^{+\infty}k_1(x|\theta)\,G_1(d\theta) \quad (3)$$

, and

$$f(y|G_2)=\int_{-\infty}^{+\infty}k_2(y|\eta)\,G_2(d\eta). \quad (4)$$

It can be shown that the random variables X and Y are independent if and only if $\alpha=0$. Let $G(\theta,\eta)=G_1(\theta)G_2(\eta)$. We assume a Dirichlet process prior for G, that is, $G\in DP(cG_0)$, where $G_0(\theta,\eta)=G_{01}(\theta)G_{02}(\eta)$, G_{01} and G_{02} are specified distribution functions and $c>0$ is a precision parameter. It follows that $G_1\in DP(cG_{01})$ and $G_2\in DP(cG_{02})$. Hence, (3) and (4) corresponds to two mixture of Dirichlet process models. We also consider the following prior distribution for α:

$$\begin{cases} \alpha|\alpha\neq 0 \;\sim\; \pi(.) \\ P(\alpha=0)=p_0 \end{cases}, \quad (5)$$

where $\pi(.)$ is a specified density function and $0 < p_0 < 1$. (Our prior belief about the hypothesis $H_0 : \alpha = 0$ is reflected byp_0.)

Suppose that $D = \{(X_1, Y_1), ..., (X_n, Y_n)\}$ is a random sample from (2). Following Mukhopadhyay and Gelfand (1997), we obtain a sample from the posterior distribution of α. For this purpose, we restate the model as follows. Consider latent (θ_i, η_i) associated with (X_i, Y_i), for $i = 1, ..., n$and suppose that

$$\begin{aligned}(X_i, Y_i)|\theta, \eta, \alpha, G_1, G_2 &\sim k_1(x_i|\theta_i)k_2(y_i|\eta_i) \\ \times\{1 + \alpha(1 - 2K_1(x_i|\theta_i))(1 - 2K_2(y_i|\eta_i))\};& \end{aligned} \tag{6}$$

for $i = 1, ..., n$, where $\theta = (\theta_1, ..., \theta_n)$ and $\eta = (\eta_1, ..., \eta_n)$. Given α, G_1 and G_2, the random vectors θ and η are assumed to be independent. We further suppose that

$$\theta_1, ..., \theta_n|\alpha, G_1, G_2 \overset{i.i.d.}{\sim} G_1(.) \tag{7}$$

and

$$\eta_1, ..., \eta_n|\alpha, G_1, G_2 \overset{i.i.d.}{\sim} G_2(.). \tag{8}$$

Then, $(X_1, Y_1), ..., (X_n, Y_n)$ are *i.i.d.* random vectors from (2). In what follows, we use the bracket notation of Gelfand and Smith (1990) to write the distribution of random variables. For example the prior and posterior distribution of α are, respectively, denoted by $[\alpha]$ and $[\alpha|D]$.

It can be shown that

$$[\alpha|\theta, \eta, D] \propto \prod_{i=1}^{n}\{1 + \alpha(1 - 2K_1(x_i|\theta))(1 - 2K_2(y_i|\eta))\}[\alpha], \tag{9}$$

$$\begin{aligned}[\theta_i|\alpha\ , \{\theta_j;\ j \neq i\}, \eta, D] \ &\propto\ q_{i0}g_{01}(.)\, k_1(x_i|\ .\ ) \\ \{1 + \alpha(1 - 2K_1(x_i|.))(1 - 2K_2(y_i|\eta_i))\} + &\sum_{\substack{j=1 \\ j \neq i}}^{n} q_{ij}\delta_{\theta_j}(.)\ ; \end{aligned} \tag{10}$$

and

$$[\eta_i|\alpha\ ,\theta,\{\eta_j;\ j\neq i\},D]\ \propto\ q_{i0}^{'}g_{02}(.)\,k_2(y_i|\,.\,)$$
$$\{1+\alpha(1-2K_1(x_i|\theta_i))(1-2K_2(y_i|.))\}+\sum_{\substack{j=1\\ j\neq i}}^{n}\ q_{ij}^{'}\delta_{\eta_j}(.)\ ;\ (11)$$

where, $g_{01}(.)$ and $g_{02}(.)$ are, respectively, the densities associated with $G_{01}(.)$ and $G_{02}(.)$,

$$q_{i0}\propto c(\int g_{01}(\theta)k_1(x_i|\theta)\{1+\alpha(1-2K_1(x_i|\theta))(1-2K_2(y_i|\eta_i))\}d\theta)^{-1}, \quad (12)$$

$$q_{i0}^{'}\propto c(\int g_{02}(\eta)k_2(y_i|\eta)\{1+\alpha(1-2K_1(x_i|\theta_i))(1-2K_2(y_i|\eta))\}d\eta)^{-1}, \quad (13)$$

$$q_{ij}\propto k_1(x_i|\theta_j)\{1+\alpha(1-2K_1(x_i|\theta_j))(1-2K_2(y_i|\eta_i))\}, \quad (14)$$

and

$$q_{ij}^{'}\propto k_2(y_i|\eta_j)\{1+\alpha(1-2K_1(x_i|\theta_i))(1-2K_2(y_i|\eta_j))\}, \quad (15)$$

subject to $\sum_{j\neq i}q_{ij}=1$ and $\sum_{j\neq i}q_{ij}^{'}=1$. Using the conditional distributions (9), (10) and (11), a Gibbs sampler can be implemented to provide a sample, $\alpha_1,...,\alpha_m$; from the posterior $[\alpha|D]$. The Bayes estimate of α, with respect to the squared error loss function, is given by

$$\hat{\alpha}=\frac{1}{m}\sum_{i=1}^{m}\alpha_i. \quad (16)$$

3 Test of independence

Here, we propose a method for testing $H_0 : \alpha = 0$ versus $H_1 : \alpha \neq 0$. The hypothesis H_0corresponds to the following Bayesian nonparametric model:

$$M_0 = \{f(x, y|0, G_1, G_2),\ G_1 \in DP(cG_{01}),\ G_2 \in DP(cG_{02})\}. \quad (17)$$

Note that the parameter α has the prior density $\pi(.)$ when H_1 is true. So the Bayesian semiparametric model

$$\begin{aligned} M_1 \ &= \ \{f(x, y|\alpha, G_1, G_2),\ \alpha \neq 0,\ \alpha \sim \pi(.), \\ &\quad G_1 \in DP(cG_{01}),\ G_2 \in DP(cG_{02})\}, \end{aligned} \quad (18)$$

corresponds to the hypothesis H_1. According to (5), the prior probabilities of the models M_0 and M_1 are p_0 and $1 - p_0$, respectively.

Now, the problem of testing hypotheses can be viewed as a model selection problem, in which there are only two possible Bayesian models. Gutierrez-Pena and Walker (2001) suggest a Bayesian predictive approach, based on a logarithmic score utility function, for selecting a suitable model from a finite set of specified parametric models. Let $f(x, y|D)$ be the Bayes estimate of (2) where $G_1 \in DP(cG_{01})$, $G_2 \in DP(cG_{02})$, and α is distributed according to (5) and let $f_i(x, y|D)$ be the posterior predictive density under the model M_i, for $i = 0, 1$. Our approach is also based on the following posterior expected logarithmic score utility

$$U(M_i) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \log\{f_i(x, y|D)\} f(x, y|D) dx dy, \qquad i = 0, 1. \quad (19)$$

We reject the hypothesis of independence if and only if

$$U(M_0) < U(M_1). \quad (20)$$

Here, we explain how to approximate the posterior predictive densities $f(x, y|D)$, $f_0(x, y|D)$, and $f_1(x, y|D)$. We first suppose that α has the prior distribution (5) and the random distributions G_1 and G_2 have Dirichlet process priors. Based on the conditional distributions (9), (10) and (11), we use a Gibbs sampler to draw a sample point

$(\alpha^*, \theta_1^*, ..., \theta_n^*, \eta_1^*, ..., \eta_n^*)$ from the posterior $[\alpha, \theta_1, ..., \theta_n, \eta_1, ..., \eta_n|D]$.

Define

$$G_{01}^*(.) = (c+n)^{-1}(cG_{01}(.) + \sum_{i=1}^{n} \delta_{\theta_i^*}(.)) \tag{21}$$

and

$$G_{02}^*(.) = (c+n)^{-1}(cG_{02}(.) + \sum_{i=1}^{n} \delta_{\eta_i^*}(.)), \tag{22}$$

where δ_a is a degenerate distribution at a. Suppose that $\theta_1^{'}, ..., \theta_l^{'}$ is a sample from

$$G_1^*(.) = \sum_{j=1}^{\infty} w_j \delta_{\theta_j}(.), \tag{23}$$

and $\eta_1^{'}, ..., \eta_l^{'}$ is a sample from

$$G_2^*(.) = \sum_{j=1}^{\infty} w_j^{'} \delta_{\eta_j}(.), \tag{24}$$

where

$$\begin{aligned} w_1 = z_1, \quad w_1^{'} = z_1^{'}, \quad w_j = z_j(1-z_{j-1})...(1-z_1), \\ w_j^{'} = z_j^{'}(1-z_{j-1}^{'})...(1-z_1^{'}), \quad j = 2, 3, ..., \end{aligned} \tag{25}$$

$\{Z_j, j = 1, 2, ...\}$ and $\{Z_j^{'}, j = 1, 2, ...\}$ are *i.i.d.* from $Beta(1, c+n)$, and $\{\theta_j, j = 1, 2, ...\}$ and $\{\eta_j, j = 1, 2, ...\}$ are *i.i.d.* from (21) and (22), respectively. For sufficiently large l,

$$\frac{1}{l^2} \sum_{i=1}^{l} \sum_{j=1}^{l} k_1(x_0|\theta_i^{'}) k_2(y_0|\eta_j^{'}) \{1 + \alpha^*(1 - 2K_1(x_0|\theta_i^{'}))(1 - 2K_2(y_0|\eta_j^{'}))\} \tag{26}$$

is a realization from $[f(x_0, y_0|\alpha, G_1, G_2)|D]$. (See Gelfand and Kottas (2002) for further details.) So we can generate a sample from this

posterior and approximate $f(x_0, y_0|D)$ by the sample mean. This approach is also applicable to approximate $f_0(x_0, y_0|D)$ and $f_1(x_0, y_0|D)$ if we take appropriate prior distributions for α. For example, we use the method to obtain $f_0(x_0, y_0|D)$ when α is assumed to have a degenerate distribution at zero.

4 Example

Here we work with the bird data given in Johnson and Wichern (1988, page 208). The variables are X =tail length (in millimeters) and Y =wing length (in millimeters). They were measured for a sample of $n = 45$ female hook-billed kites.

We consider the Bayesian model, given in section 2, where $K_1(x|\theta) = \Phi(x - \theta)$, $K_2(y|\eta) = \Phi(y - \eta)$, $c = 1$, $p_0 = 0.5$, $\pi(.)$ is the density function associated with the uniform distribution on the unit interval and $g_{01}(.)$ and $g_{02}(.)$ are the density functions associated with the uniform distribution on the interval $(-a, a)$. (a is taken to be sufficiently large.) We have $\hat{\alpha} = 0.707$, $U(M_0) \approx -2980.076$ and $U(M_1) \approx -384.1688$. Since $U(M_0) < U(M_1)$, the hypothesis of independence is rejected and we conclude that X and Y are dependent.

References

[1] Ahmad, I. A. and Li, Q. (1997) Testing independence by nonparametric kernel method. *Statistics and Probability letters*, 34, 201-210.

[2] Blum, J. R., Kiefer, J. and Rosenblatt, M. (1961) Distribution free tests of independence based on the sample distribution function, *Annals of Mathematical Statistics,* 32, 485-498.

[3] Farlie, D. J. G. (1960) The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, 47, 307-323.

[4] Geisser, S. and Eddy, W.F.(1979) A predictive approach to model selection. *Journal of the American Statistical Association*, 74,153-160.

[5] Gelfand, A.E. (1995) Model determination using sampling-based methods. In: Gilks, W., Richardson, S. and Spiegelhalter, D. (Eds.), Markov chain Monte Carlo in practice. Chapman & Hall, London, 145-161.

[6] Gelfand, A.E., Dey, D.K., Chang, H. (1992) Model determination using predictive distributions with implementation via sampling-based methods. In: Bernardo, J.M., Berger, J.O., David, A.P. and Smith, A.F.M. (Eds.), Bayesian Statistics, Vol. 4, Oxford University Press, Oxford, 147- 167.

[7] Gelfand, A.E. and Ghosh, S. (1998) Model choice: A minimum posterior predictive loss approach. *Biometrika*, 85, 1-11.

[8] Gelfand, A.E. and Kottas, A. (2002) A Computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 11, 284-305.

[9] Gelfand, A.E. and Smith, A.F.M. (1990) Sampling-Based approaches to calculating marginal densities. *Journal of the American Statistical Association,* 85, 398-409.

[10] Gumbel, E. J. (1960) Bivariate exponential distributions. *Journal of the American Statistical Association*, 55,698-707.

[11] Gutierrez-Pena, E. and Walker, S. G. (2001) A Bayesian predictive approach to model selection. *Journal of Statistical Planning and inference*, 93, 259-276.

[12] Hoeffding, W. (1948) A non-parametric test of independence. *Annals of Mathematical Statistics*, 19, 546-557.

[13] Johnson, N. L. and Kotz, S. (1975) On some generalized Farlie-Gumbel-Morgenstern distributions. *Communications in Statistics, Theory and methods*, 4,415-427.

[14] Johnson, N. L. and Kotz, S. (1977) On some generalized Farlie-Gumbel-Morgenstern distributions- II: Regression, Correlation and further generalizations. *Communications in Statistics, Theory and methods*, 6,485-496.

[15] Johnson, R. A. and Wichern, D. W. (1988) Applied Multivariate Statistical Analysis. Second edition, Prentice Hall International Inc.

[16] Laud, P.W. and Ibrahim, J.G. (1995) Predictive model selection. *Journal of the Royal Statistical Society*, B, 57, 247-262.

[17] Morgenstern, D. (1956) Einfache Beispiele zweidimensionaler verteilung. *Mitteislingsblatt fur Mathematische Statistik*, 8,234-235.

[18] Mukhopadhyay, S. and Gelfand, A.E. (1997) Dirichlet process mixed generalized linear models. *Journal of the American Statistical Association,* Vol. 92, No. 438, 633-639.

[19] Robinson, P. M. (1991) Consistent nonparametric entropy-based testing. *Review of economic*

studies, 58, 437-453.

[20] Rosenblatt, M.(1975) A quadratic measure of deviation of two dimensional density estimates and a test of independence, *Annals of Statistics*, 3, 1-14.

[21] Rosenblatt, M. and Whalen, B. E. (1992) A nonparametric measure of independence under the hypothesis of independent components. *Statistics and Probability letters*, 15, 245-252.

[22] San Martini, A. and Spezzaferri, F. (1984) A predictive model selection criterion. *Journal of the Royal Statistical Society*, B, 49,296-303.

[23] Zheng, J. X. (1997) A consistent specification test of independence. *Journal of Nonparametric Statistics*,7, 297-306.

On Some Statistical Aspects of Agreement Among Measurements

S. Hedayat[‖**]

Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago IL, 60607, U.S.A.

Abstract: Study of agreement between two comparable sets of measurements is needed in many areas of scientific investigations. There is an impressive literature on the topic - covering both the qualitative and quantitative features of study variables. To ascertain the extent of agreement between two quantitative measurements X and Y, coverage probability (CP) is defined as the probability of (X, Y) falling into a strip along the direction of "$X = Y$" of an arbitrarily specified tolerable width $2d$ but symmetric with respect to X and Y. For a given d, the higher CP value is, the better agreement it indicates. In this paper we address two problems. The first one is assessing and testing the agreement in terms of coverage probability for more than two raters or methods. The second one is to provide simultaneous lower confidence limit for coverage probability among three or more raters.

Keywords: Target, Measurement, Deviation, Tolerable Limits, Coverage Probability, Bivariate Normal Distribution, Multivariate Techniques.

1 Introduction

Assessment of agreement between two sets of measurements is called for in many areas of scientific investigations. Examples include assay or instrument validation, process or method comparisons, statistical process control and IBE (Individual Bio-Equivalence), etc. The classic example occurs when two or more raters evaluate the experimental units in a study and we need to find out how well these raters agree. Here raters could be experts or specialists or could even relate to two or more distinct methods of measuring the same response variable.

‖Invited speaker
**Email: hedayat@uic.edu

We contemplate on a situation wherein we have available a reference gold standard against which several competitors are to be compared with respect to their performance as judged by the CP. We also address the problem of judging the performance of the competitors in the absence of any prescribed gold standard. We deal with situations involving continuous measurements only.

Cohen's Kappa (1960) and weighted Kappa (1968) are the most popular indices for measuring agreement when the responses are nominal. Weighted Kappa statistic has been proposed by Landis and Koch (1977) which is appropriate for assessing agreement when the categories of response are ordinal. We will not deal with such data in this article.

Lin (1989) introduced Concordance Correlation Coefficient (CCC) for measuring agreement which is more appropriate when the data are measured on a continuous scale. A weighted CCC was proposed by Chinchilli, Martel, Kumanyika and Lloyd (1996) for repeated measurement designs and a generalized CCC for continuous and categorical data was introduced by King and Chinchilli (2001). Lin (2000) also introduced Total Deviation Index (TDI) for measuring individual agreement with applications in lab performance and bio-equivalence. Lin, Hedayat, Sinha and Yang (2002) (henceforth abbreviated as LHSY (2002)) proposed methods for checking the agreement in terms of coverage probability (CP) when the two measurements are quantitative in nature. LHSY (2002), Lou (2006), and Lin, Hedayat and Wu (2007) serve as the key references to our work.

2 Concept and Use of Coverage Probability

Broadly speaking, we would say the observations and the target values possess good agreement if a large proportion of observations are within a predetermined and acceptable boundary from the respective target values. To achieve the agreement between target value (X) and

observation (Y), the concept of coverage probability (CP) was introduced by LHSY(2002) and it is defined as the probability of (X, Y) falling into a strip along the direction of "$X = Y$" of an arbitrarily specified width $2d$ but symmetric with respect to X and Y. For a given d, the higher CP value is, the better agreement it indicates. A widely acceptable criterion is that the observation has good agreement with the target value if CP is at least 90% for give d value.

Following LHSY(2002), we start with the definition of CP given by

$$CP(d) = Pr(|Y-X| < d) = Pr(\frac{-d-\mu_D}{\sigma_D} < \frac{Y-X-\mu_D}{\sigma_D} < \frac{d-\mu_D}{\sigma_D}) \tag{1}$$

In the above, $D = Y - X$, $\mu_D = \mu_y - \mu_x$ is the mean of D and $\sigma_D^2 = \sigma_y^2 + \sigma_x^2 - 2\rho\sigma_y\sigma_x$ is the variance of D.

At this stage, we will assume bivariate normal (BVN) distribution of (Y, X). This is usually achieved in practice by suitable transformations of the data underlying the variables X and Y. Log-transformation is often a guiding factor giving good results. LHSY(2002) discussed the problem of estimation of and inference on CP based on n paired sample observations $(y_i, x_i), 1 \leq i \leq n$, drawn from a bivariate normal distribution. In terms of $\Phi(.)$,

$$CP(d) = \Phi(\frac{d-\mu_D}{\sigma_D}) - \Phi(\frac{-d-\mu_D}{\sigma_D}). \tag{2}$$

The estimates of μ_D and σ_D^2 are $\widehat{\mu_D} = M_D = M_Y - M_X$, $S_D^2 = \frac{n}{n-1}(S_x^2 + S_y^2 - 2S_{xy})$, where $M_Y = \overline{y} = \frac{\sum y_i}{n}$, $M_X = \overline{x} = \frac{\sum x_i}{n}$, $S_x^2 = \frac{\sum(x_i-\overline{x})^2}{n}$, $S_y^2 = \frac{\sum(y_i-\overline{y})^2}{n}$, and $S_{xy} = \frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{n}$.

It is well known that under BVN distribution of X and Y, M_D and S_D^2 are independent. LHSY (2002) estimated $CP(d)$ as

$$\widehat{CP(d)} = \Phi(\frac{d - M_D}{S_D}) - \Phi(\frac{-d - M_D}{S_D}). \quad (3)$$

and further computed large sample approximation to $\mathrm{Var}(\widehat{CP(d)})$ as

$$Var(\widehat{CP(d)}) \approx \frac{1}{n}([\phi(\frac{d - \mu_D}{\sigma_D}) - \phi(\frac{-d - \mu_D}{\sigma_D})]^2 + \frac{1}{2}[\frac{d - \mu_D}{\sigma_D}\phi(\frac{d - \mu_D}{\sigma_D}) + \frac{d + \mu_D}{\sigma_D}\phi(\frac{-d - \mu_D}{\sigma_D})]^2). \quad (4)$$

Consequently, the plug-in estimator of $\mathrm{Var}(\widehat{CP_d})$ is

$$\widehat{Var(\widehat{CP_d})} \approx \frac{1}{n}([\phi(\frac{d - M_D}{S_D}) - \phi(\frac{-d - M_D}{S_D})]^2 + \frac{1}{2}[\frac{d - M_D}{S_D}\phi(\frac{d - M_D}{S_D}) + \frac{d + M_D}{S_D}\phi(\frac{-d - M_D}{S_D})]^2).$$

As observed in LHSY (2002), because CP is bounded by 0 and 1, it is better to use the logit transformation for inference on the CP. Let $\zeta = ln(\frac{\hat{CP}}{1-\hat{CP}})$. Its asymptotic mean is $ln(\frac{CP}{1-CP})$, and its asymptotic variance is $\frac{Var(\hat{CP})}{\hat{CP}^2(1-\hat{CP})^2}$.

3 Coverage Probability Involving Gold Standard vs Its Competitors

LHSY(2002) have discussed the problem of evaluating agreement via coverage probability where there are two rater groups. However, in practice, we may have more than two rater groups. For example, we may have multiple sites for testing new medical devices, or multiple readings at one site. Here we want to address two problems. The first one is assessing and testing the agreement in terms of coverage probability for more than two rater groups or methods. The second one is to provide the simultaneous lower confidence limit for coverage

probability among three or more rater groups. The key reference to this work is an unpublished doctoral dissertation (Lou (2006)).

There are two typical cases that can arise in practice. Case 1 is gold standard versus treatment rater groups. In this case, we will focus on the coverage probability involving each treatment method and the gold standard. Case 2 is more general where the K (≥ 2) rater groups are equally important. So our interest will be in all the pairwise coverage probability. We will discuss case 1 here and deal with case 2 in next section.

We will introduce tests for equality of coverage probability in Section 3.1, discuss the problem of specification of simultaneous lower limit in Section 3.2 and then present tests for minimum assurance level of CP in Section 3.3.

Before concluding this section, we present some general expressions and results which will be used in the sequel. We will assume joint normality of all the raters' measurements.

For two rater groups X_i and X_j,

$$CP_{ij}(d) = Pr(|X_i - X_j| < d) = Pr(\frac{-d - \mu_{D_{ij}}}{\sigma_{D_{ij}}} < \frac{X_i - X_j - \mu_{D_{ij}}}{\sigma_{D_{ij}}} < \frac{d - \mu_{D_{ij}}}{\sigma_{D_{ij}}}), \quad (5)$$

where $D_{ij} = X_i - X_j$, $\mu_{D_{ij}} = \mu_{X_i} - \mu_{X_j} = \mu_i - \mu_j$ is the mean of D_{ij} and $\sigma^2_{D_{ij}} = \sigma^2_{X_i} + \sigma^2_{X_j} - 2\rho_{ij}\sigma_{X_i}\sigma_{X_j} = \sigma^2_i + \sigma^2_j - 2\rho_{ij}\sigma_i\sigma_j$ is the variance of D_{ij}.

Refer to (2.3), (2.4) and (2.5) for $\widehat{CP_{ij}(d)}$, Var($\widehat{CP_{ij}(d)}$) and $\widehat{Var(CP_{ij}(d))}$.

For three rater groups X_i, X_j, X_k given d, the covariance of $\widehat{CP_{ij}}, \widehat{CP_{ik}}$ is

$$
\begin{aligned}
COV_{ij,ik} = \; & \frac{1}{n\sigma_{D_{ij}}\sigma_{D_{ik}}}[\sigma_i^2 - \rho_{ij}\sigma_i\sigma_j - \rho_{ik}\sigma_i\sigma_k + \rho_{jk}\sigma_j\sigma_k]* \\
& \{[\phi(\tfrac{d_{ij}-\mu_{D_{ij}}}{\sigma_{D_{ij}}}) - \phi(\tfrac{-d_{ij}-\mu_{D_{ij}}}{\sigma_{D_{ij}}})][\phi(\tfrac{d_{ik}-\mu_{D_{ik}}}{\sigma_{D_{ik}}}) - \phi(\tfrac{-d_{ik}-\mu_{D_{ik}}}{\sigma_{D_{ik}}})] \\
& + \frac{1}{2\sigma_{D_{ij}}\sigma_{D_{ik}}}(\sigma_i^2 - \rho_{ij}\sigma_i\sigma_j - \rho_{ik}\sigma_i\sigma_k + \rho_{jk}\sigma_j\sigma_k)* \\
& [\tfrac{d_{ij}-\mu_{D_{ij}}}{\sigma_{D_{ij}}}\phi(\tfrac{d_{ij}-\mu_{D_{ij}}}{\sigma_{D_{ij}}}) + \tfrac{d_{ij}+\mu_{D_{ij}}}{\sigma_{D_{ij}}}\phi(\tfrac{-d_{ij}-\mu_{D_{ij}}}{\sigma_{D_{ij}}})]* \\
& [\tfrac{d_{ik}-\mu_{D_{ik}}}{\sigma_{D_{ik}}}\phi(\tfrac{d_{ik}-\mu_{D_{ik}}}{\sigma_{D_{ik}}}) + \tfrac{d_{ik}+\mu_{D_{ik}}}{\sigma_{D_{ik}}}\phi(\tfrac{-d_{ik}-\mu_{D_{ik}}}{\sigma_{D_{ik}}})]\}
\end{aligned} \tag{6}
$$

where
$\sigma_{D_{ij}}^2 = \sigma_i^2 + \sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j$, $\sigma_{D_{ik}}^2 = \sigma_i^2 + \sigma_k^2 - 2\rho_{ik}\sigma_i\sigma_k, \sigma_{D_{jk}}^2 = \sigma_j^2+\sigma_k^2-2\rho_{jk}\sigma_j\sigma_k$, $\mu_{D_{ij}} = \mu_i-\mu_j$, $\mu_{D_{ik}} = \mu_i-\mu_k$, $\mu_{D_{jk}} = \mu_j-\mu_k$.

The proofs of all technical results are given in Lou (2006) and are readily available.

Consequently, the correlation coefficient $\rho_{ij,ik}$ is:

$$\rho_{ij,ik} = \frac{COV_{ij,ik}}{\sqrt{Var(CP_{ij})}, \sqrt{Var(CP_{ik})}}. \tag{7}$$

The estimate of the covariance between $\widehat{CP_{ij}}, \widehat{CP_{ik}}$ can be taken to be

$$
\begin{aligned}
\widehat{COV_{ij,ik}} = \; & \frac{1}{nS_{D_{ij}}S_{D_{ik}}}[S_i^2 - r_{ij}S_iS_j - r_{ik}S_iS_k + r_{jk}S_jS_k]* \\
& \{[\phi(\tfrac{d_{ij}-M_{D_{ij}}}{S_{D_{ij}}}) - \phi(\tfrac{-d_{ij}-M_{D_{ij}}}{S_{D_{ij}}})][\phi(\tfrac{d_{ik}-M_{D_{ik}}}{S_{D_{ik}}}) - \phi(\tfrac{-d_{ik}-M_{D_{ik}}}{S_{D_{ik}}})] \\
& + \frac{1}{2S_{D_{ij}}S_{D_{ik}}}(S_i^2 - r_{ij}S_iS_j - r_{ik}S_iS_k + r_{jk}S_jS_k)* \\
& [\tfrac{d_{ij}-M_{D_{ij}}}{S_{D_{ij}}}\phi(\tfrac{d_{ij}-M_{D_{ij}}}{S_{D_{ij}}}) + \tfrac{d_{ij}+M_{D_{ij}}}{S_{D_{ij}}}\phi(\tfrac{-d_{ij}-M_{D_{ij}}}{S_{D_{ij}}})]* \\
& [\tfrac{d_{ik}-M_{D_{ik}}}{S_{D_{ik}}}\phi(\tfrac{d_{ik}-M_{D_{ik}}}{S_{D_{ik}}}) + \tfrac{d_{ik}+M_{D_{ik}}}{S_{D_{ik}}}\phi(\tfrac{-d_{ik}-M_{D_{ik}}}{S_{D_{ik}}})]\}
\end{aligned}
$$

where $S_{D_{ij}}^2 = S_i^2 + S_j^2 - 2r_{ij}S_iS_j$, $S_{D_{ik}}^2 = S_i^2 + S_k^2 - 2r_{ik}S_iS_k, S_{D_{jk}}^2 = S_j^2 + S_k^2 - 2r_{jk}S_jS_k$, $M_{D_{ij}} = M_i - M_j$, $M_{D_{ik}} = M_i - M_k$, $M_{D_{jk}} =$

$M_j - M_k$.

3.1 Testing Equality of Coverage Probability Involving Gold Standard vs Treatments

In practice, we desire to test whether the coverage probability between the gold standard and each treatment group is the same for all treatment groups. Here we assume group 0 to represent the gold standard and groups 1 to K to represent the treatment groups.

Our null hypothesis is:

$$H_0: \quad \begin{cases} CP_{01} = CP_{02} \\ CP_{01} = CP_{03} \\ CP_{01} = CP_{04} \\ \quad \cdots \\ CP_{01} = CP_{0K} \end{cases} \qquad H_1: \quad not \quad H_0$$

This is a $(K-1)$- dimensional hypothesis. We implicitly assume that all the CP's are judged with reference to a fixed and specified d, say d_0. Further, all through our analysis, we use large sample normal approximation to estimate CP.

Assume under H_0, $CP_{01} = CP_{02} = \cdots = CP_{0K} = \theta$ and let $\eta' = [\widehat{CP_{01}}, \widehat{CP_{02}}, \cdots, \widehat{CP_{0K}}]$. Under null hypothesis, $\underline{\eta}$ follows asymptotic multivariate normal distribution $N(\theta\underline{1}, W)$. W is the variance-covariance matrix of $\underline{\eta}$, where $w_{gg} = var(\widehat{CP_{0g}})$ and $w_{gg'} = COV(\widehat{CP_{0g}}, \widehat{CP_{0g'}})$, 1 ≤g≤ K.

We know under H_0, $\theta^* = \frac{\underline{\eta}'\widehat{W}^{-1}\underline{1}}{\underline{1}'\widehat{W}^{-1}\underline{1}}$ is BLUE of θ. Further, θ^* follows $N(\theta, \frac{1}{\underline{1}'W^{-1}\underline{1}})$ under H_0. Thus $Q = \underline{\eta}'(\widehat{W}^{-1} - \frac{\widehat{W}^{-1}\underline{1}\underline{1}'\widehat{W}^{-1}}{\underline{1}'\widehat{W}^{-1}\underline{1}})\underline{\eta}$ has an approximate chi-square distribution with (K-1) degrees of freedom

under H_0. Here $\hat{W}$ is the plug-in estimate of W. For given d_0, with significance level of α, we reject H_0 when $Q > \chi^2_{(K-1)}(\alpha)$. For a proof, refer to Rao (1973).

For simplicity, we will use CP_g as an abbreviation for CP_{0g}, $1 \leq g \leq K$ in the next section.

3.2 Simultaneous Lower Confidence Limit for All CP's

For two or more new treatment groups and one gold standard, we want to find out the simultaneous lower confidence limit for all coverage probabilities between new treatment groups and the gold standard. In other words, for the K pairs of coverage probability, we try to find the value of L_i such that $Pr(CP_1 \geq L_1, CP_2 \geq L_2, \cdots, CP_g \geq L_g, \cdots, CP_K \geq L_K) \geq (1-\alpha)$.

Let $Z_g = \frac{\widehat{CP_g} - CP_g}{\sqrt{VAR(\widehat{CP_g})}}$ and $\underline{Z'} = [Z_1, Z_2, ..., Z_K]$. $\underline{Z}$ follows asymptotic multivariate normal distribution $N(\underline{0}, \Lambda)$. Λ is the variance-covariance matrix of $\underline{Z}$, where $\Lambda_{gg} = 1$ and $\Lambda_{gg'} = \frac{COV(\widehat{CP_g}, \widehat{CP_{g'}})}{\sqrt{VAR(\widehat{CP_g})VAR(\widehat{CP_{g'}})}}$.

We start by solving for z such that $Pr(Z_1 \leq z, Z_2 \leq z, \cdots, Z_g \leq z, \cdots, Z_K \leq z) = (1-\alpha)$. Then, using the above expressions for Z_1, Z_2, ..., Z_K, we obtain the simultaneous lower confidence limits for CP_g, g=1,2,...,K as

$$L_g = \widehat{CP_g} - z\sqrt{\widehat{VAR(\widehat{CP_g})}}, 1 \leq g \leq K. \tag{8}$$

Table 1 lists the values of z for some selected values of the correlation coefficient ρ for K=2. These are calculated by software SAS (version 8 or higher).

Table 1: Values of z with changes of ρ for K=2

	ρ								
α	0	0.1	0.5	0.7	0.8	0.9	0.95	0.99	1
0.01	2.5751	2.5739	2.5574	2.5322	2.5092	2.4689	2.4347	2.379	2.3263
0.05	1.9545	1.951	1.917	1.8773	1.8460	1.7976	1.7581	1.6987	1.6449
0.10	1.6322	1.6257	1.5770	1.5287	1.4935	1.4396	1.3980	1.3359	1.2815

For α=0.05, from Table 1, z=1.7976. i.e., $Pr(Z_1 \leq 1.7976,\ Z_2 \leq 1.7976) \geq 95\%$.

Thus the simultaneous 95% confidence lower limits for CP_1 and CP_2 are:

$$L_1 = \widehat{CP_1} - 1.7976\sqrt{\widehat{VAR(\widehat{CP_1})}}$$
$$L_2 = \widehat{CP_2} - 1.7976\sqrt{\widehat{VAR(\widehat{CP_2})}}$$

3.3 Testing Minimum Assurance Level of CP for All Groups

In practice, we may desire to have minimum assurance level of CP (such as 90% for many medical device cases) based on previous studies and we wish to test whether CP maintains the minimum assurance level for each treatment group against the gold standard.

We formulate the null hypothesis as
H_0: $CP_g \geq P_g\ \forall$ g, vs H_1: $CP_g < P_g$, for at least one, g=1,2,..., K

When there is only one new treatment, this is a simple one sided test. For more than one new treatment groups, we will use the simultaneous lower confidence limits approach to accomplish the above test.

Remark: Normally, we may choose the same value of d for each test treatment and consequently, the same value of P_g, say P_0 for all i.

With given d and α, we can use (3.4) to estimate the simultaneous

lower confidence limit for all the CP_g's as discussed in Section 3.2. With the significance level of α, we will reject H_0 if any of $L_g \leq P_0$, $1 \leq g \leq K$.

Acknowledgement I am thankful to my doctoral student Dr Congrong Lou for sharing her dissertation with me. I would also like to express my sincere thanks to my long-time collaborator Professor Bikas K Sinha of Indian Statistical Institute, Kolkata for taking keen interest on this problem.

References

[1] Chinchilli, V.M., Martel, J.K., Kumanyika, S. and Lloyd, T. (1996). A Weighted Concordance Correlation Coefficient for Repeated Measurement Desings. *Biometrics*, **52**, 341-353.

[2] Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Education and Psychological Measurement*. **20**, 37-46.

[3] Cohen, J. (1968). Weighted Kappa: Normal Scale Agreement with Provision for Scaled Disagreement or Partial Credit. *Psychological Bulletin*, **70**, 213-220.

[4] King, Tonya S., Chinchilli, V.M. (2001). A Generalized Concordance Correlation Coefficient for Continuous and Categorical Data. *Statistical in Medicine*, **20**, 2131-2147.

[5] Landis, J.R., Koch, G.G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, **33**, 159-174.

[6] Lin, L. I.-K. (1989). A Concodance Correlation Coefficient to Evaluate Reporducibility. *Biometrics,* **45**, 25-268.

[7] Lin, L. I.-K. (2000). Total Deviation Index for Measuring Individual Agreement with Applications in Laboratory Performance and Bioequivalence. *Statistics in Medicine*, **19**, 255-270.

[8] Lin, L.I.-K., Hedayat, A.S., Sinha, B., Yang, M. (2002). Statistical Methods in Assessing Agreement: Models, Issues and Tools. *Journal of the American Statistical Association*, **97**, 257-270.

[9] Lin, L.I.-K., Hedayat, A.S., and Wu, W. (2007). A Unified Approach for Assessing Agreement for Continuous and Categorical Data. *Journal of Biopharmaceutical Studies*, **17**, 629-652.

[10] Lou, Congrong (2006). Assessment of Agreement. PhD dissertaion, Dept. of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, USA.

[11] Rao, C.R. (1973). *Linear Statistical Inference and Its Applications (Second Edition),* New York: John Wiley & Sons, Inc.

D_b-optimal Design for Rank-Order Nested Logit Models in discrete choice experiment

Habib Jafari

Department of Mathematics and Statistics, Otto-Von-Geureick University Magdeburg, 39106 , Germany

Abstract: It has been explored in a classical discrete choice experiment, the respondent is asked to choose an alternative with the highest utility among alternatives in choice set. But in a rank-order discrete choice experiment, the respondent is asked to rank a number of alternatives instead of preferred one.

Also it has been showed that D_b -optimal criterion in rank-order MNL model is greater than the D_b-optimal criterion of classical MNL model, which has been based on determinant of the information matrix.

We know that MNL model, when the assumption IIA is hold, is useful to analyes; otherwise it is better to use the other logit models like $NMNL$ model to analyes data which are relaxed from IIA assumption.

In this paper, we study the information matrix of rank-order $NMNL$ model and D_b-optimal design for that in discrete choice experiment.

Keywords: Conjoint analysis ,discrete Choice Experiment, D_b-Optimality, Rank-Order nested multinomial logit model.

1 Introduction

A discrete choice experiment measures the importance of the features of a good or service in making a purchase decision. This is achieved by asking each respond to choose his/her preferred alternative from a number of choice set. A Rank-Order conjoint experiment measures the importance of the features of a good or service by asking the respond to rank a certain number of alternatives within the choice sets. Data from a Rank-Order experiment can be analyzed by the rank-ordered exploded Logit (MNL, $NMNL$, ...) models (*Begg et al. , 1981, Hausman and Ruud, 1987).*

Vermulen , Goos and Vandebroek have proposed to use the D-optimality criterion which focuses on the accuracy of the estimates of the rank-ordered *multinomial logit* (MNL) model its parameters.

And in this paper we study to use the D- optimality criterion to estimate rank-ordered *Nested Multinomial logit* ($NMNL$) model its

parameters. The central question is then whether the corresponding Bayesian D_b-optimal ranking design results in significantly more precise estimates and predictions than commonly used design strategies in marketing.

In the next section, we review the rank-ordered multinomial logit model, then in section 3 we obtain the information matrix related to rank-ordered Nested multinomial logit model and we define a special class of design and obtaining D_b-optimal design for it and at the end we come to conclusion.

2 D_b-optimal Design in the Rank-Ordered MNL Model

The rank of an alternative is determined by its utility. The utility of alternative j in choice set s experienced by respondent i is modeled as

$$U_{sij} = \underbrace{\sum_{k=1}^{K}\sum_{\ell=1}^{L_k-1} X_{sijk\ell}\,\beta_{k\ell}}_{X_{sij}^T} + \varepsilon_{sij} \begin{cases} i = 1,2,...,I(individuals) \\ j = 1,2,...,J(alternatives) \\ s = 1,2,...,S(\text{choice set}) \end{cases} \quad (1)$$

Where, X_{ij}^T is the characteristics of attributes(there are K attributes each of them with $L_k; \forall k = 1,2,...,K$ levels) related to alternative j(Main-effects model) which is chosen by individual i and $\beta = (\beta_1^T, \beta_2^T, ..., \beta_K^T)$ is $p(= \sum_{k=1}^{K}(L_k - 1))$-dimensional vector of parameters and ε_{sij} denote the stochastic component of utility function which have $i.i.d$ Extreme Value distribution type II (Gumbel).

Now, if $Y_{(1)}, Y_{(2)}, ..., Y_{(J)}$ denote the Rank-alternatives of choice set then $Y_{(1)} = j$ means that alternative j has the first rank ($R_j = 1$, R denote the rank of alternative) which it has the highest utility and $Y_{(2)} = j'$ means that alternative j' has the second rank ($R_{j'} = 2$) which its utility is less than the utility of j and greater than remain alternatives and so on.

Simplicity, we suppose that there are three alternatives in choice set that those are denoted by $1, 2, 3$ Then the information matrix of

rank-order MNL models is obtained follow (*Vermulen et al.*)

$$M_{R(MNL)}(\beta|C) = M_{MNL}(\beta|C) + \sum_{j=1}^{3} P(Y_{(1)} = j) \times M_{MNL}(\beta|C_{(j)}) \quad (2)$$

with

$$M_{MNL}(\beta|C) = X^T(\underbrace{P - pp^T}_{D})X$$

$$M_{MNL}(\beta|C_{(j)}) = X_{(j)}^T(\underbrace{P_{(j)} - p_{(j)}p_{(j)}^T}_{D_{(j)}})X_{(j)}$$

where

- $M_{MNL}(\beta|C)$*:The information matrix of a discrete choice experiment by choice set with size three(*Sandor and Wedel ,2001*)*
- $C_{(j)}$ *denotes a choice set without alternative* $j = 1, 2, 3$.
- $p_{(j)}$*: is a 2-dimensional vector containing the probabilities of ranking alternative* j' *second (j'=1,2,3 and* $j' \neq j$*), given that* j *was ranked first.*
- $P_{(j)}$*: is a diagonal matrix with the elements of* $p_{(j)}$ *on its diagonal.*
- $X_{(j)}$*: is the* $(2 \times K)$ *design matrix containing all attribute levels of the profiles in choice set , except those the first-ranked profile* j.

Similarity to discrete Choice Experiment (*Sandor and Wedel (2001,2005)* and *Kessels et al. (2006a)*), in this situation (Rank-Oreder MNL) it is used the Bayesian approach to obtain D_b-optimal criterion where the Bayesian version of the D_b-error is given by

$$\begin{aligned} D_b - error &= E_\beta[\{det(M_{R(MNL)}(\beta|C))\}^{-\frac{1}{p}}] \\ &= \int_{\Re^p}\{det(M_{R(MNL)}(\beta|C))\}^{-\frac{1}{p}}\pi(\beta)d\beta \end{aligned} \quad (3)$$

Since there is no analytical way to compute the integral(3), it has to be approximated by using of Monte-Carlo technique. The design which minimizes the average of D_b-error over all draws based on β with size $R(\beta^{(1)}, \beta^{(2)}, ..., \beta^{(R)})$ is called the Bayesian D_b-optimal ranking design, means that

$$D_b \simeq \frac{1}{R}\sum_{r=1}^{R}\{det(M_{R(MNL)}(\beta^{(r)}|C))\}^{-\frac{1}{p}} \tag{4}$$

Now, we suppose that there are $L_1 \times L_2 \times ... \times L_K$ possible alternatives(L_1, L_2 $, ..., L_K$ are the levels of K attributes) , then there will be $S = \binom{L_1 \times L_2 \times ... \times L_K}{3}$ choice sets each of them include three alternatives.

Note: In most non-Bayesian linear problem an upper bound on the number of support points in an optimal design is available, see *Pukelsheim (1993,pp. 188-189).* The D-optimality criterion in linear models typically leads to an optimal number of support points that is the same as the number of unknown parameters and the design takes an equal number of observations at each point(*Silvey ,1980,pp.42*).

The bound also applies to most local optimality criteria and Bayesian criteria for linear models(see, *Chernoff,1972,pp.27*). In contrast for nonlinear models there is no such bound available on the number of support points.

Chaloner and Larntz(1986, 1989) have given the first examples of how the number of support points in an optimal Bayesian design increases as the prior distribution becomes more dispersed. They found that for prior distributions that have supported over a very small region the Bayesian optimal designs are almost the same as the locally optimal design and they have the same number of support points as the number unknown parameters. For more dispersed prior distributions there are more support points. This is a useful feature for a design as, if there are more support points than unknown parameters, the model assumptions can be checked with data from the experiment. In locally D-optimal designs for various non-linear models follow *Caratheodory*

theorem that for p-parameter nonlinear model, the number of support points is between p and $\frac{p(p+1)}{2}$. When we search for a D-optimal design, we only need to search for the optimal design in the class of design measures with number of support points between p and $\frac{p(p+1)}{2}$ for which the information matrices are nonsingular.

Therefore, according to the number of parameters,p, we are considering $N = \binom{S}{p'}$, discrete choice experiment (design) each of them with p' support points as follows

$$\xi_n = \left\{ \begin{array}{cccc} C_{n1} & C_{n2} & \cdots & C_{np'} \\ w_{n1} & w_{n2} & \cdots & w_{np'} \end{array} \right\} \in \Xi_n; n = 1, 2, ..., N \qquad (5)$$

where $p \leq p' \leq \frac{p(p+1)}{2}$. In this situation the information matrix of design (9) is as follow

$$M_{R(MNL)}(\beta|\xi_n) = \sum_{s=1}^{p'} w_{ns} M_{R(MNL)}(\beta|C_{ns}) \qquad (6)$$

where,

$$\sum_{s=1}^{p'} w_{ns} = 1; 0 \leq w_{ns} \leq 1 : \forall s = 1, 2, ..., p' \& n = 1, 2, ..., N$$

thus ,if

$$D_b(\xi_n^*) = \min_{\xi_n \in \Xi_n} D_b(\xi_n); n = 1, 2, ..., N$$

where

$$D_b(\xi_n^*) = \frac{1}{R} \sum_{r=1}^{R} det \left\{ \sum_{s=1}^{p'} w_{ns}^* M_{R(MNL)}(\beta^{(r)}|C_{ns}) \right\}^{-\frac{1}{p'}} \qquad (7)$$

then

$$\xi_n^* = \left\{ \begin{array}{cccc} C_{n1} & C_{n2} & \cdots & C_{np'} \\ w_{n1}^* & w_{n2}^* & \cdots & w_{np'}^* \end{array} \right\} \qquad (8)$$

is D_b-optimal in $\Xi_n; n = 1, 2, ..., N$.

In this situation and based on (8) there are N optimal design for spacial classes (5), thus it can be told the most suitable design is a design corresponding to

$$\min_n D_b(\xi_n^*)$$

3 D-Optimal Design for Rank-Order Nested Multinomial Logit models

Simplicity, we suppose that there are a nested logit models with two nests and three alternatives that one of them have two alternatives and another nest has just one alternative. To obtain the log-likelihood function, we are considering the following notations to denote alternatives and nests which include them in a choice set

$$(j, m), (j', m'), (j'', m'')$$

Then, we consider the combinations of them as $(j, m)(j', m')(j'', m'')$; $(j, j',\ j'' = 1, 2, 3(j \neq j' \neq j'')$, $m, m'm'' = I, II)$ means that alternative j in nest m has the first rank $(R_{(j,m)} = 1)$, alternative j' in nest m' has the second rank $(R_{(j',m')} = 2)$ and alternative j'' in nest m'' has the third rank $(R_{(j'',m'')} = 3)$ with the following assumptions

$$P(j, m) = \begin{cases} P(j|m) \times P(m) & \text{if } j \in m \\ 0 & o.w \end{cases}$$

and so on.

In this situation we define the following variables to denote observations

$$Y_{((j,m)(j',m')(j'',m''))} = \begin{cases} 1 & if\ Y_{(1)} = j, Y_{(2)} = j'\ and\ Y_{(3)} = j'' \\ 0 & o.w \end{cases}$$

Now, according to above definitions and log-likelihood function , the information matrix is calculated as follow

$$M_{R(NMNL)}(\theta|C) = M_{NMNL}(\theta|C) + \sum_{m=I}^{II} \sum_{j=1}^{3} P(Y_{(1)} = j, M = m) \times G(j, j', j'') \tag{9}$$

with

$$G(j, j', j'') = \sum_{m', m''=I}^{II} \underbrace{\sum_{j', j''=1}^{3}}_{j' \neq j''} P((Y_{(1)} = j', M' = m')|(Y_{(2)} = j, M = m))$$

$$\times \frac{-\partial^2 Ln(P((Y_{(1)} = j', M' = m')|(Y_{(2)} = j, M = m)))}{\partial\theta\partial\theta^T}$$

where

$$P(Y_{(1)} = j|M = m) = \begin{cases} \frac{exp(\frac{x_j^T\beta}{\lambda_m})}{exp(\frac{x_j^T\beta}{\lambda_m}) + exp(\frac{x_{j'}^T\beta}{\lambda_m})} & \text{if } m = \{j, j'\} \\ 1 & \text{if } m = \{j\} \end{cases} \tag{10}$$

$$P(M = m) = \begin{cases} \frac{[exp(\frac{x_j^T\beta}{\lambda_m}) + exp(\frac{x_{j'}^T\beta}{\lambda_m})]^{\lambda_m}}{[exp(\frac{x_j^T\beta}{\lambda_m}) + exp(\frac{x_{j'}^T\beta}{\lambda_m})]^{\lambda_m} + exp(x_{j''}^T\beta)} & \text{if } m = \{j, j'\} \\ \frac{exp(x_j^T\beta)}{[exp(\frac{x_{j'}^T\beta}{\lambda_{m'}}) + exp(\frac{x_{j''}^T\beta}{\lambda_{m'}})]^{\lambda_{m'}} + exp(x_j^T\beta)} & \text{if } m = \{j\} \end{cases} \tag{11}$$

$$P((Y_{(2)} = j')|(Y_{(1)} = j), M' = m') = \begin{cases} 1 & \text{if } m' = \{j, j'\} \\ \frac{exp(\frac{x_{j'}^T\beta}{\lambda_{m'}})}{exp(\frac{x_{j'}^T\beta}{\lambda_{m'}}) + exp(\frac{x_{j''}^T\beta}{\lambda_{m'}})} & \text{if } m' = \{j', j''\} \\ 1 & \text{if } m' = \{j'\} \end{cases} \tag{12}$$

$$P(M'=m'|(Y_{(1)}=j))=\begin{cases}\frac{exp(x_{j'}^T\beta)}{exp(x_{j'}^T\beta)+exp(x_{j''}^T\beta)} & \text{if } m'=\{j,j'\}\\ 1 & \text{if } m'=\{j',j''\}\\ \frac{exp(x_{j'}^T\beta)}{exp(x_{j'}^T\beta)+exp(x_{j''}^T\beta)} & \text{if } m'=\{j'\}\end{cases} \quad (13)$$

where $j,j',j''=1,2,3;(j\neq j'\neq j'')$ and $m,m',m''=I,II$.

(A special case): we suppose that there are two nests where one nest includes two alternatives 1,2 and another nest consists just alternative 3,(Table2)

First nest(I)	Second nest(II)
1 , 2	3

Table2:

By noting to (9) $M_{R(NMNL)}(\theta|C)$ for this special case can be rewritten follow

$$M_{R(NMNL)}(\theta|C)=\begin{pmatrix}(X^TDX) & (X^T\Pi)\\ (\Pi^TX) & c\end{pmatrix}$$
$$+P(Y_{(1)}=1,M=I)\begin{pmatrix}X_{(1)}^TD_{(1)}X_{(1)} & 0\\ 0 & 0\end{pmatrix}$$
$$+P(Y_{(1)}=2,M=I)\begin{pmatrix}(X_{(2)}^TD_{(2)}X_{(2)} & 0\\ 0 & 0\end{pmatrix}$$
$$+P(Y_{(1)}=3,M=II)P((Y_{(2)}=1,M'=I)|(Y_{(1)}=3,M=II))\times A$$
$$+P(Y_{(1)}=3,M=II)P((Y_{(2)}=2,M'=I)|(Y_{(1)}=3,M=II))\times B$$

To obtain the matrices in the second, the third , the fourth and the fifth rows of equation (above) it is used of *Goos et al., 2007.*

where

$$\mathbf{A}= \begin{pmatrix} (X_{(3)}^T D_{(3)}^{[1]} X_{(3)}) & (X_{(3)}^T \Pi_{(3)}^{[1]}) \\ (\Pi_{(3)}^{[1]T} X_{(3)}) & c_{(3)}^{[1]} \end{pmatrix}$$

$$\mathbf{B}= \begin{pmatrix} (X_{(3)}^T D_{(3)}^{[2]} X_{(3)}) & (X_{(3)}^T \Pi_{(3)}^{[2]}) \\ (\Pi_{(3)}^{[2]T} X_{(3)}) & c_{(3)}^{[2]} \end{pmatrix}$$

$X_{(j)}^T D_{(j)} X_{(j)}$: Calculating without considering alternative $j = 1, 2, 3$.

$(X_{(3)}^T D_{(3)}^{[1]} X_{(3)})$: Calculating when alternative 1 has the second rank.

$(X_{(3)}^T D_{(3)}^{[2]} X_{(3)})$: Calculating when alternative 2 has the second rank.

Totally, we suppose that there is a $NMNL$ model with two nests which one of them have J_1 alternatives and another nest have J_2 alternatives(Table 3)

nest(I)	nest(II)
1, 2 , ..., J_1	$J_1 + 1$, $J_1 + 2$, ... , $J_1 + J_2$

Table 3: two nest with J_1 and J_2 alternatives,respectively.

According to the dimension of parameters, β (p-dimensional)) and λ (2-dimensional) there are a $(p + 2)$-dimensional vector parameters. Therefore, based on the number of alternatives in each nest, there will be

$$S = \binom{J_1}{1} \times \binom{J_2}{2} + \binom{J_1}{2} \times \binom{J_2}{1} + \binom{J_1}{3} \times \binom{J_2}{0} + \binom{J_1}{0} \times \binom{J_2}{3}$$

choice sets, each of them have three alternatives. Then, it can be made $N = \binom{S}{p''}$ discrete choice experiments (design) that each of them include p'' support points as follow

$$\xi_n = \left\{ \begin{array}{cccc} C_{n1} & C_{n2} & \cdots & C_{n(p'')} \\ w_{n1} & w_{n2} & \cdots & w_{n(p'')} \end{array} \right\} \in \Xi_n; n = 1, 2, ..., N \qquad (14)$$

where $p \leq p'' \leq \frac{(p+2)(p+3)}{2}$. Similarity to rank-order MNL model, the information matrix corresponding to ξ_n is calculated with

$$M(\theta|\xi_n) = \sum_{s=1}^{p''} w_{ns} M_{R(NMNL)}(\theta|C_{ns})$$

based on D_b-optimal criterion

$$D_b(\xi_n) \simeq \frac{1}{R} \sum_{r=1}^{R} det \left\{ \sum_{s=1}^{p''} w_{ns} M_{R(NMNL)}(\theta^{(r)}|C_{ns}) \right\}^{-\frac{1}{p+2}}$$

if

$$D_b(\xi_n^*) = \min_{\xi_n \in \Xi_n} D_b(\xi_n); n = 1, 2, ..., N \tag{15}$$

then

$$\xi_n^* = \left\{ \begin{array}{cccc} C_{n1} & C_{n2} & \cdots & C_{n(p'')} \\ w_{n1}^* & w_{n2}^* & \cdots & w_{n(p'')}^* \end{array} \right\}$$

are D_b-optimal design in $\Xi_n; n = 1, 2, ..., N$ and by noting to (15) a design corresponding to

$$\min_n D_b(\xi_n^*)$$

is the most suitable design.

4 Conclusion

In a rank-order choice experiment, the respondent is asked to rank a number of alternatives in each choice set. This way of performing a conjoint experiment offers the important advantage that extra information is extracted about the preferences of the respondent which results in better estimated part-worths , dissimilarity parameters and better predicted probabilities.

References

[1] Chaloner, K. and Larntz, K. (1986). Optimal Bayesian designs applied to logestic regression experiments. *Universiry of Minnesota Technical Report.*

[2] Chernoff,H. (1972). Sequentail analysis and optimal design. *Society for Industrial and Applied Mathematics , Philadelphia PA.*

[3] Goos, P., Vermeulen, B. and Vandebroek, M. (2007), D-optimal conjoint choice designs for a nested logit model when a no-choice option is used, Unpublished manuscript

[4] Kessels, R. Goos, P. and Vandebroek, M. (2006a). A Comparison of Criteria to design Efficient Choice experiments, *Journal of Marketing Research, 43 , 409-419.*

[5] Pukelsheim, F. (1993). Optimal design of experiments. *Wiley, New York.*

[6] Sandor, Z. and Wedel, M. (2001), Designing conjoint choice experiments using manager's prior belifes, *Journal of Marketing Research , 38 , 430–444.*

[7] Vermeulen, B. Goos, P. Vandebroek, M. Rank-Order Conjoint Experiments: Efficiency and design. *Preprint.*

Search when the lie depends on the target

Gyula O.H. Katona[††‡‡]

Rényi Institute, Hungarian Academy of Sciences, Budapest, HUNGARY

Abstract: In the first half of the lecture we give a short introduction to the theory of *Combinatorial Search* (called also group testing). This includes the model with lies. In the traditional model with lie, the answer to the question "is $x \in A$?" can be wrong (say once during the whole search) independently of x and A. In the second half of the lecture we show a new model in which the answer can be wrong only for some elements x. Solutions for both the adaptive and non-adaptive cases are given.

Keywords: combinatorial search, search with lies, group testing.

1 Introduction

1.1 Introductory examples

1. When soldiers, infected by syphilis were searched, during World War II, Dorfman [4] realized that one does not have to check the blood sample of every single soldier to find the infected ones. Instead, one can pour together (group) several samples, check them together. If the result is negative, we can claim that none of them is infected, one can continue the search without them. On the other hand, if the result of the test is positive, this set of samples (group) has to be subdivided and searched by parts. In this way the number of tests can be remarkable reduced.

2. When Alfréd Rényi, the famous Hungarian mathematician got his first car in Hungary in the early nineteen sixties, the level of car repair was not very high there. When the car broke down by some electric problem, he had to find the source of error, himself. He "had to develop" a theory. It is practically impossible to check each part one by one. However, if one tries to operate a part of the car, and it

[††]Invited speaker
[‡‡]Email: ohkatona@renyi.hu

works, one can suppose that the defective part is not there. Again, tests are performed on a subset of element and the (single) defective element should be found.

3. *Chemical analysis.* A solution contains one metal. It should be determined by using chemical tests, say adding some other chemicals and checking the changes in color. Then, again, we want to find one unknown element of a set (of all metals) and test a subset (of metals changing the color of the solution).

4. *Criminal investigation.* Given a crime, we have a set S of possible perpetrators. The real perpetrator, $x \in S$ should be found. Each evidence restricts x to be in a set $A \subset S$. For instance if a witness says that the perpetrator is bold then we know $x \in A$, where A is the set of bold ones among the possible perpetrators.

1.2 The model of Combinatorial Search

The above examples can be mathematically formulated in the following way. Let X be a finite set and $\subset 2^X$ a family of its subsets. x is an unknown element of X, our aim is to find this element by asking questions of type "is $x \in F$?" where F is chosen from the family . There are two very different ways of choosing the Fs. The choice of the next "question set" might or might not depend on the previous answers.

Adaptive search. If the answer for the starting question "is $x \in F$?" is "no" then the next question is "is $x \in F(0)$?" otherwise "is $x \in F(1)$?". Similarly, the third question is denoted by $F(00), F(01), F(10)$ or $F(11)$ depending on the previous answers. In general, let $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_r$ be a 0,1 sequence. Suppose that the answer for the question "is $x \in F$?" is ε_1, the answer for the next question "is $x \in F(\varepsilon_1)$?" is ε_2, and so on, the answer for the question "is $x \in F(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{r-1})$?" is ε_r. Now two different situations may occur. Either the knowledge of all these answers determines x or not. In the previous case the search is finished with r questions, in the latter case a new question is asked: "is $x \in F(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_r)$?". The procedure can be illustrated by a binary rooted tree with set-labels on its vertices. The label of

the root of the tree is F, the labels of its children are $F(0), F(1)$. In general, if the collected knowledge after answering the question "is $x \in F(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{r-1})$?" (the answer is ε_r) determines x, then the vertex labelled with $\{x\} = F(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_r)$ is a leaf of the tree, otherwise the children of the vertex labelled with $F(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{r-1})$ are labeled with $F(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{r-1}, 0)$ and $F(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{r-1}, 1)$, respectively. The search is moving on the tree from the root towards a leaf. The length of the search is the number of questions needed to arrive to a unique x, in other words the length of the path (measured by the number of vertices) leading to a given leaf. This depends on x.

This binary rooted tree solving the given problem can be called an *algorithm* and will be denoted by $A = A()$. The number of steps (questions) using A is $A(x)$ if the unknown element is x. The *length* of the algorithm is

$$\ell(A) = \max_{x \in X} A(x)$$

(*worst case*). The *complexity* of the problem is

$$c() = \min \ell(A)$$

where the min is taken for all algorithms solving .

Sometimes the length of the algorithm is defined as the average of the values $A(x)$. Both the worst case and the average have practical significance, but the worst case is mathematically treatable easier. This is why most mathematical results use the worst case for defining the length of the algorithm.

Non-adaptive search. Here the choice of the next question cannot depend on the previous answers. That is, the question sets $F_1, F_2, \ldots, F_\ell \in$ are given in advance. The answers for the questions "is $x \in F_i (1 \leq \ell)$?" must determine the unknown element x. It is easy to see that this happens iff for any given pair $x, y \in X, x \neq y$ one of the F_is *separates* x from y, that is either $x \in F_i, y \notin F_i$ or $x \notin F_i, y \in F_i$ holds. Such a subfamily of is called a *separating system*. The length of the "algorithm" here is trivially ℓ, independently on x. The *non-adaptive complexity* of the problem is

$$n() = \min \ell$$

where the minimum is taken for all separating subfamilies $F_1, F_2, \ldots, F_\ell \in$.

Although in the latter case "only" families of subsets are considered and in the adaptive case a more complicated structure, very often it is more difficult to determine $n()$ than $c()$. An example is when consists of all sets of size $\leq k$ [7].

$$\log n \leq c() \leq n().$$

The lower bound is called the *information theoretical lower bound*, the log denotes the one of basis 2 thorough the paper.

Let us give one more example for illustration from computer science. Given distinct integers $z_1, \ldots, z_m$, find their natural order by pairwise comparisons "is $z_i < z_j$?". This problem can be easily included into our model described above. Let X be the set of all permutations of the given integers, that is, of the numbers $1, \ldots, m$. The goal is to find one unknown element of this set X of $n = m!$ elements. A question set is the set of permutations where z_i has a position before z_j. (Its size is $m!/2$.) Proposition 1 can be used: $\log m! \leq c$ holds for the adaptive complexity of this problem. By the Stirling formula $\log m!$ is approximately $m \log m + c_1 m$. On the other hand there is an easy algorithm with length $m \log m + c_2 m$. The true value of the coefficient of m is unknown. The non-adaptive problem here is trivial. One has to compare every pair z_i, z_j, since the two permutations where z_i is the first, z_j is the second and the one where z_j is the first, z_i is the second cannot be separated otherwise. The non-adaptive complexity is $\binom{m}{2}$.

1.3 Variants

There are many variants of the search problems described by the model above. For instance, when the number of unknown elements is not one. E.g. there are typically more than just one infected soldier in the first example. However there must be restrictions either on the maximum number of the unknowns, or on their structure. These problems can be often transformed to fit our model by changing X and suitably.

Another variant is when is replaced by functions f defined on X. Then the questions of type "is x is F?" are replaced by questions of form "what is $f(x)$?". In the third example, e.g. the chemical test might have several different results (color is unchanged, becomes red, becomes blue). The basic model is obtained back if the range of f is $\{0,1\}$, that is the function is 1 on F and 0 on $\overline{F}$.

There is a variant between the adaptive and non-adaptive case. Namely when the number of "rounds" is limited. E.g. the search with two rounds consist a set of questions (subsets) in the first round, then depending on the list of answers another set of questions can be asked. The unknown element has to be found after this. The complexity of the algorithm in this case is the number of questions in the first round + the largest number of questions in the second round.

The "approximate search" does not aim to find the unknown element exactly, only a member of the family of "neighborhoods" containing the searched element. (See [11].)

For practical motivations, other variants and a large number of nice results see the ancient (rather obsolete) survey [8] and the monographs [2], [1], [5].

1.4 Search with lies: Rényi-Ulam game

However the model "with liar" is really different. Then the answers are unreliable: the answer for the question "is $x \in F$?" can be wrong. Typically it is supposed that the number of wrong answers is bounded by a constant. However the result should be exact, no probabilistic. That is, we want to find the unknown element precisely. This problem was independently proposed by Rényi [10] and Ulam [12]. There are many results along these lines, too. See for instance one paper of the speaker [9] in which the sizes of the question sets and the number of possible lies are bounded from above. Two good surveys are written by Hill [6] and Deppe [3].

2 Our model

In the models Search with Lies briefly introduced in subsection 1.4 every question has the same chance to be incorrectly answered. In other words, the occurence of a lie does not depend on the relationship of the question set A and the unknown element x. In our present model this is not true. For a given question there are certain unknowns x triggering the possibility of a false answer. If the unknown x is different from these then the answer must be correct. Let us show some examples continuing our examples in the previous section.

Chemical analysis. The outcome of the chemical test might sensitively depend on a parameter we cannot well control or sense. But only in the case of certain metals. For "good" metals the result of the test is correct, for the "bad" metals however it might be wrong.

Criminal investigation. The officer asks the witness if the perpetrator is bold. The witness might lie only if it is in his/her interest: the perpetrator is his/her relative or friend.

In the first section a question A divided S into two parts: into A and $\overline{A}$. If the answer was "yes" we learned that $x \in A$, if it was "no" then the conclusion was $x \in \overline{A}$. Here a question is a partition of S into three classes: (A, L, B). If $x \in A$ then the answer is "yes" (or 1), if $x \in B$ then the answer is "no" (or 0), finally if $x \in L$ then the answer can be either "yes" or "no". In other words, if the answer "yes" is obtained then we know that $x \in A \cup L$ while in the case of "no" answer the conclusion is $x \in B \cup L$.

The obvious problem is what the fastest algorithm using such questions is. If there is no limitation on the choice of these 3-partitions, then the easy answer is that only partitions with $L = \emptyset$ should be used and we are back to the old, trivial model. Therefore a natural assumption is that every L is large, that is, $|L| \geq k$ holds for every partition we can use.

The adaptive case will be solved in Section 3 by exhibiting the the best algorithm. The non-adaptive case is more difficult. In Section 4 we reduce the problem to a graph theoretical problem: a nearly perfect matching should be found which satisfies the additional condition

that the number of edges in the matching is the same in all directions.

3 The adaptive search

start with the description of an algorithm. The starting question is an arbitrary partition (A, L, B) satisfying $|L| = k, |A| = \lceil \frac{n-k}{2} \rceil, |B| = \lfloor \frac{n-k}{2} \rfloor$. After obtaining the answer the unknown x will be restricted either to $A \cup L$ or to $B \cup L$ where $|A \cup L| = \lceil \frac{n+k}{2} \rceil$,$|B \cup L| = \lfloor \frac{n+k}{2} \rfloor$.

Suppose that x is already limited to a set $Z \subset S$ at a certain stage of the search. The next step of the algorithm will be determined distinguishing two cases depending on the size of Z. However in both cases the new L is chosen to minimize $|Z \cap L|$ since the incorrect answer in L is not interesting outside of Z.

1. $|Z| \leq n-k$. Choose an L of size k to be disjoint to Z. Divide Z into two parts U and V of sizes $\left\lceil \frac{|Z|}{2} \right\rceil$ and $\left\lfloor \frac{|Z|}{2} \right\rfloor$, respectively. Let the next question (A, L, B) in the algorithm be defined by $A = U, B = V \cup (S - Z - L)$.

2. $|Z| > n - k$. Choose L of size k in the following way: $S - Z \subset L$. Divide $Z - L$ into two parts A and B of sizes $\left\lceil \frac{|Z-L|}{2} \right\rceil$ and $\left\lfloor \frac{|Z-L|}{2} \right\rfloor$, respectively. Let the next question (A, L, B) in the algorithm be defined.

After receiving the answer to this last question the unknown element x is restricted to a set Z' of size either $\left\lceil \frac{|Z|}{2} \right\rceil$ or $\left\lfloor \frac{|Z|}{2} \right\rfloor$ in the first case and of size either $\lceil \frac{n-k}{2} \rceil$ or $\lfloor \frac{n-k}{2} \rfloor$ in the second case. (Observe that all these four values are less than $|Z|$.)

The algorithm stops when $|Z|$ becomes 1.

Theorem 1 *This algorithm is the fastest additive search.*

4 The non-adaptive search

In this case the "algorithm" consists of a series of questions

$$(A_1, L_1, B_1), (A_2, L_2, B_2), \ldots, (A_m, L_m, B_m) \tag{4.1}$$

such that the answers to these questions uniquely determine x in all cases. Take two distinct elements $x, y \in S$. If

$$\text{either} \quad x \in A_i, y \in B_i \quad \text{or} \quad x \in B_i, y \in A_i \tag{4.2}$$

holds for the question (A_i, L_i, B_i) we say that this question *really separates* x and y. If (4.2) holds then the answer to this question will be different when x is the unknown element and when it is y. In other words this question distinguishes x and y. On the other hand, if both x and y are in A_i (B_i) then the answer to the question is the same in the two cases (when x is the unknown or it is y). Finally, if one or both x and y are in L_i then we might obtain the same answer in the two cases, this question does not necessarily distinguishes x and y.

One can see form this that the answers to the set of questions (4.1) uniquely determine the unknown x iff (4.2) holds for every pair $x, y \in S$. We say in this case that (4.1) is a *really separating* set of questions. Our goal is to minimize m under the conditions that (4.1) is really separating and $|L_i| \geq k$, for given n, k. Let this minimum be denoted by $N(n, k)$.

It is useful to consider the "characteristic matrix" of the set of questions. The characteristic vector associated with the question (A, L, B) is a vector containing 1, $*$, and 0 in the jth coordinate if the jth element of S is in A, L, B, respectively. Let the $m \times n$ question-matrix Q have the characteristic vector associated with (A_i, L_i, B_i) in its ith row. Condition (4.2) is equivalent to the condition that for any pair of distinct columns of Q there is a row where the entries are 0, 1 or 1, 0 in the crossing points of this row and the two given columns. We say that that such a matrix is $*$*-less separating*. In these terms $N(n, k)$ is the minimum number of rows in an $m \times n$, $*$-less separating 0,$*$,1-matrix containing at least k stars in each row.

It is not hard to prove the following two lemmas.

Lemma 1 *If Q is an $m \times n$, $*$-less separating 0,$*$,1-matrix containing at least k stars in each row then*

$$2km \leq 2^m \tag{4.3}$$

holds.

Lemma 2 *If Q is an $m \times n$, $*$-less separating 0,$*$,1-matrix containing at least k stars in each row then*

$$n + km \leq 2^m \tag{4.4}$$

holds.

It is somewhat surprising that these two easy conditions (Lemmas 4.1and 4.2) are sufficient for the existence of a good Q.

Theorem 2 *($3 \leq m$) A Q $m \times n$, $*$-less separating 0,$*$,1-matrix containing at least k stars in each row exists if and only if both (4.3) and (4.4) hold.*

Sketching why we need here a graph construction. We only have to construct a matrix satisfying the conditions if the inequalities (4.3) and (4.4) hold. The matrix will contain one or zero $*$s in every column, and exactly k $*$s in every row. The 0,1 columns of the matrix will be considered as points of the m-dimensional cube B_m. (Here $B_m = (V, E)$ is a graph where V consists of all 0,1 sequences of length m and two such vertices are adjacent if the sequence differ in exactly one position.) A column containing one $*$ can be considered as a pair of points, namely the points corresponding to the two columns obtained by replacing the $*$ by a 0 and a 1. These points are adjacent in B_m therefore the column containing exactly one $*$ can be considered as an edge of B_m. This edge has a *direction*, namely the index of the position of the $*$. It is obvious that two such edges cannot have a common point, otherwise the two columns would not be different by all substitutions. This shows that our matrix generates a matching in B_m. Since we want to have exactly k $*$s in every row, the number of edges in the desired matching should be the same in every direction.

A subgraph (in our case a matching) of B_m is called *balanced* if the number of edges in every direction is the same. We showed how these concepts came into the picture. Let us now formulate our main tool what is needed to prove Theorem 4.3.

Theorem 3 *$B_m(m \geq 3)$ contains a balanced matching with*

$$\left\lfloor \frac{2^{m-1}}{m} \right\rfloor \tag{4.5}$$

edges in every direction.

The construction proving Theorem 4.3, using Theorem 4.4. Suppose that (4.3) and (4.4) hold. Start with the balanced matching in Theorem 4.4. By (4.3) k cannot exceed (4.5). Keep only k edges of the matching in each direction. If e is an edge of the matching in direction i then take a a corresponding column in Q having a $*$ in the ith row, its other 0,1 entries are the joint coordinates of the two endpoints of e. In this way we obtained an $m \times km$ $*$-less separating matrix. We need to add $n - km$ 0,1 columns (without a $*$) keeping the property. The existing km columns exclude $2km$ columns, what are obtained by replacing the $*$s by 0 or 1. There are $2^m - 2km$ other 0,1 columns for our disposal. However $n - km \leq 2^m - 2km$ follows from (4.4), the construction of Q can be completed.

References

[1] Ahlswede, R. and Wegener, I. (1979). *Suchprobleme*, Teubner; (1987) English translation: Search Problems, John Wiley.

[2] Aigner, M. (1988). *Combinatorial Search*, John Wiley, Chichester and Teubner, Stuttgart.

[3] Deppe, C. (2007). Coding with Feedback and Searching with Lies, in: *Entropy, Search, Complexity* (Ed. I, Csiszár, G.O.H. Katona and G. Tardos, Bolyai Society Mathematical Studies, Springer, **16**), 27-70.

[4] Dorfman, R. (1943) The detection of defective elements in a large group, *Ann. Math. Stat.*, **14**, 436-440.

[5] Du D.-Z. and Hwang F.K. (1993). *Combinatorial Group Testing*, World Scientific, Singapore.

[6] Hill R. (1995). Searching with lies. In: *Surveys in Combinatorics, (Stirling).* London Math. Soc. Lecture Note Ser., 218, Cambridge, Univ. Press, Cambridge, 41-70.

[7] Katona G. (1966). On separating systems of a finite set. *J. Combin. Theory* **1**, 174-194.

[8] Katona G.O.H. (1973). Combinatorial Search Problems, in: *A survey of combinatorial theory* (Ed. by J.N.Srivastava, North Holland/American Elsevier, Amsterdam/New York) 285-308.

[9] Katona G.O.H. (2002). Search with small sets in presence of a liar, *J. Statictical Planning and Inference*, **100**, 319–336.

[10] Rényi, A. (1961). On a problem of information theory, *MTA Mat. Int. Közl.*, **6B**, 505-516.

[11] Wiener G. (2003). Approximate search, Ph.D Thesis, Eötvös University, Budapest.

[12] Ulam, S. (1976). *Adventures of a Mathematician*, Scribner, New York.

Bayesian Estimation of the Parameters of the Bivariate Poisson Regression Models

Soraya Khafri[1], Anoshirvan Kazemnejad[2], and Farzad Eskandari[3]

[1,2]*Department of Biostatistics, Tarbiat Modares University*
[3]*Department of Statistics, Allameh Tabatabai University*

Abstract: This article is concerned with the analysis of correlated count data. A class of model in which the correlation between the counts is presented by correlated unobserved heterogeneity components. A Hierarchical Bayesian analysis is used for estimation of the parameters. A Gibbs sampling algorithm is suggested for find posterior densities of parameters. The proposed method is applied to CBC data.

Keywords: Bivariate count data, Poisson-Lognormal distribution, Hierarchical model, Metropolis-Hasting algorithm, WBC, RBC.

1 Introduction

Multivariate count data appear in a wide range of fields like epidemiology (e.g. different types of a disease), marketing (e.g. purchases of different products) and environ metrics (e.g. different kinds of plantation etc) among others, where incidences of several related events are counted. While the Poisson distribution has played a prominent role in modeling univariate count data, its multivariate counterpart has been rarely used in practice mainly due to computational difficulties in inferential procedures. The use of a multivariate normal model as an approximation to the multivariate Poisson model can be misleading, especially if the counts are small and there are many zero counts.

The multivariate Poisson distribution, while the most important among discrete multivariate distributions, has several shortcomings for its application. The main drawback of the multivariate Poisson model is the complicated form of the joint probability function. This

has led to the use of a simplified model with just one common covariance term for all pairs of variables (Tsionas(1999,2000), Karlis(2003)). In addition, the multivariate Poisson model assumes positive correlation between variables, an assumption that is not realistic for several applications. Finally, the marginal mean and variance of each variable coincide and, thus, this model is not appropriate for overdispersed data sets. Mixed multivariate Poisson models, using some mixing distribution for the parameters as, for instance, multivariate negative binomial models, can solve the problem of overdispersion. Another way to overcome the problem of overdispersion is to include stochastic error term in the model to allow for unexplained randomness in the poisson regression model. Simultaneous equation models and count analyses have been applied independently numerous times and there are only few analyses of correlated count data (Ibrahim J.,et al(2000),Chib and Winkelmann (2001)). When hypotheses of plausible interrelationships between count variables can be developed from theory and prior knowledge, a more directed approach, such as multivariate regression count models, is appropriate. There are several techniques that introduce correlations into multivariate Poisson models. One of the best known techniques employs a stochastic component that shares the specification of the count variables through either multivariate reduction techniques (Karlis(2003)) or the method of mixtures (Jung and Winkelmann(1993), Munkin and Trivedi(1999)). However, the use of an identical component restricts the effects of the factors to be the same across categories. Furthermore, this approach permits only positive correlations between two variables. A more general method allows for separate stochastic components that are correlated for each count variable. The mixing distribution in this context can be either a discrete distribution with several support points (finite mixture) or a parametric continuous distribution (continuous mixture). The finite mixture approach can flexibly accommodate extreme and/or strongly asymmetric departures from the Poisson model (Alfo and Trovato (2004)). However, finite mixture models may inadequately represent the full extent of heterogeneity and it is difficult to estimate them with

more than a half dozen or so mass points. Allenby and Rossi(1999) discuss and provide empirical evidence of the shortcomings of the finite mixture. An alternative parametric specification for the mixing distribution is the use of multivariate normal distribution, which leads to a multivariate Poisson lognormal model that can be estimated by either simulated maximum likelihood (Munkin and Trivedi(1999)) or MCMC method (Chib and Winkelmann(2001)). However, the simulated maximum likelihood method makes it difficult to extend the model beyond the case of a few outcomes, which may limit the scope of data mining applications. In contrast, Bayesian inference using the MCMC method can efficiently accommodate high-dimensional count data while still allowing the model to maintain a general correlation structure. Ma and Kockelmann(2006) used Gibbs sampling as well as Metropolis-Hasting algorithms, within a MCMC simulated framework. They used single prior model in their study. In our study, we follow their work and used the MCMC method for Estimation with a hierarchical model. The rest of the article is organized as follows: In Section 2 we present the basic model and some special cases and extensions. The fitting algorithm is developed in Section 3, while Section 4 gives real data example.

2 The Model

This section briefly reviews bivariate count models related to the main models of this paper. Suppose that we observe $y_{ij}, i = 1, 2, \ldots, n_j, j = 1, 2$ denote the value of counts for individual i and outcome j. let $y_i = (y_{i1}, y_{i2})$ denote the vector of counts for individual i over the different outcomes. Interest is in a class of models where observations are uncorrelated across individuals but correlated over outcomes:

$$cov(y_{i1}, y_{i2}) \begin{cases} = 0 & \text{for } i \neq k \\ \neq 0 & \text{for } i = k, j \neq l. \end{cases}$$

y_{ij} may be assumed to be distributed as a poisson random variables with the mean rate parameter λ_{ij} That is $(y_{ij}|x_{ij}, b_{ij}) \sim poisson(\lambda_{ij})$

where λ_{ij} is specified as:

$$\lambda_{ij} = exp(x'_{ij}\beta_j + b_{ij}) \tag{1}$$

x_{ij} represents a $k \times 1$ explanatory variable vector and β_j is a $k \times 1$ vector of corresponding parameters, including an intercept term β_0. We model the dependence between y_1 , y_2 by means of correlated unobserved heterogeneity components b_{i1} , b_{i2} and assume that $\overrightarrow{b_{\mathbf{i}}} = (b_{i1}, b_{i2})$ is distributed as a bivariate normal random variable and use a general variance-covariance structure to accommodate the correlation among the b_{ij}'s. That is $\overrightarrow{b_{\mathbf{i}}} \sim N_2(0, \Sigma)$; where Σ is an unrestricted covariance matrix. The model specification in above equations implies that λ_i follows a bivariate lognormal distribution and y_i follows a bivariate Poisson-lognormal distribution.

2.1 Likelihood Function

The probability density function for count vector y_i is given by

$$f(y_{i1}, y_{i2}|x_i, \beta_j, \mathbf{\Sigma}) = \int \prod_{i=1}^{2} f(y_{ij}|\beta_j, b_{ij}) \phi_j(\overrightarrow{b_{\mathbf{i}}}|0, \Sigma) d\overrightarrow{b_{\mathbf{i}}} \tag{2}$$

Given $\mathbf{\Sigma}$, the multiple integrals cannot be solved in a closed form. Therefore, the MCMC method under the Bayesian framework can be used to compute the integrals.

In this article we use a hierarchical Bayesian analysis of bivariate poisson log normal models. The hierarchical model can be written in two levels. At level I, Supposed that the parameters (β, Σ) independently have the prior distributions:

$$\beta \sim N_k(\beta_\beta, V_\beta^{-1}) \tag{3}$$

$$\Sigma^{-1} \sim Wishart(\upsilon_\Sigma, V_\Sigma) \tag{4}$$

And in the 2nd level, supposed:

$$\beta_\beta \sim N_k(\beta_0, Sig\beta) \tag{5}$$
$$V_\beta^{-1} \sim Wishart(\upsilon_{0_\beta}, R_0) \tag{6}$$
$$V_\Sigma^{-1} \sim Wishart(\upsilon_{0_\Sigma}, \Sigma_0), \tag{7}$$

where $(\beta_0, Sig\beta, \upsilon_{0_\beta}, \upsilon_{0_\Sigma}, \Sigma_0, R_0)$ are known parameters. In this approach, we suppose β_0 and $Sig\beta$ are independent. Figure 1 shows the model in the form of a direct acyclic graph (DAG).

2.2 Posterior Distribution

Since, the marginal distribution cannot be algebraically implemented in closed form, given the observations, the joint posterior distribution of the parameters is obtained by combining (2), (3), (4), (5), (6), and (7) via Bayes' theorem. So, we have:

$$\begin{aligned}
&\pi(\beta_j, \beta_\beta, V_\beta, \overrightarrow{b_\mathbf{i}}, \Sigma | y_i, X) \propto L(\beta_j, \overrightarrow{b_\mathbf{i}} | y_i, X) P(\beta_j, \beta_\beta, V_\beta, \overrightarrow{b_\mathbf{i}}, \Sigma, y_i) \\
&\propto L(\beta_j, \overrightarrow{b_\mathbf{i}} | y_i) P(\overrightarrow{b_\mathbf{i}} | \Sigma) p(\Sigma^{-1} | \upsilon_\Sigma, V_\Sigma) g(\upsilon_\Sigma^{-1}) P(\beta_j | \beta_\beta, V_\beta,) g(\beta_\beta) g(\upsilon_\beta^{-1}) \\
&\propto \prod_{i=1}^{n} \prod_{j=1}^{2} exp\Big(- exp(x'_{ij}\beta_j + b_{ij})\Big) \Big(exp(x'_{ij}\beta_j + b_{ij})\Big)^{y_{ij}} \\
&\times \prod_{i=1}^{n} exp(-1/2 b'_i \Sigma^{-1} b_i) |\Sigma^{-1}|^{1/2} . |Sig\beta|^{-1/2} \\
&\times \frac{|\Sigma^{-1}|^{\frac{\upsilon_\Sigma - 2}{2}} exp\Big(- 1/2 tr(V_\Sigma^{-1} \Sigma^{-1})\Big)}{|V_\Sigma|^{\frac{\upsilon_\Sigma}{2}}} \\
&\times \frac{|V_\Sigma^{-1}|^{\frac{\upsilon_{0_\Sigma} - 2}{2}} exp\Big(- 1/2 tr(V_\Sigma^{-1} \Sigma_0^{-1})\Big)}{|\Sigma_0|^{\frac{\upsilon_{0_\Sigma}}{2}}} \\
&\times \prod_{j=1}^{2} |V_\beta|^{-1/2} exp\Big(- 1/2 (\beta_j - \beta_\beta)' V_\beta^{-1} (\beta_j - \beta_\beta)\Big)
\end{aligned}$$

$$\times exp\Big(-1/2(\beta_\beta-\beta_0)'Sig\beta^{-1}(\beta_\beta-\beta_0)\Big)$$
$$\times|V_\beta^{-1}|^{\frac{\upsilon_{0_\beta}-2}{2}}.\frac{exp\Big(-1/2tr(R_0^{-1}V_\beta^{-1})\Big)}{|R_0|^{\frac{\upsilon_{0\beta}}{2}}} \quad (8)$$

3 Gibbs sampler for hierarchical model

Gibbs sampler for hierarchical model Since the posterior distribution is intractable, we seek a Markov Chain that has the posterior as its long run distribution. Sampling from this Markov Chain after an adequate burn-in period will enable us to approximate a sample from the posterior distribution (Gilks Richardson and Spiegelhalter (1996)). The conditional distributions are determined using a graphical modeling approach (Borgelt and Kruse(2002)). Using this approach in a hierarchical model, the conditional distribution of one node given all the other nodes is proportional to the product of the prior distribution of the node and the conditional distribution of all its direct child nodes and co-parent nodes. Due to the conjugate property, the complete conditional distribution of the parameters is given in two sections:

Section I; has the closed form:

$$\Sigma^{-1}|b\sim Wishart\Big((\upsilon_\Sigma+n),(V_\Sigma^{-1}+\sum_{i=1}^{n}b_i'b_i)^{-1}\Big)$$

$$V_\beta^{-1}|\beta\sim Wishart\left((\upsilon_{0_\beta}+2),\left(R_0^{-1}+\sum_{j=1}^{2}\Big[(\beta_j-\beta_\beta)'(\beta_j-\beta_\beta)\Big]\right)^{-1}\right)$$

$$V_\Sigma^{-1}|\Sigma\sim Wishart\left((\upsilon_{0_\Sigma}+\upsilon_\Sigma),\left(\Sigma^{-1}+\Sigma_0^{-1}\right)^{-1}\right)$$

Section II; has not a known closed form which makes direct inferences almost impossible. MCMC methods are techniques that have been developed to resolve this kind of problem.

Draws from this conditional density can be obtained by developing an M-H algorithm by using proposal densities as follow:

$$q(\overrightarrow{b_{\mathbf{i}}}|y_i,\beta_j,\Sigma) = N_2\Bigg\{\Big[\Sigma^{-1}\overrightarrow{b_{\mathbf{i}}} + \Big[y_i - exp(x_i\beta_j + \overrightarrow{b_{\mathbf{i}}})\Big], \Big[\Sigma^{-1} + diag(exp(x_i\beta_j + \overrightarrow{b_{\mathbf{i}}}))\Big]^{-1}\Bigg\}$$

$$q(\beta_j|y_i,X,\overrightarrow{b_{\mathbf{i}}},\Sigma) = N_k\Bigg\{\Big[-V_\beta^{-1}(\beta_j-\beta_{0j})+\sum_{i=1}^{n}(y_{ij}-exp(x'_{ij}\beta_j+b_{ij}))x_{ij}\Big], \Big[V_{\beta_{0j}}^{-1} + \sum_{i=1}^{n}(exp(x'_{ij}\beta_j + b_{ij}))x_{ij}x'_{ij}\Big]^{-1}\Bigg\}$$

$$q(\beta_\beta|\beta_j,V_\beta) = N_k\Bigg\{\Big[\sum_{j=1}^{2}\beta'_jV_\beta^{-1} + \beta'_0Sig\beta^{-1} + \beta'_\beta(-V_\beta^{-1} - 1/2Sig\beta^{-1})\Big], \Big[V_{\beta_\beta}^{-1} + 1/2Sig\beta^{-1})\Big]^{-1}\Bigg\}$$

4 Real Data

4.1 Data Description

A total of 203 children who had 6-12 years old, BMI>95 percentile for age and sex, and attended health examination of their volition from 2006 to 2007 at Sedighe Tahere Hospital were enrolled in this study. During the physical examination, body weight (kg) and height (m) of subjects were measured for computing body mass index (BMI). Plasma total triglyceride (TG), cholesterol, and HDL were measured by enzymatic method using a chemistry analyzer at the central laboratory of the hospital. Total WBC, RBC, and platelet counts were computed by means of an auto analyzer.

Metabolic syndrome (MS) is characterized by obesity, impaired glucose tolerance (or type 2 diabetes), hypertension, low high-density lipoprotein (HDL) cholesterol levels, and/or hypertriglyceridemia.

This syndrome is now recognized as one of the most common disorders worldwide. In addition, there is growing evidence that MS is associated with an increased risk of cardiovascular disease and is now considered as the secondary target for treatment of coronary heart disease. Insulin resistance and/or hyperinsulinemia has been shown in some studies to correlate with white blood cell (WBC) or red blood cell (RBC) counts. In fact, several epidemiological studies have already noted a relationship between some components of MS, such as lipid abnormalities and high blood pressure, and WBC, or RBC counts. The aim of the present study is to investigate the association between various hematological parameters with components of MS.

WBC and RBC are the main two outcomes and likely to affect each other. So, their relationship needs to be analyzed with a simultaneous equation model. A bivariate poisson regression model is plausible due to the correlation. In this study, the influence of several factors on the WBC and RBC were investigated simultaneously. These factors were: age, BMI, gender, Cholesterol, LDL, HDL, Triglyceride (TG), FBS, waist, hip. Table 1, contains descriptive statistics for the CBC count data. There is a clear indication of extra poisson variation in the data since the variances are much greater than the means.

4.2 Model Estimation

The bivariate poisson regression model was estimated using a hierarchical Bayesian approach. The starting values for β came from the method of maximum likelihood estimation (MLE). The starting values for Σ are $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

A Gibbs sampler and M-H algorithm were performed with Winbugs version 1.4.3 (an open-source statistical computing environment described at http://www.mrc-bsu.cam.ac.uk/bugs). The prior distribution for the estimation are defined by the hyper parameters $\beta_\beta = (0, 0, \ldots, 0)'$, $\beta_0 = (0, 0, \ldots, 0)'$, $V_\beta^{-1} = I_{22}$, $Sig\beta = 1.0E{-}6 \times I_{22}$, $R_0^{-1} = 0.01 \times I_{22}$ and $V_\Sigma^{-1} = 0.01 \times I_2$.

In line with the proposed model, a Bayesian analysis was car-

Table 1. Summary statistics of variables

Dependent Variable Mean ±SD	
Count of WBC	12.36± 17
Count of RBC	5.47± 4.51
Independent Variable Mean ±SD	
Age	9.20± 2.62
BMI	23.41± 3.97
Gender	1.52± 0.50
FBS	87.42± 8.75
Cholesterol	182.08 ±29.62
LDL	110.54± 27.55
HDL	46.49 ±8.97
TG	132.29 84.20
Waist	89.80± 77.66
Hip	90.58± 11.24

ried out using the Gibbs sampler. Our final inferences were based upon single long runs of length 100,000, after convergence was monitored through the use of graphical aids and the convergence diagnostic of Gelman and Rubin (16) by a chain with 50,000 iterations.

4.2 Results

Table 2 gives several summary measures of the posterior distribution. In addition to the posterior mean and standard error, we also display the 2.5th and 97.5th percentile of the marginal posterior distribution or 95% High Posterior Density (HPD) regions for parameters in the model. Figure 1 shows the marginal posterior distributions for parameters in the proposed model. The estimate for a mostly all of the parameters except for the Gender and TG appear to be bimodal; this suggest the model is uncertain and certainly model is necessary.

Parameter estimation shown in Table 2 suggests that however, the statement of the children and hormonal parameters play important roles in predicting the count of WBC, only the cholesterol and LDL

Table 2. Posterior summary for parameters based on the MCMC simulation output from purposed model

Variable definition	Mean	Std.Err.	95% HPD regions Lower	Upper
Response variable: WBC				
Constant	1.673	0.0690	0.645	2.770
Age	-0.018	0.0020	0.001	0.024
BMI	0.063	0.0020	0.034	0.100
Gender	0.187	0.0080	0.031	0.360
HDL	0.018	0.0015	0.006	0.030
TG	0.002	0.0001	0.001	0.003
Waist	0.002	0.0005	0.000	0.003
Hip	-0.025	0.0008	-0.039	-0.015
Response variable: RBC				
Constant	1.851	0.0436	1.136	2.662
Cholesterol	-0.006	0.0002	-0.010	-0.002
LDL	0.007	0.0002	0.003	0.012

among hormonal parameters are predictors for predicting the RBC rate. For example, holding all other factors fixed, more WBC are expected on increasing BMI, HDL, TG, Waist and decreasing the age of the children. Furthermore, girls have more WBC counts than boys in children between 6-12 years old.

In addition fitted model showed increasing the LDL and decreasing the cholesterol tend to increase the rate of RBC.

Based on the average parameter estimates of the bivariate poisson lognormal regression model, Table 3 provides estimates of percentage changes in the hematological parameters as a function of various statement details. For example, 1-year increases in (average) the age (from 9 to 10 years) is predicted to result in 1.8% fewer WBC count.

Table 3. Expected percentage changes in the count of WBC and RBC corresponding to changes in variables

variables	Changes In variable	Percentage changes in two outcomes	
		WBC	RBC
Age	1 (years)	-1.8%	———
BMI	2 (kg/m2)	12.6%	———
Gender	1/0	18.7%	———
HDL	2	3.6%	———
LDL	2	———	1.4%
Cholesterol	2	———	-1.2%
TG	2	0.4%	———
Waist	5	1.0%	———
Hip	2	12.5%	———

References

[1] Alfo M. and Trovato G. (2004). Semiparametric mixture models for multivariate count data with application.*Econometrics Journal*, **7**, 426-54.

[2] Allenby G.M. and Rossi P.E.(1999). Marketing models of consumer heterogeneity Journal of Econometrics 1999; 89: 57-78.

[3] Borgelt C. and Kruse R. (2002). Graphical Models: Models for data analysis and mining, Wiley, New York.

[4] Chib S.and Winkelmann R. (2001). Markov chain monte carlo analysis of correlated count data.*Journal of Business and Economic Statistics*, **19**, 428-35.

[5] Gilks W.R., Richardson S. and Spiegelhalter D.J.(1996). Introducing Markov Chain Monte Carlo in practice, Chapman and Hall, London, 1-19.

[6] Ibrahim J., Chen MH and Ryan L.(2000). Bayesian variable selection for time series count data. *Statistica Sinica*,**10**, 971-87.

[7] Jung R.C. and Winkelmann R. (1993).Two aspects of labor mobility: a bivariate Poisson regression approach Empirical. *Economics*, **18**, 543-56.

[8] Karlis D. (2003). An EM algorithm for multivariate Poisson distribution and related models.*Journal of Applied Statistics*,**1**(30): 63-77.

[9] Ma J. and Kockelman K.(2006). Bayesian multivariate Poisson regression for models of injury count. by severity Transportation Research Record, 24-34.

[10] Munkin M.K. and Trivedi P.K.(1999). Simulated maximum likelihood estimation of multivariate mixed-Poisson regression models. *with application Econometrics Journal*, **2**, 29-48.

[11] Tsionas, E.G. (1999). Bayesian analysis of the multivariate Poisson distribution. *Commun. Stat. Theory Method*, **28**, 431-51.

[12] Tsionas E.G. (2000). Bayesian multivariate Poisson regression. *Communications in Statistics-Theory and Methods*,**30**(2): 243-55.

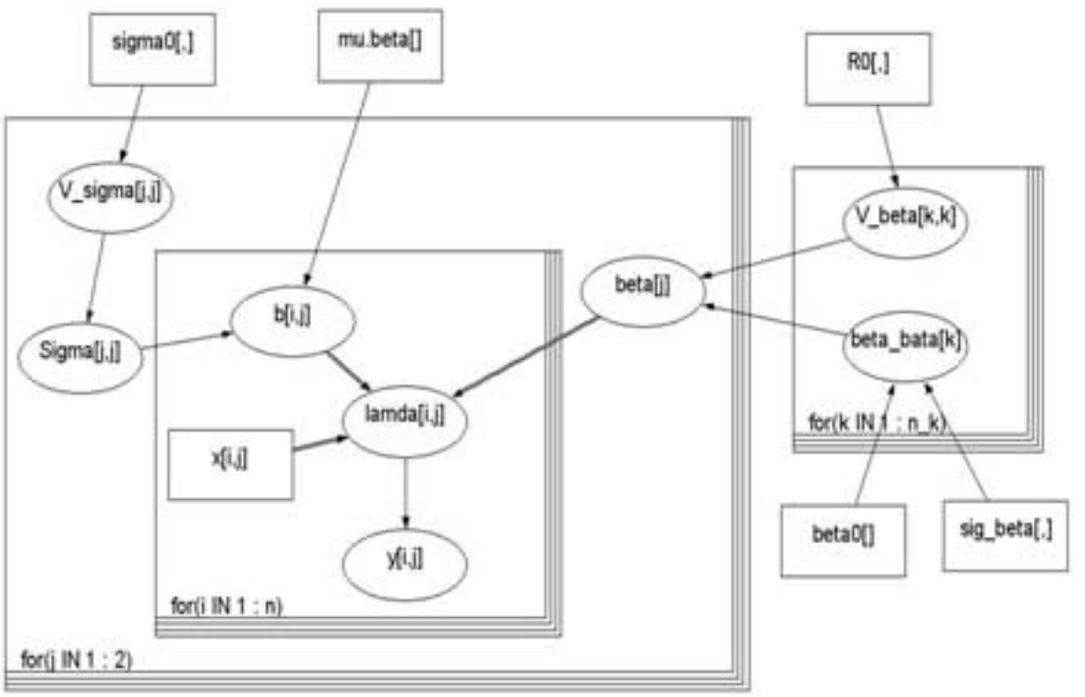


Figure. 1DAG of the bivariate poisson lognormal model.Circles represent parameters, squares represent data, large square plates represent loops, arrows represent stochastic relationships, and bold arrows represent deterministic relationships.

Inference Bayesian Nonparametric of densities under constrains of form

S. Khazaei

CEREMADE, Universit Paris Dauphine, 75016 Paris

Abstract: Recent advance in statistical computing have generated renewed interest in nonparametric inference. In this paper we propose to study and build bayesian nonparametric methods for estimating densities under form constrains. In particular, we determine general conditions on the good asymptotic properties for posterior distributions.

Keywords: Nonparametric Bayesian inference, Consistency, entropy, Kullback Leibler, k-monotone density, kernel mixture.

1 Introduction

In this paper we study the estimation problem of a monotone non increasing density. This is a well known problem, see for instance Balabdaoui and Wellner (2004) for a review on the subject. In particular it has application in fiability or as a preliminary analyses in a survival estimation problem. Decreasing densities on $\mathcal{R}^+$ have a mixture representation which allows for likelihood based inference. There is a vast litterature on maximum likelihood estimation of such densities, see for instance Balabdaoui (2006) for a review. Given the mixture representation Bayesian methods are also appropriate for their estimation. Here we study the asymptotic properties of Bayesian nonparametric estimators of a decreasing density.Non parametric Bayesian procedures presents a paradox. On the one hand they do not require precise information of the shape of the parameter (function), on the other hand, they require the construction of a distribution on an infinite dimensional space, which cannot be done in a purely subjective way. As argued by Diaconis and Freedman(1986), these nonparametric methods are a little value if they don't lead to consistent posterior distributions. Many different types of priors have been used in the litterature on Bayesian nonparametric estimation of densities; priors based

on mixture models are particularly popular. For instance, Ferguson (1983) and Lo (1984) have introduced Dirichlet mixtures, Petrone, S and Wasserman, L.(2002) has studied the properties of Bernstein polynomials, Robert and Rousseau (2003) and Rousseau (2007, 2008) have obtained consistency and rates of convergence for mixtures of Betas, Perron and Mengersen (2001) and McVinish et al. have studied mixtures of triangulars and more generally Wu and Ghosal (2008) have studied approximating properties of general mixture models.In this paper we use the mixture representation of constraint densities, namely the k-monotone densities to construct Bayesian estimators in these models. More specifically we focus on decreasing densities on $\mathcal{R}^+$. There is a theorem that implies a density on $\mathcal{R}^+$ is positive if and only if it can be written as a mixture of uniform densities densities $\mathcal{U}(0,\theta)$ in the form

$$f_P(x) = \int k(x,\theta)dP(\theta) = \int \frac{1}{\theta}\mathcal{I}_{(0,\theta)}(x)dP(\theta) \tag{1}$$

where P is the mixing distribution. Let $\mathcal{F}$ be the set of decreasing densities on $\mathcal{R}^+$, we construct a prior on $\mathcal{F}$ by determining a prior on the mixing distribution P living on $\mathcal{R}^+$.

In Section 2 we give general results to obtain weak and strong consistency of the posterior distribution and in Section 3 we apply these results to the spectial case of Dirichlet mixtures.

2 Concicteny properties of the posterior distribution

Let Π be a probability on $\mathcal{M}$, the set of probability distributions on $\mathcal{R}^+$. We denote by f_P the decreasing density with mixing distribution P:

$$f_P(x) = \int_0^\infty \frac{\mathcal{I}_{[0,\theta[}(x)}{y}dP(\theta).$$

Let $X_1, ..., X_n$ be n independent and identically distributed observations from a distribution having a decreasing density f with respect

to lebegue measure. Then the posterior probability of any measurable set A of $\mathcal{F}$ is given by

$$\Pi(A|X_1,...,X_n) = \frac{\int_{\mathcal{A}} \prod_{i=1}^n f_P(X_i) d\Pi(df_P)}{\int_{\mathcal{F}} \prod_{i=1}^n f_P(X_i) d\Pi(P)}.$$

In this Section we consider two types of consistency : the weak and the strong consistency of the posterior.

Recall the two following definitions:

Definition 1 *A prior* Π *is said to be weakly consistence at* f_0 *if with* P_0*-probabilty 1*

$$\Pi(U|x_1,...,X_n) \to 1$$

for all weak neighborhood U *of* f_0*.*

Definition 2 *A prior* Π *is said to be strongly consistence at* f_0 *if with* P_0*-probabilty 1*

$$\Pi(U|X_1,...,X_n) \to 1$$

for all strong neighborhood U *of* f_0*.*

Recall also that we say that $f \in \mathcal{F}$ is in the support of Π for some topology if the prior mass of any neighbourhood of f (with respect to the same topology) is positive. For any $f_0 \in \mathcal{F}$, we denote the Kullback- Leibler neighborhood $\left\{ f_P \in \mathcal{F} : F_0 \log \frac{f_0}{f_P} < \epsilon \right\}$ by $K_\epsilon(f_0)$, where $F_0 g = \int f_0(x) g(x) dx$ for all integrable function g. Thus f_0 is in the K-L support of Π if $\Pi(K_\epsilon(f_0)) > 0$ for all $\epsilon > 0$.

To obtain weak consistency it is enough to prove that f_0 is in the K-L support of Π, i.e.

$$\forall \epsilon > 0 \qquad \Pi(P; K(f_0, f_P) < \epsilon) > 0$$

Appendix Theorem 3.1.

Yuefeng, Wu and S.Ghosal (2008) obtained a very several results on weak consistency for decreasing densities . They proved that if $\int f_0(x) |log f_0(x) dx| < \infty$ and if the weak support of Π is $\mathcal{M}$, then

the posterior is weakly consistent.However the condition on the weak support of Π is very strong and quit difficult to prove. As an example consider a very natural prior puts mass on finite mixture of uniform densities in the form

$$dP(\theta) = \sum_{j=1}^{K} P_j \delta(\theta_j), \quad d\Pi(P) = P(k) d\Pi_{k,1}(P_1, ..., P_K) d\Pi_{k,2}(\theta_1, ..., \theta_K)$$

where $K \sim P$, and given $K, P_1, ..., P_k \sim \Pi_{k,1}$ and $\theta_1, ..., \theta_k \sim \Pi_{k,2}$. $\Pi_{k,1}$ is absolut continuous with respect to Lebesge on $\{P_1, ..., P_k : P_i \geq 0, \sum P_i = 1\}$ Then if $f_0(x) > 0$ for all $x \in \mathcal{R}^+$ and decreasing

$$\Pi(P : K(f_0, f_P = \infty) = 1$$

Because if $dP(\theta) = \sum_{j=1}^{K} P_j \delta(\theta_j), 0 < \theta_1 < ... < \theta_k$ then for all $x > \theta_k \quad f_P(x) = 0$. This implies that such a prior does not have $\mathcal{M}$ as a weak support, However it is much easier to prove that $\Pi(P : K(f_0, f_P = \infty) = 1$ then $\text{supp}(\Pi) \subset \mathcal{M}$.
In this paper we propose a new set of condition on Π to obtain weak and strong consistency under a stronger condition on f_0. To obtain strong consistency (L_1- Hellinger distance), in addition to the K-L support condition we need a condition on the entropy of the support of Π.

Definition 3 *Let $\mathcal{G} \subset \mathcal{F}$, for $\delta > 0$, the L_1-metric entropy $J(\delta, \mathcal{G})$ is defined as the logarithm of the minimum of all k such that there exist $f_1, f_2, ..., f_k$ in L_μ with the property*

$$\mathcal{G} \subset \bigcup_{i=1}^{k} \{f : \|f - f_i\| < \delta\}.$$

Theorem 1 *Let Π be a prior on $\mathcal{M}$ which puts positive mass on any weak neigborhood of P_0. Suppose $\exists 0 < L < \infty$ such that $P_0[0, L] = 1$, that f_0 is continuous on $[0, L[$and that $f_0(L) > 0$ then the posterior distribution is weakly consistent at f_0.*

Proof: By the Theorem 2.2 it is enough to show that f_0 is in the K-L support of Π. We have

$$K(f_0, f_P) = \int_0^\infty f_0(x) \log \frac{f_0(x)}{f_P(x)} dx,$$

and for any $0 < \tau_0 < L$ then

$$\begin{aligned} K(f_0, f_P) &\leq \int_0^{\tau_0} f_0(x)(\log f_0(x) - \log f_P(x))dx \\ &+ \int_{\tau_0}^\infty f_0(x) \left|\log f_0(x) - \log f_P(x)\right| dx \\ &= I_1 + I_2 \end{aligned}$$

First consider I_1:

case1: Let $\forall t > 0$ there exists $x \in [0, t[$ such that $f_0(x) > f_0(t)$.

By the lemma(3.1) in Appendix, let $\epsilon > 0$ and $\tau < \tau_0$ be such that $f_0(\tau) \geq f_0(\tau_0) + \epsilon$, then for all $x \in [0, \tau]$

$$\frac{f_P(x) - f_0(x)}{f_0(x)} > -c$$

where $c = \frac{f_0(0) - f_0(\tau_0)}{f_0(0)}$ and $P \in U_{(\epsilon,\tau)}$ and $U_{(\epsilon,\tau)}$ is defined by $U_{(\epsilon,\tau)} = \{P \in \mathcal{M} : |f_P(\tau) - f_0(\tau)| < \epsilon\}$. Then

$$-\log\left(\frac{f_P(x) - f_0(x)}{f_0(x)} + 1\right) \leq -\log(1 - c) = m.$$

Then $\forall \tau \leq \tau_0$

$$\begin{aligned} I_1 = \int_0^{\tau_0} f_0 \log\left(\frac{f_0(x)}{f_P(x)}\right) dx &= \int_0^{\tau_0} -f_0(x) \log\left(\frac{f_P(x) - f_0(x)}{f_0(x)} + 1\right) dx \\ &\leq mF_0(\tau_0) = F_0(\tau_0) \log\left(\frac{f_0(0)}{f_0(\tau_0)}\right) \end{aligned}$$

Since $\log\left(\frac{f_0(0)}{f_0(\tau_0))}\right)$ is bounded for τ_0 close to 0, we can choose τ_0 such that

$$F_0([0,\tau_0]) \leq \frac{\epsilon}{2m}$$

then

$$\int_0^{\tau_0} f_0(x)\log\frac{f_0(x)}{f_P(x)}dx \leq \frac{\epsilon}{2}.$$

case2: There exists $t_0 > 0$ such that for all $x \in [0,t[$ such that $f_0(x) = f_0(t_0)$. If $\epsilon < f_0(0)/2$ and $P \in U_{(\epsilon,t_0)}$ then for all $\tau \leq t_0$

$$f_P(\tau) \geq f_P(t_0) > f_0(\tau) - \epsilon = f_0(0) - \epsilon$$

Therefore, let $A = \{x, f_P(x) \leq f_0(x)\}$,

$$\begin{aligned}
\int_0^{\tau_0} f_0\log\frac{f_0(x)}{f_P(x)}dx &= \int_0^{\tau_0} f_0(x)\log(\frac{f_0(x)-f_P(x)}{f_P(x)}+1)dx \\
&\leq \int_0^{\tau_0} f_0(x)(\frac{f_0(x)-f_P(x)}{f_P(x)})\mathcal{I}_A(x)dx \\
&\leq \frac{f_0(0)}{f_0(0)-\epsilon}\int_0^{\tau_0}(f_0(x)-f_P(x))\mathcal{I}_A(x)dx \\
&\leq \frac{\epsilon f_0(0)}{f_0(0)-\epsilon}\tau_0 = \frac{\epsilon}{2}
\end{aligned}$$

if $\tau_0 \leq \frac{f_0-\epsilon}{2f_0}$. We consider I_2.

$$\begin{aligned}
I_2 = \int_{\tau_0}^{L} f_0(x)\log\frac{f_0(x)}{f_P(x)}dx &= \int_{\tau_0}^{L} f_0(x)\log(\frac{f_0(x)-f_P(x)}{f_P(x)}+1)dx \\
&\leq \int_{\tau_0}^{L} f_0(x)(\frac{f_0(x)-f_P(x)}{f_P(x)})\mathcal{I}_A(x)dx \\
&= \sum_{j=1}^{N}\int_{s_j}^{s_{j+1}} f_0(x)(\frac{f_0(x)-f_P(x)}{f_P(x)})\mathcal{I}_A(x)dx \\
&= \sum_{j=1}^{N}\int_{s_j}^{s_{j+1}} \mathcal{I}_A(x)\frac{f_0(x)}{f_P(x)}(f_0(x)-f_P(x))dx
\end{aligned}$$

where $A = \{x | f_0(x) \geq f_P(x)\}$ and $\tau_0 = s_1 < ... < s_{N+1} = L$. Now for every $x \in (s_j, s_{j+1})$

$$\begin{aligned} f_0(x) - f_P(x) &= f_0(x) - f_0(s_{j+1}) + f_0(s_{j+1}) - f_P(x) \\ &\leq f_0(s_j) - f_0(s_{j+1}) + f_0(s_{j+1}) - f_P(s_{j+1}). \end{aligned}$$

Define $I_{\epsilon'}(x) = \left\{y : |f_0(x) - f_0(y)| < \epsilon' f_0(x)\right\}$, by continuity of f_0, since $\forall x \in [0, L], f_0(x) > 0$, $I_{\epsilon'}(x)$ is open and by compactness on $[0, L]$ $\exists s_1, ..., s_{N+1}$ so that $\forall x \in [\tau_0, L]$ $\exists j \in \{1, ..., N+1\}$ such that $x \in I_{\epsilon'}(s_j)$. Define

$$E = \bigcap_{j=1}^{N} U_{(\epsilon' f_0(s_{j+1}), s_j)}$$

Then for every $P \in E$ and $\forall x \in [s_j, s_{j+1}]$

$$f_0(x) - f_P(x) < \epsilon' f_0(s_{j+1})$$

and

$$f_P(x) > f_P(s_{j+1}) > f_0(s_{j+1}) - \epsilon' f_0(s_{j+1}) = (1 - \epsilon') f_0(s_{j+1})$$

and

$$\frac{f_0(x)}{f_P(x)} = 1 + \frac{f_0(x) - f_P(x)}{f_P(x)} \leq 1 + \frac{\epsilon' f_0(s_{j+1})}{f_P(x)} < \frac{1}{1 - \epsilon'}.$$

Hence

$$\begin{aligned} \int_{\tau_0}^{L} f_0(x) \log \frac{f_0(x)}{f_P(x)} dx &\leq \sum_{j=1}^{N} \int_{s_j}^{s_{j+1}} \mathcal{I}_A(x) \frac{f_0(x)}{f_P(x)} \epsilon' f_0(x) dx \\ &\leq \sum_{j=1}^{N} \int_{s_j}^{s_{j+1}} \mathcal{I}_A(x) \frac{\epsilon'}{1 - \epsilon'} f_0(x) dx \\ &\leq \frac{\epsilon'}{1 - \epsilon'} \end{aligned}$$

where $\epsilon^{'} < 1$. If we define $\epsilon^{'} = \frac{\epsilon}{2+\epsilon}$ then

$$\int_{\tau_0}^{L} f_0(x) \log \frac{f_0(x)}{f_P(x)} dx \leq \frac{\epsilon}{2}$$

.

Now we consider the case where $f_0(L) = 0$.

Theorem 2 *Let $f_0(x)$ is continuous and decreasing on $\mathcal{R}^+$ such that $\exists L > 0$ with $f_0(L) = 0$. Define $E_{(L,\delta)} = \{P : f_P(L) > \delta\}$ for $\delta \in]0, 1[$. If $\forall 0 < \tau_1 < \tau_2 < ... < \tau_J < L$, $\forall \epsilon > 0$ there exists $D > 0$ such that*

$$\Pi(\tilde{E}) = \Pi\left(\bigcap_{j=1}^{J} U_{(\epsilon,\tau_j)} \bigcap E_{(L,\epsilon^D)}\right) > 0$$

then f_0 is in the KL-support of Π and the posterior is weakly consistent at f_0.

Proof:

Let $P \in \tilde{E}$ and $P_0[0, L] = 1$ then similarly to before

$$\begin{aligned} K(f_0, f_P) &= \int_0^L f_0(x) \log \frac{f_0(x)}{f_P(x)} dx \\ &= \int_0^{\tau_0} f_0(x) \log \frac{f_0(x)}{f_P(x)} dx + \int_{\tau_0}^{\tau_{J+1}} f_0(x) \log \frac{f_0(x)}{f_P(x)} dx \\ &+ \int_{\tau_{J+1}}^{L} f_0(x) \log \frac{f_0(x)}{f_P(x)} dx \\ &= I_1 + I_2 + I_3. \end{aligned}$$

The first part I_1 is dealt in the same way as in the Proof of Theorem 1 and can be bounded by $\epsilon/3$ by choosing τ_0 small enough.

We now use the uniform continuity of f_0 on $[0, L]$. Let $\epsilon > 0$, set $\tau_J = \inf\{x, f_0(x) = \epsilon\}$, then there exist $\tau_0 < \tau_1 < ... < \tau_{J-1} < \tau_J$ such that $\forall j \leq J - 1$, $\forall \tau_j \leq x \leq \tau_{j+1}$, $|f_0(x) - f_0(\tau_{j+1}| \leq \epsilon^2$. Then for all P satisfying

$$|f_0(\tau_j) - f_P(\tau_j)| \leq \epsilon f_0(\tau_j), \quad j = 1, ..., \tau_J$$

$$
\begin{aligned}
I_2 &= \int_{\tau_1}^{\tau_{J+1}} f_0(x)\log\frac{f_0(x)}{f_P(x)}dx \\
&\leq \frac{1}{f_P(\tau_J)}\sum_{j=1}^{J}\int_{\tau_j}^{\tau_{j+1}} f_0(x)(f_0(x)-f_P(x))\mathcal{I}_A(x)dx \\
&\leq \frac{1}{f_P(\tau_J)}\sum_{j=1}^{J}(f_0(x)-f_0(\tau_{j+1})+f_0(\tau_{j+1})-f_P(\tau_{j+1}))\int_{\tau_j}^{\tau_{j+1}} f_0(x)dx \\
&\leq \frac{1}{f_P(\tau_J)}\sum_{j=1}^{J}[f_0(\tau_j)-f_0(\tau_{j+1})+f_0(\tau_{j+1})-f_P(\tau_{j+1})]\int_{\tau_j}^{\tau_{j+1}} f_0(x)dx \\
&\leq \frac{1}{f_P(\tau_J)}\sum_{j=1}^{J}(\epsilon+|f_0(\tau_{j+1})-f_P(\tau_{j+1})|)\int_{\tau_j}^{\tau_{j+1}} f_0(x)dx \\
&\leq \frac{\epsilon^2}{f_P(\tau_J)} \\
&\leq \frac{\epsilon^2}{f_0(\tau_J)(1-\epsilon)} \leq \epsilon(1-\epsilon)^{-1}.
\end{aligned}
$$

We now consider I_3. For all P satisfying $f_P(L) \geq \epsilon^D$ with $D > 0$

$$
\begin{aligned}
I_3 &= \int_{\tau_J}^{L} f_0(x)\log\left(\frac{f_0(x)}{f_P(x)}\right)dx \\
&\leq \int_{\tau_J}^{L} f_0(x)|\log f_0(x)|dx - D\log(\epsilon)\epsilon(L-\tau_J) \\
&\leq -(D+1)\log(\epsilon)\epsilon(L-\tau_J)
\end{aligned}
$$

Since as ϵ goes to zero, $\epsilon\log(\epsilon)$ goes to zero and since for all $\epsilon > 0$

$$\Pi(\{P; |f_P(\tau_j)-f_0(\tau_j)| \leq \epsilon, j = 1, ..., J; f_P(L) \geq \epsilon^D\}) > 0$$

for some $D > 0$, this achieves the Proof of Theorem 3

The following Theorem gives the conditions to obtain strong consistency of the posterior.

Theorem 3 *If* Π *be a prior on* $\mathcal{F}$ *and* $\Pi(K_\epsilon(f_0)) > 0$ *for all* $\epsilon > 0$. *If* $n \in \mathcal{N}, \quad \forall \epsilon > 0 \exists r > 0$ *such that* $\Pi\{P : f_P(0) > M_n\} \leq e^{-nr}$, *where* $M_n < e^{n\epsilon}$, *then the posterior distribution is strongly consistent at* f_0 *in the* L_1 *distance.*

Proof: The Proof is based on Barron *et all.* 's (1999) result.
We only need to contrôl the upper bracketing entropy of a sequence of sieves $(\mathcal{F}_n)_{n\in\mathcal{N}} \subset \mathcal{F}$ such that $\pi(\mathcal{F}_n^c) \leq e^{-nr}$ for some $r > 0$. Recall that the bracketing entropy is defined as the logarithm of the minimum of upper brackets f_j satisfying $\forall f_P \in \mathcal{F}_n, \exists j$ such that $f_P \leq f_j$ with $\int_0^1 f_j(x)d\mu(x) \leq 1 + \delta$.
Define $\mathcal{F}_n = \{P : f_P(0) \leq M_n\}$. By the Lemma 3.2 in the Appendix

$$\forall \epsilon > 0 \quad \mathcal{H}(\epsilon, \mathcal{F}_n) \leq \frac{c \log M_n}{\epsilon} \leq \frac{c \log e^{n\epsilon}}{\epsilon} = nc$$

and

$$\Pi(\mathcal{F}_n^c) \quad = \quad \Pi\{f_P : f_P(0) > M_n\} \leq e^{-nr}$$

So the posterior is strongly consistent.

2.1 Aplication to Dirichlet Priors

In this section we prove that a Dirichlet Prior on $\mathcal{M}$ satisfies the assumptions of Theorems 1 and 3. Assume that the prior Π is a Dirichlet prior with base measure α on $\mathcal{R}^+$ having positive density with respect to Lebesgue measure. We need to prove that there exists $D > 0$ small enough such that if

$$E(\tau_1, ..., \tau_J, \epsilon) = \left\{ P : \left| \int_{\tau_j}^{\infty} \frac{1}{\theta} d(P - P_0)_{(\theta)} \right| < \epsilon, j = 1, ..., J \right\} \cap \left\{ P : f_P(L) > \epsilon^D \right\}$$

then for all $\epsilon > 0$, $\Pi(E(\tau_1, ..., \tau_J, \delta)) > 0$ for all $0 < \tau_1 <, ..., \tau_J < \infty$ and $j = 1, ..., J$. Choose point $0 < a_1 < a_2 < ... < a_K < L$ such that

$$a_j^{-1} - a_{j+1}^{-1} < \epsilon/3, \quad j = 1, ..., K$$

, with $a_1 < \tau_1$ and $a_K > \tau_J$ and $P_0[a_K, L] + P_0[0, a_1] \leq \epsilon$. Set $w_j = (a_j, a_{j+1}]$ for $j = 1, ..., K$, then $\alpha(w_j) > 0$ and $\alpha(w_1 \bigcup w_2 \bigcup ... \bigcup w_{K+1}) < \alpha(\mathcal{R}^+)$,

$$(P(w_1), ..., P(w_K), 1 - \sum_{j=1}^{K} P(w_j)) \sim D(\alpha(w_1), ..., \alpha(w_{K+1}), \alpha(w_{K+1}))$$

where $w_K = [a_K, 2L]$. Since $0 < a_1 < a_2 < ... < a_K < L$, $P_0(w_j) > 0$ for all $j > 0$, let

$$P \in E'(\epsilon', \epsilon) = \{P \in \mathcal{M}; |P(w_j) - P_0(w_j)| \leq \epsilon' P_0(w_j)(\tau_j \wedge 1)/3, \forall j \leq K, P([L, 2L]) \geq \epsilon^D\},$$

then for all $j \leq J$,

$$\begin{aligned}
&\left|\int_{\tau_j}^{\infty} \frac{1}{\theta} d(P - P_0)\right| \leq \left|\int_{\tau_j}^{a_K} \frac{1}{\theta} d(P - P_0)\right| + \left|\int_{a_K}^{\infty} \frac{1}{\theta} d(P - P_0)(\theta)\right| \\
&\leq \sum_{l, a_l > \tau_j} |a_{l-1}^{-1} - a_l^{-1}|(P(w_{l-1}) + P_0(w_{l-1})| \\
&+\tau_j^{-1} \sum_{l, a_l > \tau_j} |P(w_l) - P_0(w_l)| \left|\int_{a_K}^{L} \frac{1}{\theta} d(P - P_0)(\theta)\right| + \int_L^{\infty} dP(\theta) \\
&\leq 2\epsilon/3 + \left|\int_{a_K}^{L} \frac{1}{\theta} d(P - P_0)(\theta)\right| + \int_L^{\infty} dP(\theta) \\
&\leq 2\epsilon/3 + a_K^{-1}[\epsilon + P[a_K, \infty[]
\end{aligned}$$

Since

$$P[a_K, \infty[= 1 - P[0, a_K[\leq 1 - P_0[a_0, a_K[+ \epsilon \leq P_0[0, a_0] + P_0[a_K, L] \leq \epsilon$$

the above term is bounded by $C\epsilon$, for some constant $C > 0$ independent of ϵ. We also have that if $P([L, 2L]) \geq \epsilon^D$ $f_P(L) \geq \epsilon^D/(2L)$ therefore $E'(\epsilon', \epsilon) \subset E(\tau_1, ..., \tau_J, \epsilon)$ for $\epsilon' = \epsilon/C$. We also have that E' contains

$$\{P; P_0(w_j)(1 - \epsilon/C) \leq P(w_j) \leq P_0(w_j)(1 - \epsilon/4C), j = 1, ..., K; P([L, 2L[) \geq \epsilon^2\},$$

for ϵ small enough, which has positive probability under a Dirichlet prior with base measure α having positive density. This proves that under a Dirichlet prior the posterior is weakly consistent as soon as there exists $L > 0$ such that $f_0(L) = 0$.

We now study the strong consistency. Assume that the base measure satisfies

$$\int_0^\infty \frac{\alpha(\theta)}{\theta} d\theta < \infty. \tag{2}$$

Now, let $M_n = e^{n\epsilon}$

$$\begin{aligned} \Pi\{f_P : f_P(x) \geq M_n\} &\leq \Pi\{f_P : f_P(0) \geq e^{n\epsilon}\} \\ &\leq e^{-n\epsilon}\mathcal{E}_\Pi[f_P(0)] \\ &= e^{-n\epsilon}\int_0^\infty \frac{1}{\theta}\alpha(\theta)d(\theta) \\ &\leq Ce^{-n\epsilon} \end{aligned}$$

for some $C > 0$ where the second inequality follows from Markov inequality and the last inequality follows from condition (2) on α.

3 Appendix

Lemma 1 *Let $\epsilon > 0$ and $\tau_0 > 0$ and f_0 , f_P be continuous and decreasing function on $\mathcal{R}^+$ such that, $f_0(\tau_0) > 0$ and assume there exists $\tau \in [0, \tau_0[$ such that $f_0(\tau) \geq f_0(\tau_0) + \epsilon$. Let*

$$U_{(\epsilon,\tau)} = \{P \in \mathcal{M} : |f_P(\tau) - f_0(\tau)| < \epsilon\}$$

then $\forall 0 < c < 1$ where $c \geq \frac{f_0(0)-f_0(\tau_0)}{f_0(0)}$ and $\forall P \in U_{(\epsilon,\tau)}$ and $x \leq \tau$

$$f_P(x) > (1 - c)f_0(x)$$

Proof: Let $\tau \in [0, \tau_0]$ then $f_0(\tau) \geq f_0(\tau_0) + \epsilon$, now if $P \in U_{(\epsilon,\tau)}$ then

$$f_P(\tau) > f_0(\tau) - \epsilon \geq f_0(\tau_0) = \frac{f_0(\tau_0)}{f_0(0)} f_0(0).$$

If consider $c = \frac{f_0(0)-f_0(\tau_0)}{f_0(0)}$ then $f_P(\tau) > (1-c)f_0(0)$, and hence for any $x \leq \tau$

$$f_P(x) > (1-c)f_0(0) > (1-c)f_0(x)$$

Theorem 4 *(schwartz(1965)) Let* Π *be a prior on* $\mathcal{M}$*. If* f_0 *is in the K-L support of* Π*, then the posterior is weakly consistent at* f_0*.*

Theorem 5 *Let* Π *be a prior on* L_μ*. Suppose* $f_0 \in L_\mu$ *is in the K-L-Support of* Π *and let* $U = \{f : \|f - f_i\| < \delta\}$*. If for each* $\epsilon > 0$*, there is a* $\delta < \frac{\epsilon}{4}, \quad c_1, \quad c_2 > 0, \quad \beta < \frac{\epsilon^2}{8}$ *and* $\mathcal{F}_n \subset L_\mu$ *such that, for all* n *large*
1. $\Pi(\mathcal{F}_n^c) < c_1 e^{-nc_2}$
2. $J(\delta, \mathcal{F}_n) < n\beta$
then the posterior is strongly consistent at f_0*.*
(S.GHOSAL,and all(1999)).

Lemma 2 *suppose* $\mathcal{F}_\mathcal{M} \subset L_\mu$ *is the family of decreasing densities* f *on* $[0,1]$ *such that* $f \leq M$*, and* $\Pi(\mathcal{F}_\mathcal{M}) < 1$ *for all* $M > 0$ *then exist a* $c > 0$ *such that* $J(\epsilon, \mathcal{F}_\mathcal{M}) < c\epsilon^{-1}$*.Groenboom P.(1986).*

References

[1] Balabdaoui, F. and Wellner,J.A.(2004d). Estimation of a k-monotone density ,part 4:limit distribution theory and the spline connection.Technical Report 462,University of Washington Department of statistics.

[2] Balabdaoui, F.(2006). Estimation non-paramtrique d'une densit k-monotone: Une nouvelle thorie de distribution asymptotique. PhD thesis.

[3] Barron,A.,Schervish,M.J. and Wasserman,L.(1999). The consistency of posterior distributions in nonparametric problems.ann.statist.27 536-561.

[4] Clair Alston, Sama Low Choy, Petra Kuhnert, Ross McVinish.(2005) Bayesian Model Comparison: Review and Discussion.International Statistical Institute 55th session.

[5] Diaconis, P. and Freedman, D(1986). on inconsistent bayes estimates of location .Ann. Statist., 14(1):6-!, 1986.

[6] Fergussen, T.S. (1983). Bayesian density estimation by mixtures of normal distributions. In Recent Advances in statistics(M. Razvi, J. Rustagi, and D. Siegmund, Eds.)287-302. Academic Press, New york.

[7] Ghosh,J.K. and Ramamroothi R.V.(2003). Bayesian Nonparametrics.Springer,New York.

[8] Grnenboom Piet (1986). Some current Development in Density Estimation.

[9] Lo,A.Y.(1984). On a class of Bayesian nonparametric estimates I: density estimates.Ann.statist. 1 38-58.

[10] Petrone, S. and Wasserman, L. (2002). Consistency of Bernstein polynomial posteriors. J.Roy. Statist. Soc., Ser. B 64 79-100

[11] Perron F. Mengersen K.(2001). Bayesian Nonparametric Modelling using Mixtures of Triangular Distributions, Biometrics, 57, 2, pp. 518-529.

[12] Robert, C. And Rousseau, J.(2003). Bayesian Goodness of Fit -A Mixture Approach to Bayesian Goodness of Fit

[13] Rousseau, J.(2007).Rates of convergence for the posterior distributions of mixture of Betas and adaptive nonparametric estimation of the densities. Valencia Proceedings, Bayesian Statistics 8, 417-452.

[14] Rousseau, J.(2008).(preprint)

[15] Schwartz, L.(1965). On Bayes Procedures.Z.Wahrsch.Verw.Gebiete 4 10-26

[16] Yuefeng Wu S. Ghosal.(2008). Kullback Leibler property of kernel mixture priors in Bayesian density estimation.Electronic Journal of Statistics .Vol. 2 298-331 ISSN 1935-7524

Exact likelihood equations for multivariate symmetric Kotz-type autoregressions

Z. Khodadadi, and B. Tarami

Islamic Azad University-Tehran Science & Research Branch, Yasoj university

Abstract: Family of the Kotz-type distribution have been found to be very useful in construction models when the usual normality assumption cannot be applicable for a data. The multivariate symmetric Kotz-type distribution provides a viable framework for modeling time-series data, it includes the multinormal and power exponential as special cases. For multivariate symmetric Kotz-type autoregressive models, we derive the exact likelihood equations for the model parameters, they are related to the Yule-Walker equations and involve simple function of the data. The maximum likelihood estimator are obtained by alternately solving two linear systems and illustrated using the simulation data.

Keywords:Causality; covariance function; innovation distribution; stationary process; Yule-Walker equations; Kotz-type distribution.

1 Introduction

Kotz-type distribution in the class of the elliptical distributions can be interpreted as being generated by the Weibull or Type III extreme value distribution. It includes two important distribution, multivariate normal and power exponential as special cases. Since 1990, there has been a surge of activity relating to the symmetric Kotz-type distribution. It has attracted applications in area such as Bayesian statistics, ecology, mathematical finance, shape theory and signal processing.

Consider a stationary stochastic processes $\{X_t\}$ defined by its autoregressive model

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + \varepsilon_t, \qquad t = 0, \pm 1, \pm 2, \ldots, \tag{1}$$

where $\varepsilon_t^{'}s$ are uncorrelated white noise with $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = \sigma^2$ ([2], [3], [14]). It is said to be weakly (strongly) causal if [3] its future innovations and past observations are uncorrelated (independent). In this case, the parameters $\sigma^2, \phi_1, \ldots, \phi_p$ and the autocovariance function $\{\gamma_k\}$ satisfy the well-known Yule-Walker equations

$$\begin{aligned} \gamma_0 &= \sum_{j=1}^{p} \phi_j \gamma_j + \sigma^2, \\ \gamma_i &= \sum_{j=1}^{p} \phi_j \gamma_{j-i}, \qquad p \geq i \geq 1. \end{aligned} \tag{2}$$

Its parameters can be estimated by replacing the $\gamma_k^{'}s$ by their estimators:

$$\hat{\gamma_k} = \frac{1}{n} \sum_{t=1}^{n-k} x_t x_{t+k}, \qquad k = 0, 1, \ldots, p.$$

Letting $\{\boldsymbol{\varepsilon}_t\}$ be identically independent distributed with a density function $f(\cdot)$, one may find the maximum likelihood estimators ([10], [14], [16], [17]) of the model parameters σ^2 and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^{'}$. The derivatives of the likelihood function with respect to the covariance parameters of $\mathbf{X} = (X_1, \ldots, X_n)^{'}$ is difficult to obtain, the MLE is usually found using numerical nonlinear optimization routines [3]. However Miller (1995) and Tarami and Pourahmadi (2003) have obtained the exact Likelihood equations for independent Gaussian and uncorrelated multivariate t-distribution innovations, respectively.

In this article, we extend their works to the uncorrelated Kotz-type distribution. Roughly speaking, we study AR processes whose observed sample $\mathbf{X} = (X_1, \ldots, X_n)^{'}$, $n > 1$ is distributed as

$$f(\mathbf{x}|\boldsymbol{\theta}) = \frac{s\Gamma(\frac{n}{2})}{\pi^{\frac{n}{2}}\Gamma(\frac{2N+n-2}{2s})} \quad a^{\frac{2N+n-2}{2s}} \; |\Sigma|^{-\frac{1}{2}} \; Q^{N-1} \; \exp\big[-aQ^s\big], \tag{3}$$

where $\boldsymbol{\theta} = (\sigma^2, \phi_1, \ldots, \phi_n)$ are the model parameters, Σ is the scaling matrix and $Q = \mathbf{x}'\Sigma^{-1}\mathbf{x}$. We obtain the simple exact likelihood equation

$$\frac{\partial Q}{\partial \boldsymbol{\theta}} = \Big[\frac{-2a^{-\frac{1}{s}}}{n} \quad \frac{\Gamma(\frac{2N+n}{2s})}{\Gamma(\frac{2N+n-2}{2s})}\big((N-1)Q^{-1} - asQ^{s-1}\big)\Big]^{-1} E_N(\frac{\partial Q}{\partial \boldsymbol{\theta}}). \tag{4}$$

This paper devoted in 4 sections. In section 2 presents theorem on exact likelihood equations of autoregressive models. A simulation method for solving the maximum likelihood equations of AR(2) is offered in the section 3. Finally in section 4 we prove the theorem and corollary.

2 Kotz-type distrinution

A random vector $\mathbf{X}$ is said to be distributed as a symmetric Kotz-type ([4], p.76) and write $\mathbf{X} \sim KTD_n(\boldsymbol{\mu}, \Sigma, N, a, s)$, if it has a density generator

$$g_n(u) = C_{n,N}\, u^{N-1} \exp(-au^s), \quad a, s > 0,\ 2N + n > 2, \tag{5}$$

where

$$C_{n,N} = \frac{s\Gamma(\frac{n}{2})}{\pi^{\frac{n}{2}}\Gamma(\frac{2N+n-2}{2s})} \quad a^{\frac{2N+n-2}{2s}},$$

and $\Gamma(\cdot)$ is the gamma function. thus, its density is given by

$$C_{n,N}|\Sigma|^{-\frac{1}{2}}\big[(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\big]^{N-1} \exp\big[-a[(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})]^s\big]. \tag{6}$$

Note that for $N = 1$, $s = 1$ and $a = \frac{1}{2}$, this reduces to the multivariate normal distribution and when $N = 1$, $a = \frac{1}{2}$ it is a multivariate power exponential distribution ([5], [8]).

If $\mathbf{X} \sim KTD_n(\boldsymbol{\mu}, \Sigma, N, a, s)$, then it has a stochastic representation ([4], p.9) as

$$\mathbf{X} \overset{d}{=} \boldsymbol{\mu} + R\, A'\, \mathbf{U}^{(k)} \tag{7}$$

where $\mathbf{U}^{(k)}$ is uniformly distributed on the unit sphere of $\mathbb{R}^k$ and is independent of $R > 0$, $\Sigma = A'\,A$ and $k = rank(\Sigma)$. The density of R is given by ([4], p. 35),

$$f_R(r) = \frac{2s}{\Gamma(\frac{2N+n-2}{2s})} \quad a^{\frac{2N+n-2}{2s}}\; r^{2N+n-3} \exp\left(-ar^{2s}\right), \quad r > 0, \qquad (8)$$

If $E(R^2) < \infty$, then

$$E(\mathbf{X}) = \boldsymbol{\mu}, \qquad Cov(\mathbf{X}) = \frac{E(R^2)}{rank(\Sigma)}\,\Sigma, \qquad (9)$$

in which, for a weakly causal stationary autoregressive model $rank(\Sigma) = n$ and R is defined in (2.3) and from (2.4)

$$E(R^2) = \frac{2s\; a^{\frac{2N+n-2}{2s}}}{\Gamma(\frac{2N+n-2}{2s})} \int_0^\infty r^{2N+n-1} \exp\left(-ar^{2s}\right) dr = a^{-\frac{1}{s}}\; \frac{\Gamma(\frac{2N+n}{2s})}{\Gamma(\frac{2N+n-2}{2s})}. \qquad (10)$$

Symmetric Kotz-type distribution retain most of the attract properties of the multivariate normal distribution. For example, if $\mathbf{X}$ and its parameters are partitioned as ([4], p. 45)

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

with dimensions n_1 and n_2, respectively, then

$$\mathbf{X}^{(i)} \sim KTD_{n_i}(\boldsymbol{\mu}^{(i)}, \Sigma_{ii}, N, a, s), \quad i = 1, 2.$$

If B is some $k \times n$ matrix of rank $k \leq n$ and $\mathbf{b}$ is in $\mathbb{R}^k$ then

$$B\mathbf{X} + \mathbf{b} \sim KTD_k(A\boldsymbol{\mu} + \mathbf{b},\; A\Sigma A', N, a, s).$$

3 Exact maximum likelihood equations

In this section we deal with centered data and focus solely on exact maximum likelihood equations of AR(p) models (1.10). The following theorem is the main part of this article.

Theorem 6 *Let* $\mathbf{x} = (x_1, \ldots, x_n)'$ *be n observations from a stationary AR(p) model with Kotz -type distribution, then*

(I) *The exact likelihood equation for* $\boldsymbol{\theta}$ *is*

$$\frac{\partial Q}{\partial \boldsymbol{\theta}} = \left[-2l_N \left((N-1)Q^{-1} - asQ^{s-1}\right)\right]^{-1} E_N\left(\frac{\partial Q}{\partial \boldsymbol{\theta}}\right), \tag{11}$$

where

$$l_N = \frac{a^{-\frac{1}{s}}}{n} \quad \frac{\Gamma\left(\frac{2N+n}{2s}\right)}{\Gamma\left(\frac{2N+n-2}{2s}\right)}. \tag{12}$$

(II) *For a weakly causal stationary autoregressive model with symmetric Kotz-type distribution, its parameters and the autocovariances satisfy*

$$D_{i0} = \sum_{j=1}^{p} \phi_j \left[D_{ij} + \frac{1}{nl_N}\left(\frac{2(N-1)+n}{2as}\right)^{\frac{1}{s}} \; j \; \gamma_{j-i}\right],$$

$$D_{00} = \sum_{j=1}^{p} \phi_j \left[D_{j0} + j\gamma_j \; \frac{1}{nl_N}\left(\frac{2(N-1)+n}{2as}\right)^{\frac{1}{s}}\right] \tag{13}$$

$$+ \sigma^2 l_N^{-1}\left(\frac{2(N-1)+n}{2as}\right)^{\frac{1}{s}}. \tag{14}$$

The proof is postponed to section 5.

Corollary 1. For a weakly causal stationary symmetric multivariate power exponential distribution autoregressive model, its parameters

and the autocovariances satisfy

$$D_{i0} = \sum_{j=1}^{p} \phi_j \Big[D_{ij} + \frac{1}{n\, l_1} \ \Big(\frac{n}{s}\Big)^{\frac{1}{s}} \ j\ \gamma_{j-i} \Big],$$

$$D_{00} = \sum_{j=1}^{p} \phi_j \Big[D_{j0} + j\gamma_j \ \frac{1}{n\, l_1} \ \Big(\frac{n}{s}\Big)^{\frac{1}{s}} \Big] + \Big(\frac{n}{s}\Big)^{\frac{1}{s}} \ l_1^{-1}\ \sigma^2,$$

these equations with s=1 reduces to the likelihood equations of Gaussian model,

$$D_{i0} = \sum_{j=1}^{p} \phi_j \big[D_{ij} + j\ \gamma_{j-i} \big],$$

$$D_{00} = \sum_{j=1}^{p} \phi_j \big[D_{j0} + j\gamma_j \big] + n\ \sigma^2,$$

which have been obtained by Miller(1995).

Since this paper has extended the previous works to more general distribution , we follow their procedure to find the maximum likelihood estimators.

Step 1: compute the $(p+1) \times (p+1)$ matrix $\mathbf{D} = (D_{ij})_{i,j=0}^{p}$.

Step 2: Using $\mathbf{D}$ and the current estimates of $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_p)'$, solve (3.3)-(3.4) for $\boldsymbol{\theta} = (\sigma^2, \phi_1, \ldots, \phi_p)'$.

Step 3: Use the Yule-Walker equations to obtain updated estimates of $\boldsymbol{\gamma}$ given the current estimates of $\boldsymbol{\theta}$.

Step 4: Iterate steps 2-3 until convergence.

4 Simulation

In this section, we provide a method to generate autoregressive time series data with Kotz Type distribution. For generating random variable with a symmetric Kotz-type distribution, first we generate random variable with density $f_R(r)$ and from (2.3) obtain the required numbers. From ([4], p. 76) the density of $U = R^2$ is given

$$f_U(u) = \frac{s\ a^{\frac{2N+n-2}{2s}}}{\Gamma(\frac{2N+n-2}{2s})}\ u^{N+\frac{2N+n-2}{2}}\ \exp(-au^s), \qquad a > 0.$$

Set $T = U^s$, then the density of T is

$$f_T(t) = \frac{a^{\frac{2N+n-2}{2s}}}{\Gamma(\frac{2N+n-2}{2s})}\ t^{\frac{2N+n-2}{2s}-1}\ \exp(-at), \qquad t > 0,$$

which is distributed as gamma $(\frac{2N+n-2}{2s}, a)$. Since the parameters of gamma distribution must be positive, the conditions of $2N + n > 2$, $a > 0$ are essential. Now, by generating random numbers from $f_T(t)$ and using $R = \sqrt[2s]{T}$, we can generate from $f_R(r)$. Next, we use the above method to generate $n = 150$ uncorrelated innovations from the Kotz-type distribution i.e,

$$\boldsymbol{\varepsilon} \sim KTD_{150}(\mathbf{0}, \Sigma, g_{150}),$$

where

$$Cov(\boldsymbol{\varepsilon}) = 0.05\ I_{150} = \frac{a^{-\frac{1}{s}}}{150}\ \frac{\Gamma(\frac{2N+150}{2s})}{\Gamma(\frac{2N+148}{2s})}\ \Sigma.$$

These innovations are then used to simulate from the AR(2) model

$$X_t = 1.03 + 1.30X_{t-1} - 0.72X_{t-2} + \varepsilon_t.$$

To highlight the role of the parameters N and s of the multivariate Kotz-type distribution, we have solved (3.3)-(3.4) for several values of (N, a, s), by using iterative algorithm (step 1-4). The estimated values of ϕ_1, ϕ_2, σ_s^2 and σ_N^2 are reported in Table I. In which σ_s^2 and σ_N^2 are values of σ^2 for given s and N. It is evident that N, s

have very little impact on the estimates of ϕ_1, ϕ_2, σ_s^2 and σ_N^2. Clearly, larger value of N and s corresponds to more heavy-tailed distributions and hence to larger innovations variances.

TABLE I
MLE FOR THE AR(2) MODELS FITTED TO THE SIMULATED DATA UNDER KOTZ-TYPE DISTRIBUTION

(N, a)=(1, 0.5)			
s	ϕ_1	ϕ_2	σ_s^2
1	1.26356	-0.66588	0.04828
5	1.26351	-0.66583	0.04847
15	1.35132	-0.81263	0.04953
25	1.33829	-0.80111	0.05265
(a, s)=(0.5, 1)			
N	ϕ_1	ϕ_2	σ_N^2
1	1.2635645	-0.6658808	0.04828
5	1.2635649	-0.6658810	0.04833
20	1.2635643	-0.6658806	0.04846
60	1.2635649	-0.6658811	0.04867
95	1.2635646	-0.6658808	0.04878

5 Proof

Proof of Theorem 1.
(I) Since the log-likelihood function for $\boldsymbol{\theta} = (\sigma^2, \phi_1, \ldots, \phi_p)$ is

$$l(\boldsymbol{\theta}|\mathbf{x}) = -\frac{1}{2}\ln|\Sigma| + \ln C_{n,N} + (N-1)\ln Q - aQ^s. \tag{15}$$

Differentiating (5.1) with respect to $\boldsymbol{\theta}$ gives

$$\frac{\partial}{\partial\boldsymbol{\theta}} l(\boldsymbol{\theta}|x) = -\frac{1}{2}\frac{\partial}{\partial\boldsymbol{\theta}}\ln|\Sigma| + \left[(N-1)Q^{-1} - asQ^{s-1}\right]\frac{\partial Q}{\partial\boldsymbol{\theta}}. \tag{16}$$

But (1.3) is a density function and its integral over the support of $\mathbf{X}$ is equal to one. Thus

$$|\Sigma|^{-\frac{1}{2}} = \Big[\int C_{n,N} Q^{N-1}\exp\left(-aQ^s\right)d\mathbf{x}\Big]^{-1}.$$

Taking derivatives of logarithms of two sides with respect to $\boldsymbol{\theta}$ gives

$$-\frac{1}{2}\frac{\partial}{\partial\boldsymbol{\theta}}\ln|\Sigma| = E\big[(N-1)Q^{-1}\ \frac{\partial Q}{\partial\boldsymbol{\theta}} - asQ^{s-1}\ \frac{\partial Q}{\partial\boldsymbol{\theta}}\big]. \qquad (17)$$

Substituting from (5.3) into (5.2) gives

$$E\big[(N-1)Q^{-1}\frac{\partial Q}{\partial\boldsymbol{\theta}} - asQ^{s-1}\frac{\partial Q}{\partial\boldsymbol{\theta}}\big] = (N-1)Q^{-1}\frac{\partial Q}{\partial\boldsymbol{\theta}} - asQ^{s-1}\frac{\partial Q}{\partial\boldsymbol{\theta}}. \qquad (18)$$

With assumption $2(N-1)+n>2$, $2(N+s-1)+n>2$

$$\begin{aligned}
&(N-1)\ E_N\big[Q^{-1}\frac{\partial Q}{\partial\boldsymbol{\theta}}\big] - asE_N\big[Q^{s-1}\frac{\partial Q}{\partial\boldsymbol{\theta}}\big] = \\
&(N-1)\int \frac{\partial Q}{\partial\boldsymbol{\theta}}\ C_{n,N}|\Sigma|^{-\frac{1}{2}}\ Q^{(N-1)-1}\ \exp(-aQ^s)d\mathbf{x} \\
&-as\int \frac{\partial Q}{\partial\boldsymbol{\theta}}\ C_{n,N}|\Sigma|^{-\frac{1}{2}}\ Q^{(N+s-1)-1}\ \exp(-aQ^s)d\mathbf{x} \\
&= k_1\ E_{N-1}\big[\frac{\partial Q}{\partial\boldsymbol{\theta}}\big] - k_2E_{N+s-1}\big[\frac{\partial Q}{\partial\boldsymbol{\theta}}\big],
\end{aligned}$$

where

$$E_{N^*}(\frac{\partial Q}{\partial\boldsymbol{\theta}}) = \int \frac{\partial Q}{\partial\boldsymbol{\theta}}\ f_{\mathbf{X}}(\mathbf{x};N^*,a,s)\ d\mathbf{x},$$

$f_{\mathbf{X}}(\mathbf{x};N^*,a,s)$ defined in (1.3), and

$$\begin{aligned}
k_1 &= (N-1)a^{\frac{1}{s}}\ \frac{\Gamma(\frac{2N+n-4}{2s})}{\Gamma(\frac{2N+n-2}{2s})}, \\
k_2 &= \frac{2N+n-4}{2}\ \frac{k_1}{N-1}.
\end{aligned}$$

Now from (2.5) and (2.6), for a weakly causal stationary symmetric Kotz-type distribution autoregressive model

$$Cov(\mathbf{X}) = a^{-\frac{1}{s}}\ \frac{\Gamma(\frac{2N+n}{2s})}{\Gamma(\frac{2N+n-2}{2s})}\ \frac{\Sigma}{n} = l_N\Sigma, \qquad (19)$$

hence

$$E_m(X_iX_j) = l_m\ (\Sigma)_{ij},$$

where Σ_{ij} is the (ij)-th element of Σ, then

$$E_m(X_iX_j) = \frac{l_m}{l_{m'}}\ E_{m'}(X_iX_j).$$

Therefore

$$\begin{aligned} E\big[(N-1)Q^{-1}\frac{\partial Q}{\partial \boldsymbol{\theta}} - asQ^{s-1}\frac{\partial Q}{\partial \boldsymbol{\theta}}\big] &= \big[k_1\ \frac{l_{N-1}}{l_N} - k_2\ \frac{l_{N+s-1}}{l_N}\big]\ E_N\big[\frac{\partial Q}{\partial \boldsymbol{\theta}}\big] \\ &= -\frac{na^{\frac{1}{s}}}{2}\ \frac{\Gamma(\frac{2N+n-2}{2s})}{\Gamma(\frac{2N+n}{2s})}\ E_N\big[\frac{\partial Q}{\partial \boldsymbol{\theta}}\big], \end{aligned} \tag{20}$$

and from (5.4) and (5.6), we get to the required result(3.1).

(II) The entries of Σ^{-1} for autoregressive models are known to be a quadratic function of its coefficients ([2], p. 198), hence

$$Q = \mathbf{x}'\Sigma^{-1}\mathbf{x} = \frac{E(R^2)}{n\sigma^2} \cdot \tilde{\boldsymbol{\phi}}'\,\mathbf{D}\ \tilde{\boldsymbol{\phi}}, \tag{21}$$

where R defined in (2.3) and $\tilde{\boldsymbol{\phi}} = (-1, \phi_1, \ldots, \phi_p)'$, $\mathbf{D}$ is a $(p+1) \times (p+1)$ symmetric matrix with entries

$$D_{ij} = \sum_{k=i+1}^{n-j} x_k x_{k+j-i}, \quad i, j = 0, 1, \ldots, p,$$

set $S(\mathbf{x}, \boldsymbol{\phi}) = \tilde{\boldsymbol{\phi}}'\, D\tilde{\boldsymbol{\phi}}$, then

$$S(\mathbf{x}, \boldsymbol{\phi}) = D_{00} - 2\sum_{i=1}^{p} \phi_i D_{i0} + \sum_{i=1}^{p}\sum_{j=1}^{p} \phi_i\phi_j D_{ij}, \tag{22}$$

and from (5.7), (5.8) and (2.6) we obtain

$$\frac{\partial Q}{\partial \phi_i} = \frac{-2\ a^{-\frac{1}{s}}}{n\sigma^2}\ \frac{\Gamma(\frac{2N+n}{2s})}{\Gamma(\frac{2N+n-2}{2s})}\ \big[D_{i0} - \sum_{j=1}^{p} \phi_j D_{ij}\big], \quad i = 1, \ldots, p. \tag{23}$$

Since

$$E(D_{ij}) = (n - j - i)\gamma_{j-i}, \tag{24}$$

it follows from (5.9), (5.10) and Yule-Walker equations (1.2) that

$$E\big[\frac{\partial Q}{\partial \phi_i}\big] = \frac{-2\ a^{-\frac{1}{s}}}{n\sigma^2}\ \frac{\Gamma(\frac{2N+n}{2s})}{\Gamma(\frac{2N+n-2}{2s})}\ \sum_{j=1}^{p} j\phi_j\gamma_{j-i}. \tag{25}$$

Similarly, from (5.7) and (2.6)

$$\frac{\partial Q}{\partial \sigma^2} = \frac{-a^{-\frac{1}{s}}}{n\sigma^4}\ \frac{\Gamma(\frac{2N+n}{2s})}{\Gamma(\frac{2N+n-2}{2s})}\ S(\mathbf{x}, \boldsymbol{\phi}), \tag{26}$$

then using (1.2) and (5.12) we have

$$E\big[\frac{\partial Q}{\partial \sigma^2}\big] = \frac{-a^{-\frac{1}{s}}}{\sigma^2}\ \frac{\Gamma(\frac{2N+n}{2s})}{\Gamma(\frac{2N+n-2}{2s})}. \tag{27}$$

From (5.7) and (2.6)

$$Q = \sigma^{-2}\ l_N\ S(\mathbf{x}, \boldsymbol{\phi}), \tag{28}$$

and from (5.12), (5.13), (3.1) for σ^2 as the first entry of $\boldsymbol{\theta}$

$$S(\mathbf{x}, \boldsymbol{\phi}) = \big[-2l_N\big((N-1)Q^{-1} - asQ^{s-1}\big)\big]^{-1}\ n\sigma^2, \tag{29}$$

and from (5.14), (5.15)

$$Q = \big(\frac{2(N-1)+n}{2as}\big)^{\frac{1}{s}}, \tag{30}$$

and now from (3.1), (5.14) and (5.16) for ϕ_i as the second entry of $\boldsymbol{\theta}$ we have

$$\frac{\partial S(\mathbf{x},\boldsymbol{\phi})}{\partial \phi_i} = \frac{1}{nL_N}\big(\frac{2(N-1)+n}{2as}\big)^{\frac{1}{s}} E_N\big(\frac{\partial S(\mathbf{x},\boldsymbol{\phi})}{\partial \phi_i}\big). \qquad (31)$$

Now, from (5.9) and (5.11) gives (3.3).

By using Yule-Walker equations(1.2), (5.8), (5.16) and (3.3)

$$\begin{aligned}\sigma^2\big(\frac{2(N-1)+n}{2as}\big)^{\frac{1}{s}} l_N^{-1} &= D_{00} - \sum_{j=1}^{p}\phi_j D_{j0} \\ &\quad - \sum_{j=1}^{p}\frac{1}{nl_N}\big(\frac{2(N-1)+n}{2as}\big)^{\frac{1}{s}} \cdot j \cdot \phi_j\, \gamma_j\end{aligned}$$

which gives (3.4).

Proof of Corollary 1.
In (3.3) and (3.4), take $N = 1$ and $a = \frac{1}{2}$.

References

[1] BOX, E. P. G., JENKINS, G. M. and REINSEL, G. C. (1994) *Time Series Analysis*, (3rd edn). New Jersy: Prentice Hall.

[2] BROCKWELL, P. J. and DAVIS, R. A. (1991) *Time Series and Methods* , (2nd edn). New York: Spring- Verlag.

[3] FANG, K. T., KOTZ, S., NG, K. W. (1990) *Symmetric Multivariate and Related Distributions*, London, New York: Chapman & Hall.

[4] GOMEZ, E., GOMEZ-VILLEGAS, M.A., MARIN, J.M. (1998) A multivariate generalization of the power exponential family of distributions. *Commun. Statist. Theory Method* 27, 589-600.

[5] GUPTA, A. K., VARGA, T., (1993) *Elliptically Contoured Models in Statistics*, Dordrecht: Kluwer Academic Publishers.

[6] JOHNSON, N. L. and KOTZ, S. (1972) *Distribution in Statistics: Continuous Multivariate Distributions*. New York: John Wily & Sons, Inc.

[7] KANO, Y.(1994) Consistency property of elliptical probability density function. *J. Multiva. Anal* 51, 139-147.

[8] MILLER, J. W. (1995) Exact maximum likelihood estimation in autoregressive processes. *Jornal of Time Series Analysis* 16(6), 607-615.

[9] MUIRHEAD, R. J. (1982) *Aspects of Multivariate Statistical Theory*. New York: John Wily & Sons, Inc.

[10] PETER, C.B. PHILIPS, KE-LIXU (2006) Inference in autoregression under heteroskedasticity.*Jornal of Time Series Analysis* 27(2), 289-308.

[11] TARAMI, B., POURAHMADI, M. (2003) Multi-variate t autoregression. innovations, prediction variance and exact likelihood equations. *Jornal of Time Series Analysis* 24(6), 739-754.

[12] TIKU, M. L., WONG, W. K., VAUGHAN, D. C. and BIAN, G. (2000) Time series modeles in non-normal situations: symmetric innovations. *Jornal of Time Series Analysis* 21, 571-596.

[13] YAO, Q., BROCKWELL, P.J. (2006) Gaussian maximum likelihood estimation for ARMA models.*Jornal of Time Series Analysis* 27(6), 857-875(19).

[14] ZHOU, J., BASAWA, I.V. (2005) Maximum likelihood estimation for a first order bifurcating autoregressive process with exponential erroes.*Jornal of Time Series Analysis* 26(6), 825-842.

Recurrence relations for distributions of skew-t and a linear combination of order statistics from bivariate-t

M. Khosravi, A. Jamalizadeh

Department of Statistics, Shahid Bahonar University of Kerman

Abstract: In this paper, we derive recurrence relations for the cdf of a skew-t distribution and also the distribution of a linear combination of order statistics arising from a bivariate-t distribution. These recurrence relation are over ν, the degrees of freedom, and starting from the results for $\nu = 1$ and 2, they will allow for the recursive evaluation of the distribution function from any other value of ν.

Keywords: Skew-normal distribution, skew-t distribution, order statistics, bivariate-t distribution, recursive evaluation.

1 Introduction

A random variable Z_λ is said to have a *skew-normal distribution* with parameter $\lambda \in R$, denoted by $Z_\lambda \sim SN(\lambda)$ if its pdf is [see Azzalini(1985)]

$$\varphi(z;\lambda) = 2\varphi(z)\Phi(\lambda z), \qquad z \in R, \tag{1}$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and cdf, respectively.This distribution has been studied quite extensively in the literature by several authors including Azzalini (1986), Henze (1986), Azzalini and Dalla (1996), and Arnold and Beaver (2002). A comprehensive survey of various developments on this skew-normal distribution and its multivariate form is due to Azzalini (2005).

Azzalini and Capitanio (2003) presented a multivariate skew-t distribution through a multivariate skew-normal distribution. In this work, we consider the univariate case of these multivariate skew-t distributions. Specifically, we say that a random variable $W_{\nu,\lambda}$ has a

skew-t distribution with parameters $\nu > 0$ (degrees of freedom) and $\lambda \in R$, denoted by $W_{\nu,\lambda} \sim St(\nu, \lambda)$, if

$$W_{\nu,\lambda} \stackrel{d}{=} V^{-\frac{1}{2}} Z_\lambda, \tag{2}$$

where $Z_\lambda \sim SN(\lambda)$ independently of $V \sim \frac{X^2_{(\nu)}}{\nu}$. Let us denote the pdf and the cdf of $W_{\nu,\lambda} \sim St(\nu,\lambda)$ by $g(t;\nu,\lambda)$ and $G(t;\nu,\lambda)$,respectively; further, in the special when the shape parameter $\lambda = 0$, we shall denote them by $g(t;\nu)$ and $G(t;\nu)$, which incidentally are the pdf and cdf of the usual Student-t distribution with ν degrees of freedom.

For the general skew-t distribution, it is known from Azzalini and Capitanio (2003) that

$$g(t;\nu,\lambda) = 2g(t;\nu)G\left(\lambda t\sqrt{\frac{1+\nu}{t^2+\nu}};\nu+1\right), \quad t,\ \lambda \in R. \tag{3}$$

Recently, Behboodian et al. (2006) discussed some properties of this distribution in the special case when $\nu = 1$ termed as the *skew-cauchy distribution*, and specifically derived explicit expressions for its pdf,cdf, and quantiles.

In section 2, we first present an integration formula for the cdf of $W_{\nu,\lambda} \sim St(\nu,\lambda)$ and then use it to derive a recurrence relation for the cdf $G(t;\nu,\lambda)$. Next, in section 3, we consider order statistics arising from a bivariate-t distribution and show that the cdf of a linear combination of these order statistics is a mixture of skew-t distributions, and then use this fact to present a recurrence relation for the cdf of this linear combination of order statistics.These recurrence relations are over ν, the degrees of freedom, and starting from the results for $\nu = 1$ and 2, they will allow for the recursive evaluation of the cdf for any other value of ν.

2 Recurrence relation for Cdf of Skew-t

From Eq. (2) we readily have integration form for the cdf $G(t;\nu,\lambda)$ as

$$G(t;\nu,\lambda) = P(V^{-\frac{1}{2}} Z_\lambda \le t) = P(Z_\lambda \le tV^{\frac{1}{2}}) = E\left[\Phi(tV^{\frac{1}{2}};\lambda)\right], \tag{4}$$

where $\Phi(\cdot,\lambda)$ denotes the cdf of $SN(\lambda)$ and $V \sim \frac{X^2_{(\nu)}}{\nu}$, as before. Now upon substituting the density function of $V = \frac{X^2_{(\nu)}}{\nu}$ in Eq. (4), we obtain integrtion formula for the cdf $G(t;\nu,\lambda)$ as

$$G(t;\nu,\lambda) = \frac{2(\frac{\nu}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} \int_0^\infty x^{\nu-1} e^{-\frac{\nu}{2}x^2} \Phi(tx;\lambda) dx. \tag{5}$$

Theorem 1 *If $\nu > \frac{1}{2}$ and $t \in R$, then*

$$\begin{aligned} G(t;2\nu+1,\lambda) &= G\left(\sqrt{\frac{2\nu-1}{2\nu+1}}t;2\nu-1,\lambda\right) + \frac{\Gamma(\nu)(2\nu+1)^{\nu-\frac{1}{2}}}{\sqrt{\pi}\Gamma(\frac{2\nu+1}{2})} \\ &\quad \times \frac{t}{(2\nu+1+t^2)^\nu} G\left(\frac{\sqrt{2\nu}\lambda t}{\sqrt{2\nu+1+t^2}};2\nu\right), \end{aligned} \tag{6}$$

where $G(t;\nu)$ is the cdf of $W_\nu \sim t_\nu$.

Proof : From the integration formula in Eq. (5) we first of all have

$$G(t;2\nu+1,\lambda) = \frac{2(\frac{2\nu+1}{2})^{\frac{2\nu+1}{2}}}{\Gamma(\frac{2\nu+1}{2})} \int_0^\infty x^{2\nu} e^{-\frac{2\nu+1}{2}x^2} \Phi(tx;\lambda) dx.$$

Upon integrating by parts now, we obtain

$$\begin{aligned} G(t;2\nu+1,\lambda) &= G\left(\sqrt{\frac{2\nu-1}{2\nu+1}}t;2\nu-1,\lambda\right) \\ &\quad + \frac{\left(\frac{2\nu+1}{2}\right)^{\nu-\frac{1}{2}}}{\Gamma(\frac{2\nu+1}{2})} \int_0^\infty x^{2\nu-1} e^{-\frac{2\nu+1}{2}x^2} \left\{\frac{\partial}{\partial x}\Phi(tx;\lambda)\right\} dx \\ &= G\left(\sqrt{\frac{2\nu-1}{2\nu+1}}t;2\nu-1,\lambda\right) \\ &\quad + \frac{\left(\frac{2\nu+1}{2}\right)^{\nu-\frac{1}{2}}}{\Gamma\left(\frac{2\nu+1}{2}\right)} t \int_0^\infty x^{2\nu-1} e^{-\frac{2\nu+1}{2}x^2} \varphi(tx;\lambda) dx. \end{aligned}$$

By using the fact that $\varphi(tx;\lambda) = 2\Phi(\lambda tx)\varphi(tx)$ and changing variables, we get

$$\begin{aligned} G\left(t; 2\nu+1, \lambda\right) &= G\left(\sqrt{\frac{2\nu-1}{2\nu+1}}t; 2\nu-1, \lambda\right) + \frac{\Gamma(\nu)(2\nu+1)^{\nu-\frac{1}{2}}}{\sqrt{\pi}\Gamma(\frac{2\nu+1}{2})} \\ &\times \frac{t}{(2\nu+1+t^2)^{\nu}} \frac{2\nu^{\nu}}{\Gamma(\nu)} \int_0^{\infty} x^{2\nu-1} e^{-\nu x^2} \Phi\left(\frac{\sqrt{2\nu}\lambda t x}{\sqrt{2\nu+1+t^2}}\right) dx \\ &= G\left(\sqrt{\frac{2\nu-1}{2\nu+1}}t; 2\nu-1, \lambda\right) + \frac{\Gamma(\nu)(2\nu+1)^{\nu-\frac{1}{2}}}{\sqrt{\pi}\Gamma(\frac{2\nu+1}{2})} \\ &\times \frac{t}{(2\nu+1+t^2)^{\nu}} G\left(\frac{\sqrt{2\nu}\lambda t}{\sqrt{2\nu+1+t^2}}; 2\nu\right), \end{aligned}$$

where the last equality follows from Eq. (5) with $\lambda = 0$.

Remark 1 It is important to mention that the recurrence formula in Eq. (6) is valid for any real $\nu > \frac{1}{2}$, i.e., for integral and non-integral values of ν.

Remark 2 In the special case when $\lambda = 0$, Theorem 1 yields a recurrence relation for the cdf of the usual student's-t_ν distribution as

$$\begin{aligned} G(t; 2\nu+1) &= G\left(\sqrt{\frac{2\nu-1}{2\nu+1}}t; 2\nu-1\right) \\ &+ \frac{\Gamma(\nu)(2\nu+1)^{\nu-\frac{1}{2}}}{2\sqrt{\pi}\Gamma(\frac{2\nu+1}{2})} \times \frac{t}{(2\nu+1+t^2)^{\nu}} \end{aligned} \qquad (7)$$

for $t \in R$ and $\nu > \frac{1}{2}$.

The relation in (7) holds for integral as well as non-integral values of ν (see Remark 1 above). Now, by using the known expressions of

$$G(t;1) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(t) \quad \text{and} \quad G(t;2) = \frac{1}{2}\left(1 + \frac{t}{\sqrt{2+t^2}}\right), \qquad (8)$$

the relation in (7) can be recursively used to produce expression for other integer values of ν; for example, we obtain

$$G(t;3) = G\left(\frac{t}{\sqrt{3}};1\right) + \frac{\sqrt{3}t}{\pi(3+t^2)} = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{t}{\sqrt{3}}\right) + \frac{\sqrt{3}t}{\pi(3+t^2)}, \quad (9)$$

$$G(t;4) = \frac{1}{2} + \frac{t^3+6t}{2(4+t^2)^{\frac{3}{2}}}. \quad (10)$$

As a matter of fact, the recurrence relation in (7) can be used to produce the expressions

$$G(t;\nu) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(\frac{t}{\sqrt{\nu}}) + \frac{1}{2\sqrt{\pi}}\sum_{i=1}^{\frac{\nu-1}{2}} \frac{\Gamma(i)\ \nu^{i-\frac{1}{2}}}{\Gamma(i+\frac{1}{2})}\frac{t}{(\nu+t^2)^i}$$

for odd values of ν, and

$$G(t;\nu) = \frac{1}{2} + \frac{1}{2\sqrt{\pi}}\sum_{i=1}^{\frac{\nu}{2}} \frac{\Gamma(i-\frac{1}{2})\ \nu^{i-1}}{\Gamma(i)}\frac{t}{(\nu+t^2)^{i-\frac{1}{2}}}$$

for even values of ν; see Janson, Kotz and Balakrishnan (1995) for general results of this form.

For the skew-Cauchy distribution, Behboodian et al. (2006) derived an explicit expression for the cdf $G(t;1,\lambda)$ as

$$G(t;1,\lambda) = \frac{1}{\pi}\left\{\tan^{-1}(t) + \cos^{-1}\left(\frac{\lambda}{\sqrt{(1+\lambda^2)(1+t^2)}}\right)\right\}. \quad (11)$$

Also, from Eq.(5), upon carrying out the integration, we can derive an explicit expression for the cdf $G(t;2,\lambda)$ as

$$G(t;2,\lambda) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}(\lambda) + \frac{t}{\sqrt{2+t^2}}\left\{\frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{\lambda t}{\sqrt{2+t^2}}\right)\right\}, \quad (12)$$

Now, upon setting $\nu = 1, \frac{3}{2}, ...$ in the recurrence relation in (6), and using the expressions of $G(t;1,\lambda), G(t;2,\lambda)$ in (11) and (12), respectively,and the expressions of $G(t;1)$, $G(t;2)$, $G(t;3)$ and $G(t;4)$ presented in Eqs. (8) - (10) we can derive explicit expressions for $G(t;3,\lambda), G(t;4,\lambda)$, and so on, in a simple recursive manner. For example, we obtain in this manner

$$
\begin{aligned}
&G(t;3,\lambda) = G\left(\frac{t}{\sqrt{3}};1,\lambda\right) + \frac{2\sqrt{3}t}{\pi(3+t^2)} G\left(\frac{\sqrt{2}\lambda t}{\sqrt{3+t^2}};2\right) \\
&= \frac{1}{\pi}\left\{\tan^{-1}(\frac{t}{\sqrt{3}}) + \cos^{-1}\left(\frac{\sqrt{3}\lambda}{\sqrt{(1+\lambda^2)(3+t^2)}}\right)\right\} \\
&+ \frac{\sqrt{3}t}{\pi(3+t^2)}\left(1 + \frac{\lambda t}{\sqrt{3+(1+\lambda^2)t^2}}\right) \\
&= \frac{1}{\pi}\left\{\tan^{-1}(\frac{t}{\sqrt{3}}) + \cos^{-1}\left(\frac{\sqrt{3}\lambda}{\sqrt{(1+\lambda^2)(3+t^2)}}\right) + \frac{\sqrt{3}t}{(3+t^2)}\right\} \\
&+ \frac{1}{\pi}\left\{\frac{\sqrt{3}\lambda t^2}{(3+t^2)\sqrt{3+(1+\lambda^2)t^2}}\right\},
\end{aligned}
$$

and

$$
\begin{aligned}
&G(t;4,\lambda) = G\left(\frac{t}{\sqrt{2}};2,\lambda\right) + \frac{2t}{(4+t^2)^{\frac{3}{2}}} G\left(\frac{\sqrt{3}\lambda t}{\sqrt{4+t^2}};3\right) \\
&= \frac{1}{2} - \frac{1}{\pi}\tan^{-1}(\lambda) + \frac{t}{\sqrt{4+t^2}}\left\{\frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{\lambda t}{\sqrt{4+t^2}}\right)\right\} \\
&+ \frac{2t}{(4+t^2)^{\frac{3}{2}}}\left\{\frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{\lambda t}{\sqrt{4+t^2}}\right) + \frac{\lambda t\sqrt{4+t^2}}{\pi[4+(1+\lambda^2)t^2]}\right\} \\
&= \frac{1}{2} - \frac{1}{\pi}\tan^{-1}(\lambda) + \frac{t}{\sqrt{4+t^2}}\left(\frac{6+t^2}{4+t^2}\right)\left\{\frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{\lambda t}{\sqrt{4+t^2}}\right)\right\} \\
&+ \frac{2\lambda t^2}{\pi(4+t^2)[4+(1+\lambda^2)t^2]}.
\end{aligned}
$$

Of course, we can proceed similarly to derive explicit expressions for $G(t;5,\lambda)$, $G(t;6,\lambda)$, and so on.

3 Recurrence Relation for Cdf of Linear Combination of order Statistics from a Bivariate-t

A random vector $\mathbf{U} = (U_1, U_2)^T$ has a *bivariate student's-t distrivution* with ν degrees of freedom and parameters $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, where $\sigma_1, \sigma_2 > 0$ and $|\rho| < 1$, if its density function is

$$f(\mathbf{u};\nu,\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{2\pi\,|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\left(1+\frac{(\mathbf{u}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{u}-\boldsymbol{\mu})}{\nu}\right)^{-\frac{\nu+2}{2}}, \qquad \mathbf{u}\,\epsilon\,R^2,$$

see Fang et al. (1990). We shall use the notation $\mathbf{U} = (U_1, U_2)^T \sim t_2(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ for the bivariate student's-t distribution. Then the following lemma is well-known[see Fang et al (1990) for example].

Lemma 1 *If* $\mathbf{U} = (U_1, U_2)^T \sim t_2(\nu, \mathbf{0}, \boldsymbol{\Sigma})$, *then*

$$\mathbf{U} \overset{d}{=} V^{-\frac{1}{2}}\mathbf{X},$$

where, $\mathbf{X} = (X_1, X_2)^T \sim N_2(\mathbf{0}, \boldsymbol{\Sigma})$ *and* $V \sim \frac{X^2_{(\nu)}}{\nu}$ *independently of* X.

In this section we consider a linear function of order statistics arising from the bivariate student's-t distribution and show that its cdf is in fact a mixture of skew-t distributions. We will then use this result to derive a recurrence relation for the cdf of this linear function of order statistics. For this purpose, let $Y = a_1X_{(1)} + a_2X_{(2)}$ and $Y^* = a_1U_{(1)} + a_2U_{(2)}$, where $\left(X_{(1)}, X_{(2)}\right)^T$ and $(U_{(1)}, U_{(2)})^T$ are the order statistics from $\mathbf{X} = (X_1, X_2)^T \sim N_2(\mathbf{0}, \boldsymbol{\Sigma})$ and $\mathbf{U} = (U_1, U_2)^T \sim$

$t_2(\nu, \mathbf{0}, \mathbf{\Sigma})$, respectively. Let us further denote the cdfs of Y and Y^* by $F(t; \mathbf{\Sigma})$ and $F^*(t; \mathbf{\Sigma})$, respectively.

We first present the following result, due to Genc (2006), which will be used in the subsequent derivations.

Lemma 2 *The cdf of* Y *, when* $a_1 + a_2 \neq 0$*, is the mixture*

$$F(t; \mathbf{\Sigma}) = \frac{1}{2}\left\{\Phi\left(\frac{t}{\sqrt{\xi_1}}; \lambda_1\right) + \Phi\left(\frac{t}{\sqrt{\xi_2}}; \lambda_2\right)\right\},$$

where $\Phi(\cdot; \lambda)$*, denotes the cdf of* $SN(\lambda)$ *as before, and*

$$\begin{aligned} \xi_1 &= a_1^2\sigma_1^2 + 2\rho a_1 a_2 \sigma_1\sigma_2 + a_2^2\sigma_2^2, \\ \xi_2 &= a_2^2\sigma_1^2 + 2\rho a_1 a_2 \sigma_1\sigma_2 + a_1^2\sigma_2^2, \end{aligned}$$

$$\lambda_1 = \frac{a_2\sigma_2^2 - a_1\sigma_1^2 - \rho\sigma_1\sigma_2(a_2 - a_1)}{|a_2 + a_1|\,\sigma_1\sigma_2\sqrt{1-\rho^2}},$$

and

$$\lambda_2 = \frac{a_2\sigma_1^2 - a_1\sigma_2^2 - \rho\sigma_1\sigma_2(a_1 - a_2)}{|a_2 + a_1|\,\sigma_1\sigma_2\sqrt{1-\rho^2}}.$$

Upon using the result in Lemma 2, we now derive the distribution of Y^* as presented in the following theorem.

Theorem 2 *The cdf of* Y^**, when* $a_1 + a_2 \neq 0$*, is the mixture*

$$F^*(t; \nu, \mathbf{\Sigma}) = \frac{1}{2}\left\{G\left(\frac{t}{\sqrt{\xi_1}}; \nu, \lambda_1\right) + G\left(\frac{t}{\sqrt{\xi_2}}; \nu, \lambda_2\right)\right\}, \qquad (13)$$

where $G(\cdot; \nu, \lambda)$ *is the cdf of* $St(\nu, \lambda)$ *as before, and* ξ_1,ξ_2, λ_1 *and* λ_2 *are all as given in lemma 2.*

Proof : We first observe that

$$Y^* \stackrel{d}{=} V^{-\frac{1}{2}}Y, \qquad (14)$$

where $V \sim \frac{X^2_{(\nu)}}{\nu}$ and is independent of Y. Now, upon using Eq. (14) along with lemma 2, we obtain

$$\begin{aligned} F^*(t;\nu,\mathbf{\Sigma}) &= P(Y^* \le t) = P(V^{-\frac{1}{2}}Y \le t) = E\left[F(tV^{\frac{1}{2}};\mathbf{\Sigma})\right] \\ &= \frac{1}{2}\left\{E\left[\Phi\left(\frac{tV^{-\frac{1}{2}}}{\sqrt{\xi_1}};\lambda_1\right)\right] + E\left[\Phi\left(\frac{tV^{-\frac{1}{2}}}{\sqrt{\xi_2}};\lambda_2\right)\right]\right\}. \end{aligned}$$

The result in Eq. (13) then readily follows by (4).

Upon using Theorem 2 in conjunction with Theorem 1, we can present readily a recurrence relation for the cdf $F^*(t;\nu,\mathbf{\Sigma})$ as given in the following theorem.

Theorem 3 *For $t \in R$ and $\nu > \frac{1}{2}$, we have*

$$\begin{aligned} F^*(t;2\nu+1,\mathbf{\Sigma}) &= F^*(\sqrt{\frac{2\nu-1}{2\nu+1}}t;2\nu-1,\mathbf{\Sigma}) + \frac{\Gamma(\nu)(2\nu+1)^{\nu-\frac{1}{2}}}{2\sqrt{\pi}\Gamma(\frac{2\nu+1}{2})} \\ &\times \{\frac{\xi_1^{\nu-\frac{1}{2}}t}{[(2\nu+1)\xi_1+t^2]^\nu}G\left(\frac{\sqrt{2\nu}\lambda_1 t}{\sqrt{(2\nu+1)\xi_1+t^2}};2\nu\right) \\ &+ \frac{\xi_2^{\nu-\frac{1}{2}}t}{[(2\nu+1)\xi_2+t^2]^\nu}G\left(\frac{\sqrt{2\nu}\lambda_2 t}{\sqrt{(2\nu+1)\xi_2+t^2}};2\nu\right)\}, \quad (15) \end{aligned}$$

where $G(t;\nu)$ is the cdf of Students-t distribution with ν degrees of freedom as before.

Proof : In the expression of the cdf of Y^* in Theorem 2, if we use the recurrence relation in Eq. (6), we easily obtain the recurrence relation in (15).

Remark 3 By setting $\nu = 1, \frac{3}{2}, 2, \frac{5}{2}, ...$ in Theorem 2 and upon using $F^*(t;1,\Sigma)$ and $F^*(t;2,\Sigma)$ from Theorem 2 and the expressions of $G(t;1,\lambda)$ and $G(t;2,\lambda)$ in Eqs. (11) and (12), we can easily obtain explicit expressions $F^*(t;3,\Sigma)$, $F^*(t;4,\Sigma)$, and so on.

Acknowledgements. The authors thank Professor Balakrishnan(Mc Master University, Hamilton, Canada) for his suggestions which improve the presentation of this paper.

References

[1] Arnold, B. C. and Beaver, R. J. (2002). Skew multivariate models related to hidden trancation and/or selective reporying (with Discussion), *Test,* ***11****, 7-54.*

[2] Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones, *Statistica*, **46**, 199-208.

[3] Azzalini, A. (2005). The skw normal distribution and related multivariate families, *Scand. J. Statist.* **32**, 159-188.

[4] Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skw-t distribution, *Journal of the Royal Statostical Society, Series B*, **65**, 367-389.

[5] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution, *Biometrika*, **83**, 715-726.

[6] Behboodian, J.,.Jamalizadeh, A. and Balakrishnan, N. (2006). A new class of skew-Cauchy distributions, *Statistics & Probability letters,* ***76****, 1488-1493.*

[7] Fang, K.-T., Kotz, S. and Ng, K.(1990) *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London.

[8] Genc, A. (2006). A note on the distribution of linear functions of two orderd correlated normal random variables, *Communications in Statistics-Theory and Methods*,**35**:989-995.

[9] Henze, N. A. (1986). A probabilistic representation of the skw-normal distribution, *Scand. J. Statist.* ***13****, 271-275.*

[10] Janson, N.L., Kotz, S. and Balakrishnan, N. (1995). Continuous Univariate Distributions-Vol. 2, Second edition, John Wiley & Sons, New York.

Regression model diagnostics under long memory

Hira L. Koul*

Michigan State University
(This talk is based on joint work done with Hongwen Guo and Donatas Surgailis)

Abstract: We shall discuss the two testing problems pertaining to regression modeling in the presence of long memory in the errors. One is to fit a parametric model to the regression function in heteroscedastic set up and when design variable has long memory. The second problem is to test a sub-hypothesis in homoscedastic multiple linear regression models. For the first problem a test based on a certain marked empirical process will be presented. For the second problem likelihood ratio type test based on the minimized Whittle quadratic form is analyzed. Asymptotic null distribution of this statistic is shown to be a chi-square distribution when design variables are random having short or long memory but have zero mean. Additionally, the estimators of the slope parameters obtained by minimizing the Whittle dispersion is seen to be $n^{1/2}$-consistent for all values of the long memory parameters of the design and error processes. In the case of non-random design, asymptotic distribution of the proposed test statistic is like a weighted chi-square and $n^{1/2}$-consistency of the minimizing slope parameter estimators no longer holds.

1 Introduction

Fitting a parametric family of functions to the regression function is a classical problem in statistics. Under the assumption of independence and/or Gaussian errors numerous tests are available in literature for this problem: cf. Hart (1997) and Li (2004). But their behavior under long memory design and/or errors has not been investigated carefully

Invited speaker

Research of this author partly supported by the NSF DMS Grant 0701430

Research of this author partly supported by the bilateral France-Lithuania scientific project Gilibert and the Lithuanian State Science and Studies Foundation grant T-15/07

to date. In this talk we shall discuss behavior of a generic stochastic process under long memory on which tests for fitting a regression model can be based. In addition, we shall also discuss tests of a subhypothesis in linear regression model when errors have long memory and design are random possibly having long memory and when design variables are non random. Let X, Y be r.v.'s, with X representing a design variable, and Y the response variable having finite second moment; $\mu(x) := E(Y|X = x)$ denote the regression function and $\sigma^2(x) := E\{(Y - \mu(x))^2|X = x\}$. Let $\bar{\mathbb{R}} := [-\infty, \infty]$, $\Omega \subset \mathbb{R}^q$ and $m(x, b)$ be a function from $\mathbb{R} \times \Omega$ to $\mathbb{R}$. Consider the hypothesis $\mathcal{H} : \mu(x) = m(x, \beta)$, for some $\beta \in \mathbb{R}^q$, $\forall\, x$, vs. $\mathcal{H}$ is not true. A process that is found useful in testing for $\mathcal{H}$ is

$$V_n(x, b) := \sum_{i=1}^{n} [Y_i - m(X_i, b)] I(X_i \leq x), \quad x \in \bar{\mathbb{R}},\, b \in \Omega.$$

This process is an extension of the partial sum process used in the one sample setup to make inference about mean to the current regression setup. von Neuman (1941) used this process for testing the hypothesis of no effect of the design, i.e. for testing $\mu(x) \equiv c$, a constant, for all x.

In 1990's, several authors found that lack-of-fit tests based on this process for regression and autoregressive model fitting have desirable level and power properties against a broad class of alternatives; see, e.g., An and Cheng (1991), Stute (1997), Stute, Thies and Zhu (1998; AS), Koul and Stute (1999), Stute and Zhu (2002), Koul (2002), Khmaladze and Koul (2004) and Koul (2006).

Let $\sigma^2(x) := E[(Y - m(X, \beta)^2)|X = x]$. Observe that under $\mathcal{H}$, $EV_n(x, \beta) \equiv 0$, and if $e_i = Y_i - m(X_i, \beta)$ are i.i.d., then

$$\begin{aligned} \mathrm{Cov}(V_n(x, \beta), V_n(y, \beta)) &= nE\{e^2 I(X \leq x \wedge y)\} \\ &= nE[\sigma^2(X) I(X \leq x \wedge y)]. \end{aligned}$$

Suppose also that the errors are homoscedastic, i.e., $\sigma^2(x) \equiv \tau^2$, a positive constant. Then this covariance is the same as that of $\tau B(G(x))$,

where B is the standard Brownian motion on $[0,\infty)$ and G is the d.f. of X.

If additionally, $\hat{\beta}$ is $n^{1/2}$-consistent linear estimator of β under $\mathcal{H}$, and m is differentiable w.r.t. b with its differential $\dot{m}(x,\beta)$ satisfying $E\|\dot{m}(X,\beta)\|^2 < \infty$, and $\Sigma := E\dot{m}(X,\beta)\dot{m}(X,\beta)'$ positive definite, then Stute (1997) and Stute, Thies and Zhu (1998) (STZ) proved that under $\mathcal{H}$, $n^{-1/2}V_n(x,\hat{\beta}) \Longrightarrow W_G$, in the Skorokhod space $D(\bar{\mathbb{R}})$ with uniform metric, where W_G is a mean zero continuous Gaussian process having the covariance function

$$\tau^2[G(x\wedge y) - E[\dot{m}(X,\beta)I(X\le x)]'\Sigma^{-1}E[\dot{m}(X,\beta)I(X\le y)].$$

Thus asymptotic null distribution of $n^{-1/2}V_n(x,\hat{\beta})$, $x\in\bar{\mathbb{R}}$ depends on the model being fitted.

In STZ a transformation of $n^{-1/2}V_n(x,\hat{\beta})$ is shown to converge weakly to a time transformed Brownian motion so that its limiting distribution is known, and hence any test based on a continuous function of this transformation will be asymptotically distribution free. An analog of this result in an autoregressive set up appears in Koul and Stute (1999). An extension to the case of higher dimension design and some other geometric aspects are discussed in Khmaladze and Koul (2004). Koul (2000) and Koul, Stute and Li (2005) discuss the case of fitting a non-smooth regression/autoregressive function.

2 Lack-of-fit testing under Long memory

A discrete time strictly stationary stochastic process with finite second moment is said to have long memory if its auto-covariances tend to zero in a hyperbolic fashion as the lag increases to infinity. Long memory processes arise in numerous physical and social sciences. See Baillie (1996), Beran (1992, 1994), Dehling, Mikosch and Sorenson (2002), and Doukhan, Oppenheimer and Taqqu (2003), for more on these processes. Regression models with long memory in design and errors are found useful in economics and finance when observing high frequency data where spot returns are regressed on forward premiums.

See e.g., Booth, Kaen and Koves (1982), Cheung (1993), Lo (1991) among others. It is hence of interest to analyze the process V_n when both design and error processes have long memory.

To keep exposition relatively transparent, we shall consider a special case of $\mathcal{H}$. Let r_j, $j = 1, \cdots, q$, be some known functions of X, $r' := (r_1, \cdots, r_q)$. Assume

$$E\|r(X)\|^2 < \infty, \ A := Er(X)r(X)' > 0.$$

Consider the problem of testing $\mathcal{H}_0 : \mu(x) = \beta' r(x)$, for some $\beta \in \mathbb{R}^q$, $\forall\, x \in \mathbb{R}$, against the alternative $\mathcal{H}_0$ is not true, based on the observations (X_i, Y_i), $i = 1, \cdots, n$ on (X, Y).

We are interested in investigating the large sample behavior of $V_n(x, \hat{\beta})$ under $\mathcal{H}_0$ when

$$Y_i = \beta' r(X_i) + \sigma(X_i)u_i, \quad u_i := \sum_{j=0}^{\infty} b_j \eta_{i-j}, \quad i \in \mathbb{Z} := \{0, \pm 1, \cdots\};$$

$$b_j \sim C j^{-(1-d)}, \ j \to \infty, \quad 0 < d < \frac{1}{2};$$

for some $a < \infty$, $b_{j+1} \le b_j(1 + j^{-1}a)$, for all sufficiently large j.

Here, η_j are standardized i.i.d. r.v.'s, independent of the X_i-process and the constant C is such that $\sum_{j=0}^{\infty} b_j^2 = 1$. Under this set up, $\sigma^2(x) = \mathrm{Var}(Y|X = x)$, $x \in \mathbb{R}$, and $E\sigma^2(X) < \infty$. For example, u_i =FARIMA $(0, d, 0)$, $0 < d < 1/2$, model satisfies the above conditions where $b_j = \Gamma(j + d) \Big/ \Gamma(j + 1)\Gamma(d)$.

We assume $\{X_i\}$ to be a Gaussian process with mean μ, auto-covariance function satisfying

$$\gamma_X(k) := \mathrm{Cov}(X_0, X_k) \sim G_X \delta(d_1) k^{-(1-2d_1)}, \quad k \to \infty,$$
$$\text{for some } 0 < d_1 < 1/2, \ G_X > 0,$$

where $\delta(a) := 2\Gamma(1-2a)\cos(\pi(1/2-a))$, $0 < a < 1/2$. It is also known that for some $G_u > 0$, the spectral densities f_u and the auto-covariance function γ_u of the u process satisfy $f_u(\lambda) \sim G_u \lambda^{-2d}$, as $\lambda \to 0_+$, and

$\gamma_u(k) \sim G_u\theta(H)k^{-(1-2d)}$, as $k \to \infty$. Because $\sum_{k=1}^{\infty}|\gamma_X(k)| = \infty = \sum_{k=1}^{\infty}|\gamma_u(k)|$, both, X- and u- processes, have long memory. The dependence between a current observation and the one at a distant future is persistent.

In the presence of long memory in design and/or errors and when $\sigma(x) \equiv \tau$, the above V_n process has been studied in Koul, Baillie and Sugailis (2004).

Now, let $\psi_1^2 := G_u\theta(d)/(d+1/2)(2d)$, Z be a $N(0,1)$ r.v., $A = E\{r(X)r(X)'\}$, and let

$$J_\sigma(x) \;:=\; E[\sigma(X)I(X \le x)] - E(\sigma(X)r(X))'A^{-1}E[r(X)I(X \le x)].$$

We also recall a result from Davydov (1970): $n^{-1/2-d}\sum_{i=1}^{n} u_i \to_D \psi_1 Z$. This result is used to prove the following theorem in Guo and Koul (2008).

Theorem 7 *Under the above set up, under $\mathcal{H}_0$ and when $\hat\beta$ is the LSE,*

$$\sup_{x\in\bar{\mathbb{R}}}|n^{-H}V_n(x,\hat\beta) - J_\sigma(x)n^{-1/2-d}\sum_{i=1}^{n}u_i| = o_p(1). \quad (\mathcal{H}_0).$$

Hence, under $\mathcal{H}_0$, $n^{-1/2-d}V_n(x,\hat\beta) \Longrightarrow J_\sigma(x)\psi_1 Z$, in $\mathcal{D}(\bar{\mathbb{R}})$, and uniform metric.

Unlike in the i.i.d. set up, the above limiting process is degenerate, i.e., there is a separation of the randomness and the functional behavior in the limit. On the one hand this makes life a bit simpler as only the normal distribution tables are used to implement the test. But on the other hand, in order to implement the tests based on this process we must have consistent estimators of $\sigma(x)$ and G_u and $\log(n)$-consistent estimator of d. To construct these estimators, let K be a density function on $[-1,1]$, $b = b_n$ be sequence of positive numbers, ϕ denote the $N(0,1)$ density, $K_b(x) \equiv K(x/b)/b$, $K_{bi}(x) := K_b(x - X_i)$,

$\hat{e}_i := Y_i - \hat{\beta}' r(X_i)$, and define

$$\hat{\Lambda}_{-i}(x) := \Big(\frac{1}{n-1}\sum_{t\neq i}^{n} K_{bt}(x)\tilde{e}_t^2\Big)^{1/2}, \quad v_i(x) := \frac{\hat{\Lambda}_{-i}(x)}{\phi^{1/2}(x)},\ i = 1, \cdots, n,$$

$$A_n := \frac{1}{n}\sum_{i=1}^{n} r(X_i)r(X_i)', \ \ \alpha_n(x) := \frac{1}{n}\sum_{i=1}^{n} r(X_i)I(X_i \leq x)$$

Think of $v_i(x)$ as an estimator of $\sigma(x)$. It is used to define an estimator of $J_\sigma(x)$:

$$n\hat{J}_n(x) = \sum_{i=1}^{n} v_i(X_i)I(X_i \leq x) - \sum_{i=1}^{n} r(X_i)' v_i(X_i) A_n^{-1}\alpha_n(x).$$

Lemma 3 *Suppose the above set up and $\mathcal{H}_0$ hold with $\mu = 0$, $\gamma = 1$ and the function σ has continuous first derivative. In addition, suppose $b \to 0$, $n^{2h-2} = O(b)$, $E\|r(2X)\|^2 < \infty$, $E\sigma^{2k}(X)r_j^4(X) < \infty$, for $j = 1, \cdots, q$, $k = 0, 1$, and $E\sigma^2(X)\phi^{1/2}(X) < \infty$.*

Then, under $\mathcal{H}_0$, $\sup_{x\in\bar{\mathbb{R}}} |\hat{J}_n(x) - J_\sigma(x)| = o_p(1)$.

Next, we describe estimators of G_u, d in the above set up under $\mathcal{H}_0$. For a process $\xi_j, 1 \leq j \leq n$, let $w_\xi(\lambda) := (2\pi n)^{-1/2}\sum_{i=1}^{n}\xi_j e^{\mathbf{i}j\lambda}$, $I_\xi(\lambda) := |w_\xi(\lambda)|^2$, $\lambda \in [-\pi, \pi]$, denote its discrete Fourier transform and periodogram, respectively, where $\mathbf{i} := (-1)^{1/2}$. Fix $1/2 < a_1 < a_2 < 1$. With $\lambda_j := 2\pi j/n$ and an integer $m \in [1, n/2)$, for $a_1 \leq \alpha \leq a_2$, let

$$Q(\alpha) := \frac{1}{m}\sum_{j=1}^{m} \lambda_j^{2\alpha-1} I_{\hat{e}}(\lambda_j), \quad R(\alpha) := \log Q(\alpha) - (2\alpha - 1)\sum_{j=1}^{m}\log\lambda_j.$$

Local Whittle estimators of G_u and d: $\hat{G}_u = Q(\hat{d})$, $\hat{d} = \text{argmin}_{\alpha\in[a_1,a_2]}$ $R(\alpha)$.

The $\log(n)$ consistency of an analog of $\hat{d}$ and consistency of an analog of $\hat{G}_u$ in nonparametric homoscedastic regression models with $X_i = i/n$, $i = 1, \cdots, n$ is proved in Robinson (1997). Using his

methodology of proof the analogues of these results under the current set up are proved in Guo and Koul (2007). Let $\hat{\psi}_1 := \hat{G}_u\theta(\hat{d})/(\hat{d}+1/2)(2\hat{d})$.

A consequence of all these results is that if $\sup_x |J_\sigma(x)| \neq 0$, the test that rejects $\mathcal{H}_0$, whenever,

$$D_n := \frac{1}{n^{\hat{d}}\hat{\psi}_1 \sup_x |\hat{J}_n(x)|} \sup_x |V_n(x,\hat{\beta})| \geq z_{\alpha/2},$$

is of the asymptotic size α. Here z_α is the $100(1-\alpha)\%$ percentile of the $N(0,1)$ d.f.

In the simple linear regression model with non-zero intercept, i.e., when $r(x) = (1,x)'$, $J_\sigma(x) \equiv 0$ if and only if $\sigma(x)$ is constant in x. In the case of a polynomial regression through the origin, $\sup_x |J_\sigma(x)| \neq 0$ and the above test is applicable when fitting a heteroscedastic polynomial.

2.1 Application to a foreign exchange data set

We shall apply the above proposed test to fit a simple linear regression model with heteroscedastic errors to some currency exchange rate data. The data are noon buying rates in New York for cable transfers payable in foreign currencies. We use the currency exchange rates of the United Kingdom Pounds (UK£) vs. US$ and the Switzerland Franc (SZF) vs. US$ from January 4, 1971 to December 2, 2005 obtained from
www.federalreserve.gov/releases/H10/hist/. We first deleted missing values and obtain 437 monthly observations. The symbols $X=$ dlUK and $Y=$ dlSZ stand for differenced log exchange rate of UK£ vs. US$ and SZF vs. US$, respectively. We obtain

$$\begin{aligned} \bar{X} &= -0.0001775461,\ s_X = 0.001701488;\ \bar{Y} = -0.00004525129, \\ & \quad s_Y = 0.001246904. \end{aligned}$$

The local Whittle estimates of d_1 and d are $\hat{d}_1 = .1610273$ and $\hat{d} = .2147475$. In computing these estimates we used $m = [n/8] =$

54. These estimates point to the presence of long memory in these processes.

Comparing the X-process with a simulated fractional Gaussian noise with $\hat{d} = 0.1610273$ and $n = 437$, Figure 1 suggests that the marginal distribution of X is Gaussian.

Next, we regress Y on X, using the normal density kernel regression function estimator and a simple linear regression model. Both of these estimates are depicted in Figure 2.

They display a negative association between X and Y. The estimated linear equation: $\hat{Y} = -0.000118775 - 0.4141107\,X$, with a residual standard error of 0.00102992.

Figure 3 plots $\hat{\sigma}(x)$, where with $\varphi_n(x) := s^{-1}\phi((x - \bar{X})/s)$,

$$\hat{\sigma}^2(x) := \frac{1}{n\varphi_n(x)} \sum_{i=1}^{n} K_{bi}(x)\hat{e}_i^2, \quad K(x) = .5(1 + \cos(x\pi))I(|x| \leq 1).$$

The estimators of d based on $\hat{\varepsilon} = Y - \hat{\beta}X$ and $\hat{u} = (Y - \hat{\beta}X)/\hat{\sigma}(X)$ are equal to 0.1046235 and 0.1246576, respectively. This again suggests the presence of long memory in the error process.

Finally, to check if the regression of Y on X is simple linear, we obtain $D_n = 0.4137897$ with the asymptotic p-value 66%. As expected, this test fails to reject the null hypothesis that there exists a linear relationship between these two processes.

3 Testing a sub-hypothesis

A classical problem in statistics is to see whether among a given set of predictor variables a subset of variables is significant or not for predicting a response variable. In the regression set up a way to do this is by first stipulating a linear regression model and then testing for the absence of some of these covariates by testing that the corresponding slope parameters are zero. This is so called the problem of testing a sub-hypothesis in linear regression set up.

Let n, p and k be known positive integers with $k \leq p$; $\boldsymbol{X}_{i1}(\boldsymbol{X}_{i2}), 1 \leq i \leq n$, be $k \times 1$ $((p-k) \times 1)$ random vectors and Y_i, $1 \leq i \leq n$ denote

Figure 1: QQ-plot of dlUK.

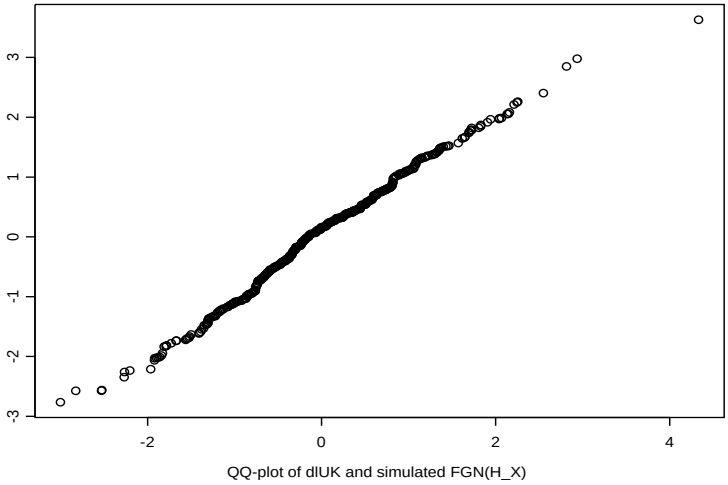


Figure 2: Kernel estimation of $r(x)$.

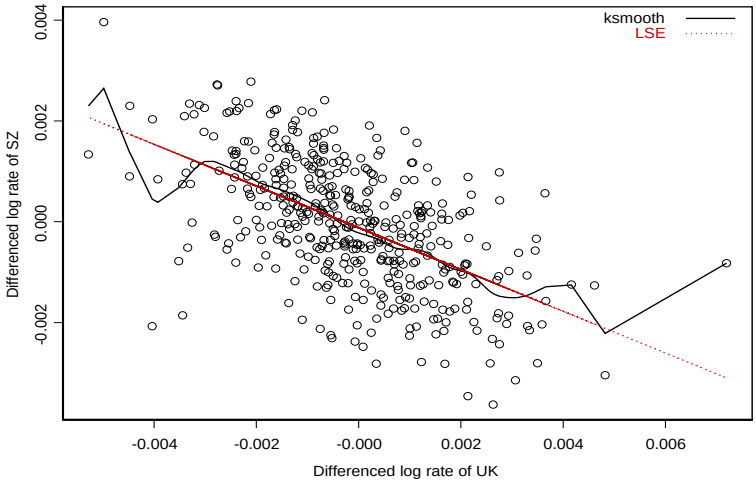
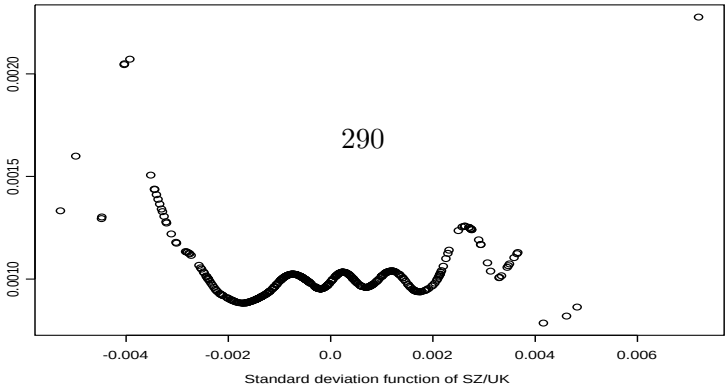


Figure 3: Kernel estimation of $\sigma(x)$.



the response variables. Consider the regression model where for some $\beta_1 \in \mathbb{R}^k$ and $\beta_2 \in \mathbb{R}^{p-k}$,

$$Y_i = \beta_1' \boldsymbol{X}_{i1} + \beta_2' \boldsymbol{X}_{i2} + \varepsilon_i, \quad i = 1, \cdots, n. \tag{1}$$

The problem of interest is to test $H_0 : \beta_2 = 0$, vs. $H_1 : \beta_2 \neq 0$.

In the case of independent homoscedastic errors $\{\varepsilon_i\}$ and when the design variables are either non-random or random and i.i.d., this problem is well studied in the literature: Rao (1985), Hart (1997) and references therein. A classical testing procedure is the likelihood ratio test when errors are Gaussian or the analysis of variance type tests via the least square theory which are asymptotically valid without the Gaussianity assumption. Here we shall investigate large sample behavior of an analog of the quasi-likelihood ratio type test for H_0 when errors form a LMMA process and covariates are random having long and/or short memory, independent of the errors.

Let Θ be an open and relatively compact subset of $\mathbb{R}^q$, $q \geq 1$. The process $\{\varepsilon_i,\, i \in \mathbb{Z}\}$ is given by said to form a parametric long memory moving average if for some functions $\alpha :\, \Theta \mapsto (0,1)$ and $c :\, \Theta \mapsto \mathbb{R}$,

$$\varepsilon_i \;:=\; \sum_{s\in\mathbb{Z}} a_{i-s}(\vartheta)\zeta_s, \quad t \in \mathbb{Z}, \quad a_i(\vartheta) \sim c(\vartheta) i^{\alpha(\vartheta)/2-1}\, (i \to \infty). \tag{2}$$

Here ζ_j, $j \in \mathbb{Z}$, are standardized i.i.d r.v.'s having finite fourth moment. Let

$$\hat{a}(u;\vartheta) \;:=\; (2\pi)^{-1} \sum_{t\in\mathbb{Z}} a(t;\vartheta) e^{-\mathbf{i}tu}, \quad f(u;\vartheta) := 2\pi\, |\hat{a}(u;\vartheta)|^2, \quad u \in \Pi,$$

$$b_j(\vartheta) \;:=\; \int_\Pi e^{\mathbf{i}\, ju} \hat{b}(u;\vartheta) du, \quad j \in \mathbb{Z},\ \vartheta \in \Theta.$$

Let $\boldsymbol{X}_i' := (\boldsymbol{X}_{i1}', \boldsymbol{X}_{i2}')$, $\beta' := (\beta_1', \beta_2')$ and define

$$\begin{aligned}
\Lambda_n(\vartheta, \beta) &:= \sum_{t,s=1}^{n} b_{t-s}(\vartheta)(Y_t - \boldsymbol{X}_t'\beta)(Y_s - \boldsymbol{X}_s'\beta), \\
&\quad (\hat{\theta}_n, \hat{\beta}_n) := \text{argmin}_{(\vartheta,\beta)\in\Theta\times\mathbb{R}^p}\Lambda_n(\vartheta, \beta), \\
\Lambda_{n1}(\vartheta, \beta_1) &:= \sum_{t,s=1}^{n} b_{t-s}(\vartheta)(Y_t - \boldsymbol{X}_{t1}'\beta_1)(Y_s - \boldsymbol{X}_{s1}'\beta_1), \\
(\hat{\theta}_{n1}, \hat{\beta}_{n1}) &:= \text{argmin}_{(\vartheta,\beta_1)\in\Theta\times\mathbb{R}^k}\Lambda_{n1}(\vartheta, \beta_1).
\end{aligned}$$

Analog of the likelihood ratio test for H_0 would be based on

$$\mathcal{Q}_n := -2[\Lambda_n(\hat{\theta}_n, \hat{\beta}_n) - \Lambda_{n1}(\hat{\theta}_{n1}, \hat{\beta}_{n1})].$$

Strictly speaking, in the case of Gaussian errors the exact likelihood ratio test would have the elements of the inverse of the covariance matrix as weights in the quadratic forms Λ_n's instead of $\{b_{t-s}(\vartheta)\}$. The above quadratic forms are their Whittle approximations.

Assumptions about f: In the sequel, θ, β_0 denote the true values of ϑ, β, respectively, and Θ_0 denotes an arbitrarily small neighborhood of θ. For any twice differentiable function $h(\vartheta)$ let

$$\begin{aligned}
&\nabla_\vartheta h(\vartheta) := (\partial h(\vartheta)/\partial\vartheta_i)_{i=1,\cdots,q}; \\
&\nabla^2_{\vartheta\vartheta} h(\vartheta) := (\partial^2 h(\vartheta)/\partial\vartheta_i\partial\vartheta_j)_{i=1,\cdots,q;\, j=1,\ldots,q}.
\end{aligned}$$

Write

$$\nabla^2_{\beta\beta} h(\theta, \beta_0)$$

for $\nabla^2_{\beta\beta} h(\vartheta, \beta)|_{\vartheta=\theta,\beta=\beta_0}$, $\nabla^2_{\vartheta\beta} h(\vartheta, \beta_0)$ for $\nabla^2_{\vartheta\beta} h(\vartheta, \beta)|_{\beta=\beta_0}$, etc.

Consider the following assumptions on f:

(a.1) $f(u;\vartheta), u \in \Pi$, determines ϑ uniquely and $\int_{-\pi}^{\pi} \log f(u;\vartheta)du = 0$, $\vartheta \in \Theta_0$.

(a.2) $\int_\Pi \log f(u;\vartheta)du$ is twice differentiable under the sign of integral.

Furthermore, there exist a function α from Θ to $(0,1)$ that is continuous at θ and a constant $C < \infty$ satisfying the following conditions (a.3)-(a.6).

(a.3) $f(u,\vartheta)$ is continuous at all (u,ϑ), $u \neq 0$, $\vartheta \in \Theta_0$, f^{-1} is continuous on $\Pi \times \Theta_0$ and

$$f(u;\vartheta) \leq C|u|^{-\alpha(\vartheta)}, \qquad \forall\, (u,\vartheta) \in \Pi \times \Theta_0.$$

(a.4) $\nabla_\vartheta f^{-1}$ and $\nabla^2_{\vartheta\vartheta} f^{-1}$ are continuous on $\Pi \times \Theta_0$ and

$$\begin{aligned} |\nabla_\vartheta f^{-1}(u;\vartheta)| &\leq C|u|^{\alpha(\vartheta)}, \\ |\nabla^2_{\vartheta\vartheta} f^{-1}(u;\vartheta)| &\leq C|u|^{\alpha(\vartheta)/2}, \qquad \forall\ (u,\vartheta) \in \Pi \times \Theta_0. \end{aligned}$$

(a.5) $|\nabla_\vartheta(\partial f^{-1}(u;\vartheta)/\partial u)| \leq C|u|^{\alpha(\vartheta)-1},\ \forall\ (u,\vartheta) \in \Pi \times \Theta_0.$

(a.6) $f^{-1}(u;\vartheta) \leq C|u|^{\alpha(\vartheta)/2}, \quad \forall\, (u,\vartheta) \in \Pi \times \Theta_0.$

Conditions (a.1)-(a.5) are similar to those used in Fox and Taqqu (1986), Dahlhaus (1989), Giraitis and Surgailis (1990), and Koul and Surgailis (2000). The condition (a.6) is new.

Assumptions about random $\boldsymbol{X}$: Write $\boldsymbol{X}_i' = (X_{i,1}, X_{i,2}, \cdots, X_{i,p})$. We assume

$$X_{i,k} = \sum_{j=1}^{p} \sum_{s\in\mathbb{Z}} B_{kj}(i-s)\xi_{s,j}, \quad i \in \mathbb{Z}; \quad \sum_{s\in\mathbb{Z}} B^2_{kj}(s) < \infty, \quad (3)$$
$$k, j = 1, 2, \ldots, p.$$

Here, $\xi_s := (\xi_{s,1}, \cdots, \xi_{s,p})'$, $s \in \mathbb{Z}$ are i.i.d. standardized r.v.'s, and independent of $\zeta_s, s \in \mathbb{Z}$, implying the independence of designs and errors in (1). The model (4) includes both the short memory (in particular, i.i.d.) and the long memory random designs, but **it does not allow for a non-zero intercept parameter in (1)**.

Let $\mathbf{g}(u)$ be the $p \times p$ matrix-valued spectral density of $\boldsymbol{X}$,

$$V(\vartheta) := \int_\Pi \mathbf{g}(u) f^{-1}(u;\vartheta)du, \quad W(\vartheta) := \int_\Pi f(u;\vartheta)\nabla^2_{\vartheta\vartheta} f^{-1}(u;\vartheta)du.$$

Write

$$\mathbf{g}(u) = \begin{pmatrix} \mathbf{g}_{11}(u) & \mathbf{g}_{12}(u) \\ \mathbf{g}_{12}(u)' & \mathbf{g}_{22}(u) \end{pmatrix}, \; V(\vartheta) = \begin{pmatrix} V_{11}(\vartheta) & V_{12}(\vartheta) \\ V_{12}(\vartheta)' & V_{22}(\vartheta) \end{pmatrix},$$

where $\mathbf{g}_{12}(u)$ and $\mathbf{g}_{22}(u)$ are matrix-valued (cross-)spectral densities of $\boldsymbol{X}_{i1}$ and $\boldsymbol{X}_{i2}$ having dimensions $k\times p$ and $(p-k)\times(p-k)$, respectively. Recall

$$\mathcal{Q}_n := -2[\Lambda_n(\hat{\theta}_n, \hat{\beta}_n) - \Lambda_{n1}(\hat{\theta}_{n1}, \hat{\beta}_{n1})].$$

Theorem 1 *Assume the model (1), (a.1) - (a.6), (4) hold and $V(\theta)$ is positive definite. Then, under H_0, $(2\pi)\mathcal{Q}_n \to_D \chi^2_{p-k}$.*

The proof of the above theorem is facilitated by the following lemma which is proved in Koul and Surgailis (2008a). Let Z_θ and Z_β denote two independent random vectors with Z_θ (Z_β) having $\mathcal{N}_d(0, 16\pi^3 W(\theta))$ ($\mathcal{N}_p(0, 8\pi^3 V(\theta))$) distribution. Then $(Z_\theta', Z_\beta')'$ has $\mathcal{N}_{d+p}(0, \Gamma)$ distribution where

$$\Gamma := \begin{pmatrix} 16\pi^3 W(\theta) & 0 \\ 0 & 8\pi^3 V(\theta) \end{pmatrix}.$$

We also need to define

$$\begin{aligned} \Lambda_{n0}(\vartheta) &:= \Lambda_n(\vartheta, \beta_0) := \sum_{t,s=1}^{n} b_{t-s}(\vartheta)\varepsilon_t\varepsilon_s, \\ & Z_{n\theta} := n^{-1/2}\nabla_\vartheta \Lambda_{n0}(\theta), \quad T_n(\vartheta) := \sum_{t,s=1}^{n} b_{t-s}(\vartheta)\varepsilon_t \boldsymbol{X}_s, \\ Z_{n\beta} &:= n^{-1/2} T_n(\theta), \quad T_{n1}(\vartheta) := \sum_{t,s=1}^{n} b_{t-s}(\vartheta)\varepsilon_t \boldsymbol{X}_{s1}, \\ & * \quad Z_{n\beta 1} := n^{-1/2} T_{n1}(\theta). \end{aligned}$$

Lemma 4 *Under the conditions of Theorem 8,*

$$\Lambda_n(\hat{\theta}_n, \hat{\beta}_n) = \Lambda_n(\theta, \beta_0) - \frac{1}{2} Z'_{n\theta}(2\pi W(\theta))^{-1} Z_{n\theta} - \tag{4}$$
$$Z'_{n\beta}(2\pi V(\theta))^{-1} Z_{n\beta} + o_p(1),$$
$$\Lambda_{n1}(\hat{\theta}_n, \hat{\beta}_{n1}) = \Lambda_{n1}(\theta, b_{01}) - \frac{1}{2} Z'_{n\theta}(2\pi W(\theta))^{-1} Z_{n\theta} \tag{5}$$
$$-Z'_{n\beta 1}(2\pi V_{11}(\theta))^{-1} Z_{n\beta 1} + o_p(1). \, (H_0)$$
$$(Z'_{n\theta}, Z'_{n\beta}) \to_D (Z'_\theta, Z'_\beta), \quad n^{1/2}\|\hat{\theta}_n - \theta\| \tag{6}$$
$$+n^{1/2}\|\hat{\beta}_n - \beta_0\| = O_p(1).$$

Sketch of the Proof of Theorem 8. Since, under H_0, $\Lambda_n(\theta, \beta_0) = \Lambda_{n1}(\theta, \beta_{01})$, (5) and (6) yield

$$\mathcal{Q}_n \;=\; 2[Z'_{n\beta}(2\pi V(\theta))^{-1} Z_{n\beta} - Z'_{n\beta 1}(2\pi V_{11}(\theta))^{-1} Z_{n\beta 1}] + o_p(1).$$

Hence (6) implies

$$\mathcal{Q}_n \to_D \mathcal{Q} := 2[Z'_\beta(2\pi V(\theta))^{-1} Z_\beta - Z'_{\beta 1}(2\pi V_{11}(\theta))^{-1} Z_{\beta 1}],$$

where $Z_{\beta 1}$ is the vector of the first k components of Z_β. Now the claim follows from the following fact. Let $Z \sim \mathcal{N}_p(0, \Sigma)$ where Σ is a positive definite covariance matrix, and let Z_1 denote its first k components with covariance matrix Σ_{11}. Then, the distribution of $Z'\Sigma^{-1}Z - Z'_1\Sigma_{11}^{-1}Z_1$ is χ^2_{p-k}. See, e.g., Rao (1985). Contents of this section are from Koul and Surgailis (2008a).

4 Non-random $\boldsymbol{X}$

Here the results are different. We shall assume $\boldsymbol{X}_i = \boldsymbol{V}(i/n)$, $1 \le i \le n$, where $\boldsymbol{V}(i)' = (V_1(i), \cdots, V_p(i))$ is a vector of p continuously differentiable real-valued functions on $[0, 1]$ and linearly independent in $L^2[0, 1]$. Let $\boldsymbol{V}_1(i)' = (V_1(i), \ldots, V_k(i))$, $\boldsymbol{V}_2(i)' = (V_{k+1}(i), \ldots, V_p(i))$. Then $\boldsymbol{V}(i) = (\boldsymbol{V}_1(i)', \boldsymbol{V}_2(i)')'$. We also need to modify assumption (a.3) on f:

(a.3′) $f(u,\vartheta)$ is continuous at all (u,ϑ), $u \neq 0$, $\vartheta \in \Theta_0$, f^{-1} is continuous on $\Pi \times \Theta_0$ and

$$f(u;\vartheta) \sim C(\vartheta)|u|^{-\alpha(\vartheta)}, \quad (u \to 0, \vartheta \in \Theta_0).$$

In (a.3), $f(u;\vartheta) \leq C(\vartheta)|u|^{-\alpha(\vartheta)}$.

Next, define

$$\begin{aligned} Z_{n\beta} &:= n^{-(1-\alpha)/2} \sum_{t,s=1}^{n} b_{t-s}(\theta)\varepsilon_t \boldsymbol{V}(s/n), \quad \hat{\boldsymbol{V}}(x) := \int_0^1 e^{\mathbf{i}xt}\boldsymbol{V}(i)dt, \\ & x \in \mathbb{R}, \\ R_V(\vartheta) &:= \frac{1}{C(\vartheta)} \int_{\mathbb{R}} |x|^{\alpha} \hat{\boldsymbol{V}}(x)\overline{\hat{\boldsymbol{V}}(x)}' dx, \\ \Gamma_V &:= \frac{1}{C(\theta)} \int_{\mathbb{R}^3} |x_1|^{\alpha}|x_2|^{\alpha}|y|^{-\alpha} \times \frac{(e^{\mathbf{i}(y-x_1)}-1)(e^{\mathbf{i}(y-x_2)}-1)}{(y-x_1)(y-x_2)} \\ & \quad \hat{\boldsymbol{V}}(x_1)\overline{\hat{\boldsymbol{V}}(x_2)}' dx_1 dx_2 dy, \end{aligned}$$

Also, let $Z_{n\beta 1}$, $R_{V1}(\vartheta)$, Γ_{V1} be defined as above with $\boldsymbol{V}$ and $\hat{\boldsymbol{V}}$ replaced by $\boldsymbol{V}_1$ and $\hat{\boldsymbol{V}}_1(x) := \int_0^1 e^{\mathbf{i}xt}\boldsymbol{V}_1(t)dt$. Let $\mathcal{Q} := 2[Z'_\beta R_V^{-1} Z_\beta - Z'_{\beta 1} R_{V1}^{-1} Z_{\beta 1}]$.

Theorem 8 *Under (1), (a.1), (a.2), (a.3′) - (a.5), and under H_0, $\mathcal{Q}_n \to_D \mathcal{Q}$.*

The distribution of this $\mathcal{Q}$ is not a chi-square. The proof of this result is given in Koul and Surgailis (2008b).

References

[1] An, H.Z. and Cheng, B. (1991). A Kolmogorov-Smirnov type statistic with application to test for nonlinearity in time series. *Int. Statist. Rev.*, **59**, 287-307.

[2] Baillie, R. T,. (1996). Long memory processes and fractional integration in econometrics. *J. Econometrics*, **73**, 5–59.

[3] Booth, G. G., Kaen, R. R. and P. E. Koves. (1982). R/S analysis of foreign exchange rates under two international money regimes. *Journal of Monetary Economics*, **10**, 407-415.

[4] Beran, J. (1992). Statistical methods for data with long-range dependence. (With discussion). *Statistical Science*, **7**, 404-427.

[5] Beran, J. (1994). *Statistics for Long Memory Processes*. Monographs on Statistics and Applied Probability, Vol. 61. Chapman and Hall, New York 1994.

[6] Cheung, Y.-W. (1993). Long memory in foreign exchange rates. *Journal of Business and Economic Statistics*, **11**, 93-101.

[7] Davydov, Y. (1970). The invariance principle for stationary process. *Theory of Probab. and its Appl.* **XV** 487-498.

[8] Dehling, H., Mikosch, T. and Sørensen, M. (Eds.) (2002). *Empirical Process Techniques for Dependent Data.* Birkhäuser, Boston.

[9] Doukhan, P., Oppenheim, G. and Taqqu, M. S. (Eds.) (2003). *Theory and Applications of Long-Range Dependence.* Birkhäuser, Boston.

[10] Fox, R. and Taqqu, M. S. (1986). Large sample properties of parameter estimates for strongly dependent Gaussian time series. *Ann. Statist.* **14**, 517–532.

[11] Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its applications to the asymptotic normality of Whittle estimates. *Prob. Theory Relat. Fields*, **86**, 87-104.

[12] Guo, H. and Koul, H. L. (2007). Nonparametric regression with heteroscedastic long memory errors. *J. Statist. Plann. Inference*, **137**, 379–404.

[13] Guo, H. and Koul, H. L. (2007). Asymptotic inference in some heteroscedastic regression models with long memory design and errors. *Annals Statist.***36**, 458-487.

[14] Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer Series in Statistics. Springer-Verlag, New York.

[15] Khmaladze, E. V. & Koul, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Annals Statist.***32**, no. 3, 995–1034.

[16] Koul, H. L. (2000). Fitting a two phase linear regression model. Special issue dedicated to Professor Vasant P. Bhapkar. *J. Indian Statist. Assoc.* **38** no. 2, 331–353.

[17] Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models. Second Edition*. Springer Lecture Notes in Statistics, **166**. Springer, N.Y.

[18] Koul, H. L. (2006). Model diagnostics via martingale transforms: a brief review. *Frontiers in statistics*, 183–206, Imp. Coll. Press, London, 2006.

[19] Koul, H. L., Baillie, R. T. and Surgailis, D. (2004). Regression model fitting with a long memory covariate process. *Econometric Theory*, **20**, no. 3, 485–512.

[20] Koul, H. L. and Surgailis, D. (2000). Asymptotic normality of the Whittle estimator in linear regression models with long memory errors. *Statistical Inference for Stochastic Processes*, **3**, 129-147.

[21] Koul, H. L. and Surgailis, D. (2008a). Testing of a sub-hypothesis in linear regression models with long memory covariates and errors. *Applications of Mathematics*, **53(3)**, 235-248.

[22] Koul, H. L. and Surgailis, D. (2008b). Testing of a sub-hypothesis in linear regression models with long memory errors and non-random design. A preprint.

[23] Koul, H.L. and Stute, W. (1998). Regression model fitting with long memory errors. *J. Statist. Planning and Inference*, **71**, 19-34.

[24] Koul, H.L. and Stute, W. (1999). Nonparametric model checks for time series. *Ann. Statist.* **27**, 204-236.

[25] Koul, H. L., Stute, W. and Li, Fang. (2005). Model diagnosis for SETAR time series. *Statist. Sinica*, **15** no. 3, 795–817.

[26] Li, W.K. (2004). *Diagnostic Checks and Time Series*. Monographs on Statistics and Applied Probability, Chapman and Hall, Boca Raton, Florida.

[27] Lo, A. (1991). Long term memory in stock market prices. *Econometrica* **59**, 1279-1313.

[28] C.R. Rao: *Linear Statistical Inference and Its Applications*. Wiley, New York 1985.

[29] Robinson, P. M. (1997). Large-sample inference for nonparametric regression with dependent errors. *Ann. Statist.*, **25**, 2054-2083.

[30] Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.*, **25**, 613-641.

[31] Stute, W., Thies, S. and Zhu, L.X. (1998). Model checks for regression: An innovation process approach. *Ann. Statist.*, **26**, 1916-1934.

[32] Stute, W and Zhu, L.X. (2002). Model checks for generalized linear models. *Scand. J. Statist.* **29**, no. 3, 535–545.

[33] von Neumann, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Statistics*, **12**, 367–395.

Bayesian Model Selection for ARCH Models Using Iterated Importance Sampling

Fazlollah Lak

Department of Statistics and Mathematics, Faculty of Sciences, Persian Gulf University, Bushehr 75169, Iran

Abstract: We propose a full Bayesian analysis of ARCH models with unknown order. The order will be considered just as any other parameter in the model. This will allow to perform simultaneously the model choice and the parameter estimation.

For the implementation we use Population Monte Carlo algorithm. This method is introduced by Cappé *et al.* (2004). Our results are illustrated by a simulation. The simulated results for a special ARCH model with a known order show that estimation of order and parameters are good.

1 Introduction

The first model that provides a systematic framework for volatility modeling is the ARCH model of Engle (1982). Many extensions and applications of this class of models were proposed in the literature. Bollorslev (1986) put forward the generalized ARCH (GARCH) with combines parsimony in the parameters and flexibility in the lag structure of the conditional variance. Analyzing financial and economic time series data with ARCH and GARCH models has become very common in empirical research, with a huge literature having been established; see, for example, the review article of Bollerslev *et al.* (1992). Despite both the attractiveness and complexity of these models, the Bayesian community has not yet broadly utilized the recent advances in statistical computation to effect a practical statistical analysis "package". Geweke (1988, 1989) was first given a Bayesian treatment of the ARCH models. The method is based on Monte Carlo integration with importance sampling. Kliebergen and Van Dijk (1993) and Bauwens and Lubrano (1998) gave a Bayesian treatment of ARCH

lak@pgu.ac.ir; fazlollahlak@gmail.com

model and some extensions from a computational point view. Vrontos *et al.* (2000) used reversible-jump to choose between a GARCH and EGARCH model selection in these models. We use iterated importance sampling for model choice for ARCH models. Celeux *et al.* (2006) used this method for missing data problems. We propose to use a Population Monte Carlo (hereafter abbrivated as PMC) algorithm. This method is introduced by Cappé *et al.* (2004). This kind of algorithms is based on importance sampling, sharing its unbiasedness properties, but introducing a temporal dimensions in the selection of the importance function. In this way, an adaptive perspective can be achieved at little cost, for a potentially large gain in efficiency. The results will be showed by a simulation. The remainder of the paper is organized as follows. In Section 2 we present in some detail the ARCH model and some of its properties. Monte Carlo integration with importance sampling and PMC with the details for model choice are illustrated in Section 3. In Section 4, we illustrate our results through a simulation of an ARCH process with known order. Section 5 conclude the paper with some remarks.

2 ARCH models

In this section, we present the model, the basic assumptions and the notation used through the paper.
The variance of returns on assets tends to change over time. One way of modeling this feature of the data is to let the conditional variance be a function of the squares of previous observations and past variances. This leads to the autoregressive conditional heteroscedasticity (ARCH) based models developed by Engle (1982). One of the key ideas behind the whole family of ARCH models is that models that make use of recent available information will be able to forecast better than other models that do not take into account this information. This is one of the reasons why these models benefit particularly from focusing on establishing the difference between conditional and unconditional moments. Conventional econometrics models do not allow for

a conditional variance whose values depend on past information, so volatility clustering is not a phenomenon that can be understood with the aid of these traditional models. The basic idea of ARCH models is that the asset return y_t are dependent but serially uncorrelated, and the dependence of y_t can be described by a simple quadratic function of its lagged values. Due to all the reasons explained above, a new type of processes, called ARCH processes, were developed by Engle in an attempt to create a model in which the recent past would give information about the one-period forecast variance. The main characteristics of ARCH process are that they have a zero mean, a constant unconditional variance and, most importantly, they have non-constant variances conditional on the past. In order to model a variance that is not constant over time, a phenomenon also known as *heteroscedasticity*, the approach given by Engle was to propose the following model

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t, & (1)\\ \sigma_t^2 &= \alpha_0 + \alpha_1 y_{t-1}^2 & (2) \end{aligned}$$

where $\varepsilon_t \sim iid(0,1)$ (white noise with unit variance), and σ_t is a positive, time-varying function of Ψ_{t-1}, which is the information set available at time $t-1$. In practice, ε_t is often assumed to follow the standard normal or a standardized Student-t distribution. The model defined by equations (1) and (2) is known as an "Autoregressive Conditional Heteroscedasticity" model. The variance function described by equation (2) can be generalized in order to include information that goes further back than only one period, in this case we obtain

$$\sigma_t^2 = h(y_{t-1}, y_{t-2}, ..., y_{t-p}, \alpha), \quad (3)$$

where p is called the *order* of the ARCH process, and α is a vector of unknown parameters. Thus, the first model presented, as described by equations (1) and (2) is the first order ARCH process, where the vector of unknown parameters is $\alpha = [\alpha_0, \alpha_1]$. The variance function, equation (3) can be written in its most general form as $\sigma_t^2 = h(\Psi_{t-1}, \alpha)$. In particular, Engle (1982) suggested that one possible way of param-

eterizing σ_t^2 is:

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 \\ &= \alpha_0 + \alpha(L) y_t^2, \end{aligned} \tag{4}$$

where $\alpha_0 > 0$ and $\alpha_i \geq 0$ and L is the lag operator. Equation (4) expresses σ_t^2 as a linear function of p-past squared values the process, a reason why this process is known as the linear p-order ARCH model (ARCH(p)). As it can easily be observed, this process is a straightforward continuation of the more simple process described by equations (1) and (2), i.e. the ARCH(1) process. The mean of an ARCH process, y_t, described by equations (1) and (3), is zero as well as all its autocovariances. The unconditional variance will be given by $E(y_t^2) = E(\sigma_t^2)$, and y_t will be covariance stationary for the function h and the values α where the variance is independent of t. The covariance stationary conditions for a p^{th}-order ARCH process are given by (Engle, 1982, page 993):
"Theorem 2: *The* p^{th}-order linear ARCH processes, with $\alpha_0 > 0$, $\alpha_1, ..., \alpha_p \geq 0$, is covariance stationary if, and only if, the associated characteristic equation has all roots outside the unit circle the stationary variance is given by

$$E(y_t^2) = \frac{\alpha_0}{(1 - \sum_{j=1}^{p} \alpha_j)}.$$

Defining $\upsilon_t \equiv y_t^2 - \sigma_t^2$, the ARCH(p) model can be re-written as

$$y_t^2 = \alpha_0 + \alpha(L) y_t^2 + \upsilon_t \tag{5}$$

(Notice that $\sigma_t^2 = E(y_t^2 | \Psi_{t-1})$). Since $E(\upsilon_t | \Psi_{t-1}) = 0$, the model corresponds to an AR(p) model for the squared innovations, y_t^2. Then, the process is covariance stationary if and only if the sum of the positive autoregressive parameters is less than one, ie.

$$\alpha_1 + \alpha_2 + ... + \alpha_p < 1. \tag{6}$$

Even though y_t's are serially uncorrelated they are clearly not independent through time. In accordance with the stylized facts for assets returns $\{y_t\}$ discussed above, there is a tendency for large (small) absolute values of the process to be followed by other large (small) values of unpredictable sign. Equation (6) is sufficient condition for strict stationary and ergodicity of y_t, see Milhøj (1985). But this condition is not necessary for strict stationary. The necessary and sufficient condition for the strict stationary was established by Bougerol and Picard (1992) in terms of the top Lyapunov exponent.

2.1 Population Monte Carlo

Method of importance sampling (IS) is based on an importance function. This algorithm uses an approximation to the posterior density by generating random draws from a function called the importance function. It would be more accurate to call it "weighted sampling". The method of IS is used to evaluate

$$E_f[h(X)] = \int_X h(x)f(x)dx, \tag{7}$$

based on generating a sample $X_1, X_2, ..., X_m$ from a given distribution g and approximating

$$E_f[h(X)] \approx \frac{1}{m}\sum_{j=1}^{m}\frac{f(X_j)}{g(X_j)}h(X_j). \tag{8}$$

This method is based on the alternative representation,

$$E_f[h(X)] = \int_X h(x)\frac{f(x)}{g(x)}g(x)dx. \tag{9}$$

The estimator (11) converges to (10) for the same reason that the regular Monte Carlo estimator does, whatever the choice of the distribution, (as long as $Supp(g) \supset Supp(f)$). The crux of the method is to find a good importance function. An alternative to (11) which addresses the finite variance issue, and generally yields a more stable

estimator, is to use

$$\frac{\sum_{j=1}^{m} \frac{h(x_j)f(x_j)}{g(x_j)}}{\sum_{j=1}^{m} \frac{f(x_j)}{g(x_j)}}. \tag{10}$$

By Strong Law of Large Numbers, this estimator also converge to $E_f[h(X)]$, Robert and Casella (2004).
The PMC algorithm can be described in a very general framework: it is indeed possible to consider different proposal distributions at each iteration and for each particle with this algorithm. That is, the $Z_i^{(t)}$'s can be simulated from distributions g_{it} that may depend on past samples, $Z_i^{(t)} \sim g_{it}(z)$, independently of each other (conditional on the past samples). Thus, each simulated point $Z_i^{(t)}$ is allocated an importance weight

$$\rho_i^{(t)} = \frac{\pi(z_i^{(t)})}{g_{it}(z_i^{(t)})}, \qquad i = 1, 2, ..., n,$$

and approximation of the form

$$I_t = \frac{1}{n} \sum_{i=1}^{n} \rho_i^{(t)} h(z_i^{(t)})$$

are then unbiased estimators of $E^{\pi}[h(Z)] = \int_Z h(z)\pi(z)dz$, even when the importance distribution g_{it} depends on the entire past of the experiment. We can propose the following algorithm, which is validated by the same principles as regular importance sampling.

Algorithm: Population Monte Carlo

For t=1,2,...,T
1.For i=1,2,...,n,
(i) Select the generating distribution $g_{it}(\cdot)$.
(ii) Generate $\tilde{Z}_i^{(t)} \sim g_{it}(z)$.
(iii) Compute $\rho_i^{(t)} = \frac{\pi(\tilde{z}_i^{(t)})}{g_{it}(\tilde{z}_i^{(t)})}$.
2.Normalize the $\rho_i^{(t)}$'s with replacement, using the weights $\rho_i^{(t)}$, to create the sample $(Z_1^{(t)}, Z_2^{(t)}, ..., Z_n^{(t)})$.

3 Model Choice

It is often difficult to have even a vague idea about the order of the model that generated the data. Assume that we have a countable set M of competing models for a given set of data $\mathbf{y}$. Let model $m \in M$ have a vector $\theta_m \in \Theta_m$ of unknown parameters, the dimension of which may vary from model to model. The posterior probability of model m is given by

$$\pi(m|\mathbf{y}) = \frac{\pi(m) \int_{\Theta_m} \pi(\mathbf{y}|m, \theta_m)\pi(\theta_m|m)d\theta_m}{\sum_{m\in M} \pi(m) \int_{\Theta_m} \pi(\mathbf{y}|m, \theta_m)\pi(\theta_m|m)d\theta_m} \tag{11}$$

where $\pi(\mathbf{y}|m, \theta_m)$ is the likelihood given the model m, the parameter vector θ_m, $\pi(m)$ is the prior probability for model m, and $\pi(\theta_m|m)$ is the prior of the parameter vector θ_m given the model m. Inferences about the model selection problem may be done using the Bayes factor (BF) of model m_i against model m_j given by

$$\begin{aligned} BF &= \frac{\pi(m_i|\mathbf{y})\pi(m_j)}{\pi(m_j|\mathbf{y})\pi(m_i)} \\ &= \frac{\int_{\Theta_{m_i}} \pi(\mathbf{y}|m_i, \theta_{m_i})\pi(\theta_{m_i}|m_i)d\theta_{m_i}}{\int_{\Theta_{m_j}} \pi(\mathbf{y}|m_j, \theta_{m_j})\pi(\theta_{m_j}|m_j)d\theta_{m_j}}. \end{aligned} \tag{12}$$

Kass and Raftery (1995) gave a series of arguments that make Bayes factors appealing when compared with other model selection strategies such as the Akaike information criterion (AIC) or the Bayes information creterion (BIC). Bayes factors, however, require evalution

of the integrals in the numerator and denominator of (9), which are the marginal densities $\pi(\mathbf{y}|m_i)$ and $\pi(\mathbf{y}|m_i)$. These integrals are, in general, difficult to calculate; Kass and Raftery (1995) provided an extensive description and comparison of available numerical strategies. Green (1995) introduced a reversible-jump MCMC strategy for generating from the joint posterior $\pi(m, \theta_m|\mathbf{y})$, based on the standard Metropolis-Hastings approach. The reversible-jump MCMC was also applied by Richardson and Green (1997) for an analysis of univariate normal mixtures. During reversible-jump MCMC sampling, the constructed Markov chain moves within and between models so that the limiting proportion of visits to a given model is the required $\pi(m|\mathbf{y})$ in (8). We propose an alternative to Reversible Jump by extending the PMC method to variable dimension models. In the following subsections, we explain the PMC method and its extension.

4 Model Choice using PMC

In any bayesian model, it is necessary to elicitate the prior distributions on the parameters. In any variable dimension model, usually the priors are proposed following a natural hierarchy in which the index of the model is at the top and the other parameters have a prior conditional on the index value. We used

$$\pi(\alpha_{0:p}, p) = \pi(p)\pi(\alpha_0)\pi(\alpha_{1:p} \mid p).$$

Generally, the most natural choice for the prior on p is a uniform distribution over $1, \ldots, p_{\max}$, where $p_{\max}$ is an arbitrary integer. It is however possible to think about other possibilities. For example, for the problem at hand, we can propose a multinomial on $1, \ldots, p_{\max}$ with probabilities proportional to $\Gamma(p)$. It is difficult to think about a relation between p and α_0, so the prior for this parameter can be, as usual in the literature, an Inverse Gamma with parameters a and b, indipendently of p. Finally, in order to have a non-informative but flexible prior on $(\alpha_1, \ldots, \alpha_p \mid p)$, we choose a prior related to the Dirichlet distribution. While this distribution requires that $\alpha_{1:p}$ lie on

the symplex, we need that $\sum_{i=1}^{p} \alpha_i < 1$. So we can consider

$$\pi(\alpha_{1:p} \mid p, \gamma) = \int_0^1 f_{\mathcal{D}}(\alpha_1, \ldots, \alpha_p, \alpha_{p+1} \mid \gamma) d\alpha_{p+1}.$$

Setting $\gamma = \alpha_{p+1}$, we have that this prior is equal to $\Gamma(p)$ for each $\alpha_{1:p}$. This can justify the second prior on p. The posterior distribution has not a known form, and we cannot exploit calculations analytically. In this paper, we use a PMC algorithm to obtain a sample from the joint posterior distribution on $(p, \alpha_{0:p})$, in this way obtaining simultaneously the model choice and the estimation of the parameters.
A reasonable choice often made while using PMC is to mimic a Metropolis-within-Gibbs, exploiting this way the efficiency of the full conditional distributions, whereas they are available, and to use a random walk for the other moves. In case of multimodality of the posterior, this kind of algorithms can escape from local moves better than other algorithms. In this case, we have not any full conditional, so all the moves but the one on the order p are based on a random walk. In order to reach an "automatic" tuning of the scale of the random walk, we make use of a D-kernel algorithm , see Douc *et al.* (2007). For each component of the parameter vector, we will propose D different proposal distributions, and, in particular, D normals with different variances. The algorithm will learn, across iterations, about the optimal mixing proportions of proposals to be used.
The proposal for the order p is a mixture that gives an aribitrary probability ρ to $p^{(t+1)} = p^{(t)}$ and $(1-\rho)/(p_{\max} - 1)$ to each other possible order. Given the order, $\alpha_{0:p^{(t)}}$ are updated by a truncated normal centered on their last value and variances depending on the choosen kernel. In the case in which $p^{(t+1)} < p^{(t)}$, the new parameters are generated from a uniform distribution over $(\sum_{i=1}^{p^{(t-1)}} p_i, 1)$.

5 A Simulation Study

In this section, we show our method is used in practice. We simulate

$y_1, y_2, ..., y_n$ from an ARCH(3) model,

$$y_t \sim N(0, \sigma_t^2), \qquad \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 y_{t-2}^2 + \alpha_3 y_{t-3}^2.$$

To see the effect of sample size, we use two cases: length of time series equal to 500 and 1000. We ran the program size of particle equal to 20000 and learning equal to 5. The whole experiment was repeated 10 times. We chose $\alpha_0 = 0.5,\ \alpha_1 = 0.1,\ \alpha_2 = 0.3,\ \alpha_3 = 0.1,\ a = 0.01$ and $b = 0.01$ and $p_{\max} = 10$. Table 1 gives the posterior density of the order, p, and Table 2 gives Bayesian estimation of the parameters of $\alpha_0, \alpha_1, \alpha_2$ and α_3, resulting from 10 replications and for different sample sizes.

Table 1: The posterior density of p (the order of the ARCH model) for different sample sizes and for 10 replications.

p	n=500	n=1000
1	0.0000	0.00000
2	0.0116	0.00107
3	0.8782	0.98820
4	0.0874	0.00953
5	0.0202	0.00089
6	0.0011	0.00031
7	0.0005	0.00000
8	0.0010	0.00000
9	0.0000	0.00000
10	0.0000	0.00000

The mode of the posterior density of p has been used as the estimator of the order of the model. We present the results for estimation of the order of the model for different sample sizes in Table 1. As seen in Table 1, this method correctly captures the order of the model. Although the results are good for the sample size 500, but it performs much better for sample size 1000 or more. Table 2, gives the average value of the posterior means of the parameters for 10 replications. It is

observed that irrespective of sample size of the process, the estimates of the parameters are good. However, the sample size has an appreciable effect. The larger the sample size the better and more reasonable results.

Table 2: The means of posterior distribution of $\alpha_0, \alpha_1, \alpha_2$ and α_3.

True Value	n=500	n=1000
$\alpha_0 = 0.5$	$0.4783(0.0114)^a$	0.4922(0.0124)
$\alpha_1 = 0.1$	0.0264(0.0034)	0.03890(0.0038)
$\alpha_2 = 0.3$	0.3442(0.0246)	0.3273(0.0132)
$\alpha_3 = 0.1$	0.1548(0.0134)	0.1203(0.0153)

[a] the numbers inside the parantheses show the standard deviations of the mean of estimators.

6 Concluding Remarks

In this paper, we presented the iterated importance sampling for ARCH model, especially for determinate their order. We also applied this method to conduct the Bayesian inference, using samples of the parameters obtained by iterated importance sampling.
The results show that we can estimate the order and the values of the model parameters by this method very well. This method can be used for estimation of the order and parameters of an ARCH time series with length of at least 500.

Acknowledgment
The author is grateful to Professor Christian Robert and Antonio Parisi for their advices and guidance over this work.

References

[1] Bauwens, L. and Lubrano, M., 1998, "Bayesian inference on GARCH models using the Gibbs sampler, " *Econometric Journal,* **1,** 23–46.

[2] Bauwens, L., Lubrano, M. and Jean-Francois Richard, 1999, *Bayesian inference in Dynamic Econometric models,* Oxfored University Press.

[3] Bauwens, L. and Lurbano, M., 1996, Identification restrictions and posterior densities in cointegrated Gaussian VAR systems. In: Fomby, T.B. (Ed.), *Advances in Econometrics,*Vol. 11, Part B, pp. 3–28. London: JAI Press.

[4] Bollerslev, T., 1986, "Generalized autoregressive conditional heteroscedasticity", *Journal of Econometrics,* **31**, 307-327.

[5] Bollerslev, T., Chou, R.Y. and Kroner, K. F., 1992, ARCH modeling in finance: A review of the theory and empirical evidence, *Journal of Econometrics,* **52,** 5–59.

[6] Bougerol, P. and Picard, N. M., 1992 "Stationarity of GARCH processes and of some nonnegative time series", *Journal of Econometrics,* **52**, 115–127.

[7] Cappé, O., Guillin, A., Marin, J. M. and Robert, C., 2004, "Population Monte Carlo", *J. Compout. Graph. Statist.,* **13(4)**, 907–929.

[8] Celeux, G., Marin, J. M. and Robert, C. P., 2006, "Iterated importance sampling in missing data problems", *Compoutational Statistics and Data Analysis,* **50(12)**, 3386–3404.

[9] Douc, R., Guillin, A., Marin, J. M. and Robert, C. P., 2007, "Minimum variance importance sampling via population monte carlo", *ESIAM Probability and Statistics,* (to appear).

[10] Engle, R. F., 1982, "Autoregressive conditional heteroskedasticity with estimation of the variance of the United Kingdom inflation," *Econometrica,* **50**, 987–1007.

[11] Geweke, J., 1989, "Exact predictive densities in linear models with ARCH disturbances," *Journal of Econometrics,* **40**, 63–86.

[12] Geweke, J., 1994, "Bayesian comparison of econometric models," *Working paper 532,* Research Department, Federal Reseve Bank of Minneapolis.

[13] Geweke, J., 1988, "Antithetic acceleration of Monte Carlo integration in Bayesian inference," *Journal of Econometrics,* **38,** 73–89.

[14] Green, P. J. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination", *Biometrika,* **82,** 711732.

[15] Kass, R. E., and Raftery, A. E. 1995, "Bayes Factors", *Journal of the American Statistical Association,* **90,** 773795.

[16] Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H. and Teller E. 1953, "Equations of state calculations by fast computing machines", *Journal of Chemical Physics*, 1087-1092.

[17] Milhøj, A., 1985, "The moment structure of ARCH processes," *Scandinavian Journal of statistics,* **12,** 281–334.

[18] Richardson, S., and Green, P. J., 1997, "Bayesian Analysis of Mixtures With an Unknown Number of Components", *Journal of the Royal Statistical Society, Ser. B,* **59,** 731758.

[19] Robert, C.P., 2001, *The Bayesian Choice,* (Springer-Verlag,) Second edition.

[20] Robert, C.P. and Casella, G., 2004, *Monte Carlo Statistical Methods,* Springer-Verlag, New York, Secon edition.

[21] Vrontos, I. D., Dellaportas, P. and Politis, D. N., 2000, "Full Bayesian inference for GARCH and EGARCH models," *Journal of Business and Economics statistics,* **18**, 187–198.

A.P.O. Rules are Asymptotically Non-Deficient for Estimation under Linex Loss Function

Eisa Mahmoudi

Department of Statistics, Yazd University

Abstract: The problem considered is sequential estimation of the mean of a family of transformed chi-square distributions with linex loss function for estimation error and a cost $c > 0$ for each of an i.i.d. sequence of potential observations $X_1, X_2, \cdots$. A Bayesian approach is adopted, and natural conjugate prior distributions are assumed. For this problem, the asymptotically pointwise optimal (A.P.O.) procedure continues sampling until the posterior variance of the mean of this family of distributions is less than $c(n_0+n)$, where n is the sample size and n_0 is the fictitous sample size implicit in the conjugate prior distribution. Here the A.P.O. rule is shown to be asymptotically non-deficient, under strong regularity conditions: that is, the difference between the Bayes risk of the A.P.O. rule and the Bayes risk of the optimal procedure is of smaller order of magnitude than c, the cost of a single observation, as $c \longrightarrow 0$.

Keywords: A.P.O. rule, Bayes sequential estimation, Linex loss function, Non-deficiency, Transformed chi-square distributions.

1 Introduction

In some estimation problems, use of symmetric loss functions may be inappropriate, as has been recognized in the literature, see, for example, Ferguson (1967), Zellner and Geisel (1968), Aitchison and Dunsmore (1975), Varian (1975), and Berger (1980). That is, a given positive error may be more serious than a given negative error of the same magnitude or vice versa. The authors mentioned above, expect for Varian, have considered asymmetric linear loss functions. Varian introduced a very useful asymmetric linex loss function that rises approximately exponentially on one side of zero and approximately linearly on the other side in his applied study of real estate assessment. Underassessment results in an approximately linear loss of revenue

whereas overassessment often results in appeals with attendant, substantial litigation and other cost, see Varian (1975) for details. Also, in dam construction an underestimate of the peak water level is usually much more serious than an overestimate. It is clear that finding Bayes stopping rules is a formidable task expect in those special case where the Bayes rules reduce to a fixed-sample rules. For example Chattopadhyay (1998,2000), considered the sequential estimation using an asymmetric loss function and Takada (2000) treated the sequential estimation of normal mean under linex loss function. This paper considers Bayes sequential estimation in a family of transformed chi-square distributions under a gamma prior distribution.

Suppose, on the basis of the sample with size n and fixed cost $c > 0$ for each unit, the posterior risk is given by

$$L_n(c) = Y_n + cn, \tag{1}$$

where $\{Y_n; n \geq 1\}$ is a sequence of random variables defined on a probability space (Ω, Λ, P), where Y_n is $\Lambda - measurable$, $P(Y_n > 0) = 1$ and $Y_n \xrightarrow{a.s} 0$ as $n \longrightarrow \infty$.

Definition 1 *The stopping rule $\tau = \tau_c$ is called a Bayes stopping rule if*

$$E(L_\tau(c)) = \min_S E(L_S(c)),$$

where S is any stopping rule.

Bickel and Yahav (1967, 1968) proposed asymptotically pointwise optimal (A.P.O.) rule to the minimization problem. A more general property of A.P.O. rules was proved by Woodroofe (1981). He define asymptotic non-deficiency of A.P.O. rule in sense that

$$E(L_T(c)) - \rho(c) = o(c) \quad \text{as} \quad c \longrightarrow 0, \tag{2}$$

where $\rho(c) = E(L_\tau(c))$, is an A.P.O. rule. More details can be seen in Ghosh et al. (1997).

Varian (1975) proposed a useful asymmetric loss function, known as linex loss function which may be given by

$$L(\theta, \hat{\theta}) = \exp(a(\hat{\theta} - \theta)) - a(\hat{\theta} - \theta) - 1, \tag{3}$$

where $a \neq 0$. This loss function is useful when an overestimation is more serious than an underestimation ($a > 0$), or vice-versa ($a < 0$). Takada (2001) studied Bayes sequential estimation of the poisson mean under linex loss function. He shows that an A.P.O. rule is asymptotically non-deficient in the sense of Woodroofe (1981).

In this paper, Bayes sequential estimation in a subclass of the exponential family, that is called the family of transformed chi-square distributions, has been obtained. The plan of the article is as follows. In Section 2, the family of transformed chi-square distributions is defined. In Section 3, the A.P.O. rule is presented. Section 4 is devoted to proving asymptotic non-deficiency of the A.P.O.

2 A subclass of the scale parameter exponential family

Let $X_1, X_2, ...$ be a sequence of independent and identically distributed random variable from a one-parameter exponential family

$$f(x; \eta) = e^{a(x)b(\eta)+c(\eta)+h(x)}. \tag{4}$$

Theorem 1 *(Rahman and Gupta (1993)) In a family (4), the function $-2a(X)b(\eta)$ has Gamma$(\frac{j}{2}, 2)$ distribution if and only if*

$$\frac{2\acute{c}(\eta)b(\eta)}{b'(\eta)} = j, \tag{5}$$

where j is positive and free from η. In case j is an integer, $-2a(X)b(\eta)$ follows a central chi-square distribution with j degrees of freedom.

Definition 2 *The one-parameter exponential family of form (4) satisfying (5) is called the family of transformed chi-square distribution, provided j is a positive integer.*

For example the Gamma$(\alpha, 1/\lambda)$ distribution with known α belongs to this family with

$$a(X) = X, \quad b(\lambda) = -\lambda, \quad c(\lambda) = \alpha \ln \lambda, \quad -2a(X)b(\lambda) = 2\lambda X, \quad j = 2\alpha,$$

and the Lognormal$(0, \sigma^2)$-distribution belongs to this family with

$$a(X) = \frac{1}{2}(\ln X)^2, \quad b(\sigma) = -\frac{1}{\sigma^2}, \quad c(\sigma) = -\ln \sigma,$$
$$-2a(X)b(\sigma) = \frac{(\ln X)^2}{\sigma^2}, \quad j = 1.$$

Using the condition (5), the family of distribution (4) reduces to

$$f(x;\theta) = c(x)\theta^v e^{-\theta a(x)}, \tag{6}$$

where $c(x) = e^{h(x)+k_1}$, $\theta = -b(\eta) > 0$ and $v = \frac{j}{2} > 0$. Also note that $-2a(X)b(\eta) = 2\theta a(X)$ has Gamma$(v, \frac{1}{2})$-distribution or $a(X)$Gamma$(v, \frac{1}{\theta})$. Now, if $X_1, X_2, ..., X_n$ be a sample of size n from distribution (6), then the joint density of $X_1, X_2, ..., X_n$ is given by

$$f(\mathbf{x};\theta) = c(\mathbf{x}, n)\theta^{nv} e^{-\theta \sum_{i=1}^n a(x_i)}, \ \theta > 0 \tag{7}$$

where $c(\mathbf{x}, n) = \prod_{i=1}^n c(x_i)$, $S_n(\mathbf{X}) = \sum_{i=1}^n a(X_i)$Gamma$(nv, \frac{1}{\theta})$.

3 A.P.O. rule

Suppose $\{Y_n; n \geq 1\}$ be a sequence of random variables on probability space (Ω, Λ, P) such that $p(Y_n > 0) = 1$ and $Y_n \xrightarrow{a.s} 0$ as $n \to \infty$. Define

$$L_n(c) = Y_n + nc \qquad \text{for} \quad c > 0.$$

Then a stopping rule $T = T_c$ is said to be P.O. if $P\left(\frac{L_T(c)}{L_S(c)} \leq 1\right) = 1$,for any other stopping rule S. Bickel and Yahav (1967) named the stopping rule $T = T_c$ as A.P.O. if

$$\lim_{c \to 0} \frac{L_T(c)}{L_S(c)} \leq 1, \quad \text{a.s.}$$

for any other stopping rule S. Obviously, T is an A.P.O. rule iff

$$\lim_{c\to 0}\frac{L_T(c)}{L_\tau(c)}=1, \quad \text{a.s.}$$

where $L_\tau(c)=\min_S L_S(c)$ and τ is the Bayes stopping rule. Also, a stopping rule $T=T_c$ is said to be asymptotic optimal (A.O.) if $\lim_{c\to 0}\frac{E(L_T(c))}{E(L_S(c))}\leq 1$.

Theorem 2 *(Bickel and Yahav 1967) If $P(Y_n>0)=1$ and for some β (>0), we have $n^\beta Y_n \xrightarrow{a.s} V$ as $n\to\infty$, where V is positive random variable, then*

$$T=T_c=\inf\left\{n\geq 1; \quad \frac{Y_n}{n}\leq\frac{c}{\beta}\right\}$$

is A.P.O. Further, if $\sup_{n\geq 1}E(n^\beta Y_n)<\infty$, then the stopping rule T_c is A.O.

We study the above definitions and Theorems based on the family of transformed chi-square distribution. The posterior distribution of θ given $\mathbf{x}$, under the conjugate prior Gamma(α,β) distribution is

$$\pi(\theta|\mathbf{x})=\frac{\left(\frac{1+\beta S_n}{\beta}\right)^{(nv+\alpha)}}{\Gamma(nv+\alpha)}\theta^{(nv+\alpha)-1}e^{-\theta\left(\frac{1+\beta S_n}{\beta}\right)} \qquad \theta>0.$$

The Bayes estimation of θ under prior π and sample vector $\mathbf{x}=(x_1,...,x_n)$ is

$$\delta^\pi_{BL}=-\frac{1}{a}\log E(e^{-a\theta}|\mathbf{X}). \tag{8}$$

Therfore, using (3), the Bayes estimator of θ is $\delta^\pi_{BL}=\frac{(nv+\alpha)}{a}\log(1+\frac{a\beta}{1+\beta S_n})$, where $S_n=\sum_{i=1}^n a(X_i)$. From the definition of $L_n(c)$, we have

$$\begin{aligned} Y_n &= E\left(L\left(\theta,\delta^\pi_{BL}\right)|\mathbf{X}\right)=-aE\left(\delta^\pi_{BL}-\theta|\mathbf{X}\right)\\ &= \left\{\log E\left(e^{-a\theta}|\mathbf{X}\right)+aE\left(\theta|\mathbf{X}\right)\right\}, \end{aligned}$$

and hence

$$L_n(c) = -(nv+\alpha)\log(1+\frac{a\beta}{1+\beta S_n}) + a(nv+\alpha)\frac{\beta}{1+\beta S_n} + cn = Y_n + cn,$$

where

$$Y_n = -(nv+\alpha)\log(1+\frac{a\beta}{1+\beta S_n}) + a(nv+\alpha)\frac{\beta}{1+\beta S_n}.$$

Obviously, $P(Y_n > 0) = 1$. Using the extension $\log(1+x) = \sum_{n=0}^{\infty}(-1)^n \frac{x^{n+1}}{n+1}$, we obtain

$$nY_n = \frac{a^2}{2\left(\frac{1}{n\beta}+\frac{S_n}{n}\right)^2} \times \left[\left(v+\frac{\alpha}{n}\right) - \left(\frac{v}{n}+\frac{\alpha}{n^2}\right)\frac{a}{\frac{3}{2}\left(\frac{1}{n\beta}+\frac{S_n}{n}\right)} + \left(\frac{v}{n^2}+\frac{\alpha}{n^3}\right)\frac{a^2}{2\left(\frac{1}{n\beta}+\frac{S_n}{n}\right)^2} - ...\right] = \frac{a^2}{2\left(\frac{1}{n\beta}+\frac{S_n}{n}\right)^2}\{v+o(1)\} = \frac{a^2 v}{2\left(\frac{1}{n\beta}+\frac{S_n}{n}\right)^2}.$$

By using the SLLN, $\frac{1}{n}S_n \xrightarrow{a.s} E(a(X_i)) = \frac{v}{\theta}$ and hence $\left(\frac{1}{n\beta}+\frac{S_n}{n}\right)^2 \xrightarrow{a.s} \frac{v^2}{\theta^2}$, using this facts we have

$$nY_n \xrightarrow{a.s} \frac{a^2\theta^2}{2v} \quad \text{as} \quad n \longrightarrow \infty.$$

Therefore, if we suppose $V = \frac{a^2\theta^2}{2v}$, from Theorem 3.1 the stopping rule

$$T_c = \inf\left\{n \geq 1; \quad \frac{Y_n}{n} \leq c\right\} \tag{9}$$

is an A.P.O. rule. But since

$$E(nY_n) = \frac{a^2}{2v}E(\theta^2|\mathbf{X}) = \frac{a^2}{2v}\alpha\beta^2(\alpha+1) < \infty, \qquad \forall \ \ n = 1, 2, ...$$

then $\sup_{n\geq 1} E(nY_n) < \infty$, hence Theorem 3.1 implies that an A.P.O. rule $T = T_c$ is A.O., so that

$$\lim_{c\longrightarrow 0}\frac{E(L_T(c))}{E(L_\tau(c))} = 1. \tag{10}$$

Example 1 *Suppose $X_1, ..., X_n$ be a random sample from Gamma(α, β) distribution (α known), with $a(X) = X$, $\theta = \frac{1}{\beta}$ and $v = \alpha$. Therefore, $nY_n \xrightarrow{a.s} \frac{a^2}{2\alpha\beta^2}$ as $n \longrightarrow \infty$ and*

$$T_c = \inf\left\{ n \geq 1; \quad \frac{a^2}{2n^2\alpha\overset{\wedge}{\beta}^2} \leq c \right\} = \inf\left\{ n \geq 1; \quad n \geq (2c)^{-\frac{1}{2}} \frac{\sqrt{\alpha a^2}}{\bar{X}_n} \right\}.$$

4 Asymptotic non-deficiency

In this section we show the asymptotic non-deficiency of the A.P.O. rule $T = T_c$ in (9). First we consider the asymptotic expansion of $E(L_T(c))$ as $c \longrightarrow 0$. From (1) we have

$$L_n(c) = \frac{Z_n}{n} + cn = 2\sqrt{c}\sqrt{Z_n} + n^{-1}\left(\sqrt{Z_n} - n\sqrt{c}\right)^2, \tag{11}$$

where $\{Z_n; \quad n \geq 1\}$ is a sequence of random variables, such that $Z_n = nY_n$.
If we suppose that $U_n = E(\theta^2|\mathbf{X})$ then

$$Z_n = \frac{a^2}{2v} U_n. \tag{12}$$

Suppose that $V_n = E(\theta|\mathbf{X})$ and $W_n = U_n - V_n^2$, hence

$$\sqrt{U_n} = V_n + \frac{W_n}{\sqrt{U_n} + V_n}. \tag{13}$$

It follows from (11), (12) and (13) that for any stopping rule $S = S_c$

$$\begin{aligned} E(L_S(c) &= 2\sqrt{c}E(\sqrt{Z_S}) + E\left\{ S^{-1}\left(\sqrt{Z_S} - S\sqrt{c}\right)^2 \right\} \\ &= 2\sqrt{a^2c}\frac{1}{\sqrt{2v}} E\left\{ V_S + \frac{W_S}{\sqrt{U_S} + V_S} \right\} + E\left\{ S^{-1}\left(\sqrt{Z_S} - S\sqrt{c}\right)^2 \right\} \\ &= \sqrt{2a^2c}\frac{1}{\sqrt{v}} E(V_S) + \frac{\sqrt{2a^2c}}{\sqrt{v}} E\left\{ \frac{W_S}{\sqrt{U_S} + V_S} \right\} \\ &+ E\left\{ S^{-1}\left(\sqrt{Z_S} - S\sqrt{c}\right)^2 \right\}, \end{aligned} \tag{14}$$

and obviously

$$E(V_S) = E(\theta) = \alpha\beta. \tag{15}$$

Lemma 1 *Suppose that $\alpha > 2$. If $S = S_c$ be any stopping rule such that there exists a constant k, with*

$$\sqrt{c}S \geq k\sqrt{Z_S}, \qquad \textit{for small } \ c \tag{16}$$

and

$$\sqrt{c}S \xrightarrow{p} \sqrt{V} \quad \textit{as } c \longrightarrow 0, \tag{17}$$

then

$$\lim_{c\longrightarrow 0} \frac{\sqrt{2a^2}}{\sqrt{cv}} E\left\{\frac{W_S}{\sqrt{U_S}+V_S}\right\} = \frac{1}{4\beta^2(\alpha-1)(\alpha-2)}. \tag{18}$$

Lemma 2 *If $T = T_c$ be the A.P.O. rule then*

$$E\left\{T^{-1}\left(\sqrt{Z_T} - T\sqrt{c}\right)^2\right\} = o(c) \qquad \textit{as } \ c \longrightarrow 0.$$

Proof: Using Lemma 4 of Woodroofe (1981) it's enough to show that

$$E\left(\sup_{n\geq 1} Y_n^*\right) < \infty \qquad \text{and} \qquad Y_n^* \xrightarrow{a.s} 0 \ \text{ as } \ n \longrightarrow \infty,$$

where

$$Y_n^* = \begin{cases} n\left(\sqrt{\frac{Z_{n-1}}{Z_n}} - 1\right)^2 & \sqrt{\frac{Z_{n-1}}{Z_n}} \geq 1 \\ 0 & \sqrt{\frac{Z_{n-1}}{Z_n}} < 1 \end{cases}$$

but we have

$$\frac{Z_{n-1}}{Z_n} = \frac{(n-1)^2 v + (n-1)\alpha}{n^2 v + n\alpha}\left(\frac{1+\beta S_n}{1+\beta S_{n-1}}\right)^2 \frac{1+o(1)}{1+o(1)},$$

hence

$$\begin{aligned} \sqrt{\tfrac{Z_{n-1}}{Z_n}} \leq \left(\tfrac{1+\beta S_n}{1+\beta S_{n-1}}\right)(1+o(1)) \ &= \tfrac{(1+\beta S_{n-1}+\beta a(X_n))}{(1+\beta S_{n-1})}(1+o(1)) \\ &= 1 + \tfrac{a(X_n)}{(\frac{1}{\beta}+S_{n-1})}(1+o(1)), \end{aligned}$$

therefore

$$\sqrt{\frac{Z_{n-1}}{Z_n}} - 1 \leq \frac{a(X_n)}{(\frac{1}{\beta} + S_{n-1})} (1 + o(1))$$

thus, we will have

$$n \left(\sqrt{\frac{Z_{n-1}}{Z_n}} - 1 \right)^2 \leq n \left(\frac{a(X_n)}{(\frac{1}{\beta} + S_{n-1})} \right)^2 (1 + o(1)) .$$

Since $n \left(\frac{a(X_n)}{(\frac{1}{\beta}+S_{n-1})} \right)^2 \xrightarrow{a.s} 0$ as $n \longrightarrow \infty$, therefore $Y_n^* \xrightarrow{a.s} 0$ as $n \longrightarrow \infty$ and $E\left(\sup_{n\geq 1} Y_n^*\right) < \infty$.

Theorem 3 *Suppose that $\alpha > 2$. If $T = T_c$ be the A.P.O. rule then*

$$E\left(L_T(c)\right) = \frac{\sqrt{2a^2}\alpha\beta}{\sqrt{v}}\sqrt{c} + \frac{1}{4\beta^2\left(\alpha-1\right)\left(\alpha-2\right)}c + o(c) \quad as \quad c \longrightarrow 0. \tag{19}$$

Proof: From the Theorem 2.1 of Bickel and Yahav (1967)

$$\sqrt{c}T_c \xrightarrow{a.s} \sqrt{V} \quad \text{as} \quad c \longrightarrow 0,$$

hence from (9) the conditions (16) and (17) in Lemma 4.1 are satisfied. By substituting the results of Lemma 4.1 and 4.2 with (15) into (14) the relation (20) is satisfied.

Example 2 *In Example 3.2. with a random sample from Gamma(α, β) distribution (α known), we have*

$$E\left(L_T(c)\right) = \sqrt{2\alpha a^2}\beta\sqrt{c} + \frac{1}{4\beta^2\left(\alpha-1\right)\left(\alpha-2\right)}c + o(c) \quad as \quad c \longrightarrow 0,$$

where

$$T_c = \inf\left\{ n \geq 1; \quad n \geq (2c)^{-\frac{1}{2}} \frac{\sqrt{\alpha a^2}}{\bar{X_n}} \right\}.$$

Lemma 3 *The Bayes stopping rule $\tau = \tau_c$ satisfies (16) and (17).*

Proof: Since τ is the Bayes stopping rule on the set $\{\tau = n\}$ we must have

$$E\left(L_{n+1}(c)|\mathbf{X}\right) \geq L_n(c).$$

But $L_n(c) = \frac{Z_n}{n} + cn$, where $Z_n = \frac{a^2 v}{2\left(\frac{1}{n\beta} + \frac{S_n}{n}\right)^2} = \frac{a^2 n^2 \beta^2 v}{2(1+\beta S_n)^2}\left(1 + o(1)\right) = \frac{k' n^2}{(1+\beta S_n)^2}$, then we have

$$\frac{Z_n}{n} + cn \leq E\left\{\left[\frac{k^*(n+1)}{\left((1+\beta S_{n+1})^2\right)} + c(n+1)\right]|\mathbf{X}\right\},$$

or

$$\begin{aligned}\tfrac{Z_n}{n} &\leq k^*(n+1)E\left\{\left[\tfrac{1}{(1+\beta S_n+\beta a(X_{n+1}))^2}\right]|\mathbf{X}\right\} + c \\ &\leq k^*(n+1)E\left\{\left[\tfrac{1}{(1+\beta S_n)^2}\right]|\mathbf{X}\right\} + c,\end{aligned}$$

but

$$\frac{1}{(1+\beta S_n)^2} = \frac{k' n^2}{Z_n},$$

and hence

$$\frac{Z_n}{n} \leq \frac{k^*(n+1)}{k' n^2} Z_n.$$

Therefore we have

$$Z_n\left(n - \frac{k^*(n+1)}{k'}\right) \leq cn^2.$$

It is easy to show that $D_n = n - \frac{k^*(n+1)}{k'} = 1 + o(1)$ as $n \longrightarrow \infty$. Therefore (16) for $\tau = \tau_c$ is satisfied. Note that

$$E\left\{\frac{Z_\tau}{\sqrt{c}\tau} + \sqrt{c}\tau - 2\sqrt{Z_\tau}\right\} = \frac{\rho(c)}{\sqrt{c}} - 2E\left(\sqrt{Z\tau}\right), \qquad (20)$$

where $\rho(c) = E\left(L_\tau(c)\right)$.

It follows from (9) and (20) that

$$\frac{\rho(c)}{\sqrt{c}} \xrightarrow{p} \frac{\sqrt{2a^2}\alpha\beta}{\sqrt{v}} \quad \text{as} \quad c \longrightarrow 0. \tag{21}$$

Since

$$\sqrt{Z_\tau} \xrightarrow{a.s} \sqrt{V} \quad \text{as} \quad c \longrightarrow 0, \tag{22}$$

and $\sqrt{Z_\tau}$ is uniformly integrable then

$$2E(\sqrt{Z_\tau}) \xrightarrow{a.e} 2E(\sqrt{V}) = \frac{\sqrt{2a^2}\alpha\beta}{\sqrt{v}} \quad \text{as} \quad c \longrightarrow 0. \tag{23}$$

Substituting (22) and (23) into (21) we have

$$E\left\{\frac{Z_\tau}{\sqrt{c\tau}} + \sqrt{c\tau} - 2\sqrt{Z_\tau}\right\} \longrightarrow 0 \quad \text{as} \quad c \longrightarrow 0, \tag{24}$$

which means that

$$\frac{Z_\tau}{\sqrt{c\tau}} + \sqrt{c\tau} - 2\sqrt{Z_\tau} \xrightarrow{p} 0 \quad \text{as} \quad c \longrightarrow 0.$$

Since the inside of relation (25) is non-negative, using (23) we have $\sqrt{c\tau} \xrightarrow{p} \sqrt{V}$ as $c \longrightarrow 0$, and hence the relation (15) for $\{\tau = \tau_n\}$ is satisfied.

Theorem 4 *Suppose that $\alpha > 2$. If $T = T_c$ be the A.P.O. rule, then*

$$E\left(L_T(c)\right) - \rho(c) = o(c) \quad \textit{as} \quad c \longrightarrow 0.$$

Proof: Let $\tau = \tau_c$ be the Bayes stopping rule. Then

$$\rho(c) \geq \frac{\sqrt{2a^2}}{\sqrt{v}}\sqrt{c}E\left(V_\tau\right) + \frac{\sqrt{2a^2}}{\sqrt{v}}\sqrt{c}E\left\{\frac{W_\tau}{\sqrt{U\tau} + V_\tau}\right\}.$$

On the basis of Lemma 4.1 and 4.3 with (15) we will have

$$\rho(c) \geq \frac{\sqrt{2a^2}}{\sqrt{v}}\sqrt{c}\alpha\beta + \frac{\sqrt{2a^2}}{\sqrt{v}}c\frac{1}{4\beta^2\left(\alpha-1\right)\left(\alpha-2\right)},$$

i.e.

$$\rho(c) \geq E\left(L_T(c)\right) - o(c),$$

or

$$E\left(L_T(c)\right) - \rho(c) \leq o(c).$$

On the other hand, $E\left(L_T(c)\right) - \rho(c) \geq o(c)$. Therefore,

$$E\left(L_T(c)\right) - \rho(c) = o(c) \qquad \text{as} \quad c \longrightarrow 0,$$

and the proof is completed.

References

[1] Bickel, E., Yahav, J. (1967). Asymptotically pointwise optimal procedures in sequential analysis. *Proc. Fifth Berkeley sympos. Math. Statist. Prob.*, 1, 401-413.

[2] Bickel, E., Yahav, J. (1968). Asymptotically optimal Bayes and minimax procedures in sequential estimation. *Ann. Math. Statist.*, 39, 442-456.

[3] Chattopadhyay, S. (1998). Sequential estimation of normal mean under asymmetric loss function with a shrinkage stopping rule. *Metrika*, 48, 53-59.

[4] Chattopadhyay, S. (2000). Sequential estimation of exponential location parameter using an asymmetric loss function. *Commu. Statist. Theory and Methods*, 29(4), 783-795.

[5] Ghosh, M., Mukhopadhyay, N., Sen, P.K. (1997). *Sequential estimation.* John Wiley and Sons, 1ne., New York.

[6] Rahman, M.S., Gupta, R.P. (1993). Family of transformed chi-square distributions. *Commu. Statist. Theory and Methods*, 22, 135-146.

[7] Takada, Y. (2000). Sequential point estimation of normal mean under Linex loss function. *Metrika*, 52, 163-171.

[8] Takada, Y. (2001). Bayes sequential estimation of poisson mean under a Linex loss function. *Sequential Analysis*, 20(1), 55-64.

[9] Varian, H.R. (1975). A Bayesian approach to real estate assessment. *In: Fienberg S.E. Zellner A.(eds) Studies in Bayesian econometric and statistics in Honor of Leonard J. Savage, North Holland, Amesterdam*, 195-208.

[10] Woodroofe, M. (1981). A.P.O. rules are asymptotically non-deficient for estimation with squared error loss. *Z. Wahr. Ver. und Gebiete*, 58, 331-341.

Large Efficiency Issues in Generalized Estimating Equations for Randomized Clinical Trials

M. Mansourian , I. Kazemi, B. Soleimani

Department of Biostatistics, Tarbiat Modares University
Department of Statistics, University of Isfahan
Department of Biostatistics, Isfahan University of Medical Sciences

Abstract: In this paper we consider the generalized estimating equations (GEEs) approach in fitting models for randomized clinical trials. This estimation approach, which is popular in handling correlated response data, uses some working correlation structure to result consistent and asymptotically normal estimates of the regression coefficient. We describe an application of using the GEEs with a clinical trial data set involving categorical response variable, to help end-users appreciate the substance of the approach. In this study, it is of interest to relate the conversions of DMFS index in the three settings (fluoride gum, fluoride mouth rinse wash and a control group) for both males and females characteristics. We also set the baseline DMFS measure as a covariate. The incidence rates of carries are compared by Poisson regression models which use a log-link function to estimate the odds of achieving alleviation for comparing explanatory factors including of treatments. We use the quasi-likelihood criterion to select the optimum correlation matrix form. Results show that three groups reach statistical significance and that there is significant differences between times follow-up and interaction between time and treatments, but there is no difference between genders. Furthermore, a study of the contrast between the groups, known as the effect size, was carried out, although it was not prespecified in the trial conventions. According to one documental threshold, the odds ratio was calculated. In all follow-up sessions the odds ratio was lower than one, which indicated that Fluoride gum showed greater efficacy than Fluoride rinse.

Keywords: Longitudinal data, Poisson regression model, working correlation structure, correlated measurements.

1 Introduction

Longitudinal data present statistical problems of interest in clinical trials and epidemiologic studies. In analyzing these data, the usual model assumptions may not be valid and thus advanced estimation methods are needed. This problem arises, particularly, when different count measurements are taken from the same patient, after receiving several treatments. In fitting the repeated measurement outcomes, traditional statistical analyses have often concentrated on univariate approaches. Examples include; the mean of response, the last on-treatment observation, or the analysis of variance which are inadequate in many models fitting. These analyses, while simple to perform and interpret, do not consider the effects of time events (Diggle et al., 2002). They also ignore incorporating the correlation structure of responses. This leads to incorrect fitting of the regression model and to decreasing in the efficiency of parameter estimates (Fitzmaurice and Laird, 1995; Zorn, 2001).

Currently, advanced estimation approaches, which use the whole dataset in the estimation process, provide more comprehensive tools in fitting many complex models. These approaches have the potential to increase the efficiency of the statistical analyses for many dependent responses including binary and categorical longitudinal data. One of these approaches, the generalized estimating equations (GEEs) introduced by Liang and Zeger (1986) facilitates analysis of data collected in longitudinal or repeated measures designs. The GEE permits specification of a working correlation matrix that accounts for the form of within-subject correlation of many different responses. Thus, GEE estimates more efficient and unbiased regression parameters (Emond et al., 1997). The GEE has received extensive applications in medical and life sciences, such as epidemiology, biology and medical sciences (see e.g., Ballinger, 2004). An extensive literature on the properties and the limitations of using GEE is given by Hardin and Hilbe (2003).

In this paper, we briefly review GEE methodology, and then consider a dental clinical trial. The outcome measurements are taken for each subject at seven follow-up times. The primary of interest is to illustrate the existence of dependence between outcomes on several

covariates. The trial was a single blind, randomized, controlled, two-year caries clinical trial. In order to obtain a 95% confidence interval, a 65% difference between the groups, with an expected range of 2.23-7.2 local DMFS (Decay, Missing and Filling Surface) data, and a minimum sample size of 22 subjects in each group was required. Allowing for an expected annual dropped-out rate of 10% (23.5% increase in sample size for a 2-year follow-up), a minimal recruitment was set at 28 subjects per group (84 in total). Volunteers were recruited in Isfahan-Iran ranging from 13-15 year-old students. With respect to the high caries rates, this region is suitable for caries clinical trials. A randomization list was drawn up and the enrolled participants were randomly assigned to the three comparison groups (fluoride gum, fluoride mouth rinse, and a control group). In this study, of interest was to relate the conversions of DMFS index in the three setting groups for both males and females. We also set the baseline DMFS measure as a covariate. The incidence rates of caries were computed by fitting a Poisson regression model.

The organization of paper is as follows. In Section 2, we briefly review GEE methodology and highlight some limitations of using this approach. Section 3 introduces the longitudinal dataset in details. In Section 4, we propose a Poisson regression model for the dental clinical trial data and consider parameter estimation using the GEE approach together with various working correlation matrices. Section 5 is left for recommendations.

2 The GEE Approach

The generalized estimating equations approach provides a method of fitting a marginal model in which a mean function and a covariance structure are specified, but a full likelihood is not required. Specifically, the GEE considers a known function of the marginal mean of responses as a linear function of covariates. This offers a method

to obtain consistent estimates of regression parameters together with their standard errors under suitable regularity conditions (Zeigler et al., 1998). The necessary condition includes the correct specification of the mean structure, otherwise the GEE leads to inconsistency in the model parameter estimates (Crouchley and Davies, 2001) and hence, subsequent inferences may be invalid.

Let the vector $\mathbf{Y_i} = (\mathbf{Y_{i1}}, \mathbf{Y_{i2}}, \cdots, \mathbf{Y_{i,n_i}})^{\mathbf{T}}$, for $i = 1, 2, \cdots, k$, denote the n_i repeated measurements on individual i, and $\mathbf{X_i^T} = (\mathbf{X_{i1}}, \mathbf{X_{i2}}, \cdots, \mathbf{X_{i,n_i}})$ be a $P \times n_i$ matrix of covariates, where $\mathbf{X_{it}^T} = \left(\mathbf{X_{it_1}}, \mathbf{X_{it_2}}, \cdots, \mathbf{X_{it_p}}\right)$ is the vector of covariates corresponding to Y_{it}, for $t = 1, 2, \cdots, n_i$.

Assume that $\mu_{it} = E\left(Y_{it}|\mathbf{X_{it}}\right)$ and $\mathbf{R_i}\left(\alpha\right)$ is a positive definite matrix that describes well the association structure. Suppose that the working correlation matrix $\mathbf{R_i}\left(\alpha\right)$ is fully specified by an unknown variance-components vector α of order s. The working variance- covariance matrix of $\mathbf{Y_i}$ is given by

$$\mathbf{V_i} = \mathbf{A_i^{\frac{1}{2}}} \mathbf{R_i}\left(\alpha\right) \mathbf{A_i^{\frac{1}{2}}} \phi \tag{1}$$

where $\mathbf{A_i} = \mathbf{diag}\left\{\mathbf{V}\left(\mu_{\mathbf{it}}\right)\right\}$, $V\left(\mu_{it}\right)$ is a variance function, after having adjusted for the effects of covariates and ϕ is a scale parameter. It is shown that the correct choice of $\mathbf{R_i}\left(\alpha\right)$ will lead to more efficient estimates of β (Pan and Connett, 2002).

Suppose that μ_{it} is specified correctly, the GEEs analogous to the score function are

$$S\left(\beta\right) = \sum_{i=1}^{k} \mathbf{D_i^T V_i^{-1}} \left(\mathbf{y_i} - \mu_{\mathbf{i}}\right) = \mathbf{0}, \tag{2}$$

where $\mathbf{D_i} = \mathbf{diag}\left\{\partial \mu_{\mathbf{i}} / \partial \beta\right\}$ is a $n_i \times p$ matrix of mean derivatives, $\mathbf{V_i}$ is a $n_i \times n_i$ matrix of weights in (1), and $\mathbf{y_i} - \mu_{\mathbf{i}}$ is a $n_i \times 1$ vector of centered observations. This equation is identical to the regular score function when the distribution of Y belongs to the exponential family and the observations for each subject are independent. The statement

$\mathbf{D_i^T V_i^{-1}} (\mathbf{y_i} - \mu_\mathbf{i})$ in (1) is equivalent to the estimating function suggested by Wedderburn (1974) except that the values of $\mathbf{V_i}$ in (1) are functions of α and β. That is, an additional set of parameters α is introduced in (2) in order to account for the correlation of measurements. These parameters may consistently be estimated by a set of estimating equations and are functions of $\widehat{\beta}$ and $\widehat{\phi}$, by noting that the estimation of depends on assumptions made on the correlation structure of responses. For estimating the working correlation matrix the method of moments is usually used. The choice of $\mathbf{R_i}(\alpha)$ has been discussed for example by Liang and Zeger in some detail.

The GEE estimate of β is obtained by iterating between solutions of equation (2). This requires the use of numerical methods, such as Fisher's scoring algorithm, iterative weighted least squares, or quasi-Newton algorithms. We need also to specify the link function and the correlation structure. For example, if the responses are count then we may choose a log link and then examine the effects of overdispersion problem. Furthermore, in using GEE an important step is to correctly specify the form of correlation matrix. There are several options to select this form and depends on the nature of the data collected. The aim of selecting a correct form of the working correlation structure is to estimate β more efficiently (Hardin and Hilbe, 2003). For the longitudinal and panel data which cross sectional observations are correlated over time points, an autoregressive structure, or a compound-symmetry for equally correlated measurements may be selected in order to incorporate the within-subject correlations.

3 Data description

The trial was a single blind randomized control trial, based on a two-year caries clinical experiment. In order to obtain a 95% confidence interval (at the two sided 5% level), a 65% statistical difference between the groups together with an expected range of 2.23-7.2 local DMFS, and a minimum sample size of 22 subjects in each group was required. Allowing for an expected annual drop-out rate of 10% (23.5%

increase in sample size for a 2-year Follow-up 18), a minimal recruitment was set at 28 subjects per group (84 in total). Volunteers were recruited in Isfahan-Iran from a collection of 13 to15 year-old students. With respect to the high caries rates, this region is a good site for caries clinical trials. 20 Subjects brush their tooth at least once a day with fluoridated toothpaste. A randomization list was drawn up by a statistician and given to the independent pharmacists. Subsequently, he randomly assigned the enrolled participants to three comparison groups (fluoride gum, fluoride mouth rinse and concurrent control group).

At base-line and annual cutoffs, each participant attended in an oral health education session. Participants were supplied with fluoride gum, and each piece contained 0.25 mg fluoride. Gums were chewed for about 20 minutes twice per day (just after waking up and right before bedtime). In order to get an even fluoride distribution in the oral cavity we advised subjects about the importance of using both sides of the dentition during of chewing. A group of six designated as the 'In mouth rinse' group were asked to rinse with 5 ml of 0.2% natural NaF (fluoride mouth wash, Shahre Daru Laboratories Co., Tehran, Iran) solution for 60 seconds once a week lastly. The control group received only the dental examinations and oral hygiene education sessions.

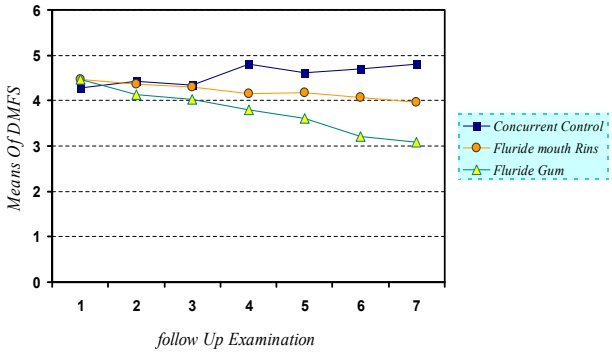
The subjects tooth health status was determined by DMFS index and using the World Health Organization (WHO) (see e.g. Ismail, 2004) criteria for detection of dental caries. Safety procedures to prevent injury and cross-infection were followed. Dental examinations were carried out at the time of randomization (base-line) and at 3-month intervals for two years by one trained, experienced dentist unaware of the trial convention. Therefore, the measurements were made over eight different stages during the two-year period.

4 Analysis of the dental clinical trial

The main interest in our study is to relate the differences in DMFS index with the three setting groups for both males and females. We set

the baseline DMFS measure as a covariate; using the three experimental treatments and follow up times as a dummy variable. Estimated marginal means of DMFS scores in three trial groups are shown in Figure 1.

Figure 1: Marginal means of DMFS on fallow up examination in different treatments.



We then fit a mixed-effects Poisson regression model with the log link function to obtain log odds of achieving alleviation. The GEE method with various working correlation structures is used to take account the dependence nature of DMFS scores. We use the procedure GENMOD in SAS version 9.1 (SAS Institute Inc., Cary, NC, USA). The incidence rates of caries and estimated marginal means of DMFS scores in three trial groups are shown in Table 1.

In order to compare the correlation matrices the quasi-likelihood criterion (QIC), a modified version of Akaike's information criterion, is used. A smaller QIC value represents a better fitting model (Pan, 2001). We observe that AR(1) and equal correlation (r=0.6319) have the minimum QIC between all of the correlation structures with values 203.936 and 203.725, respectively. Results are given in Table 1. The regression estimates and z-statistics did not significantly change for two structures. This finding agrees with Liang and Zeger (1986) theory that z-statistics are robust to the misspecification of correlation structures, provided that a correct link function is applied. The scaled deviance is nearly 1 showing that there is no overdispersion problem.

	covariate	Estimate	S.E	%95 confidence interval		Z	p-value
Main Effect	constant	۰/۵۶۷۰	۰/۱۶۸۸	۰/۲۳۶۳	۰/۸۹۷۸	۳/۳۶	۰/۰۰۰۸
	Treatment 1*	۰/۳۹۷۰	۰/۱۰۲۰	۰/۱۹۷۱	۰/۵۹۶۹	۳/۸۹	<۰/۰۰۰۱
	Treatment 2*	۰/۱۰۳۴	۰/۰۹۹۲	-۰/۰۹۱۱	۰/۲۹۷۹	۱/۰۴	۰/۲۹۷۶
	Time1	۰/۳۲۴۶	۰/۰۸۱۰	۰/۱۶۵۸	۰/۴۸۳۵	۴/۰۱	<۰/۰۰۰۱
	Time2	۰/۲۶۷۰	۰/۰۷۷۱	۰/۱۱۵۸	۰/۴۱۸۲	۳/۴۶	۰/۰۰۰۵
	Time3	۰/۲۰۱۸	۰/۰۶۰۸	۰/۰۸۲۶	۰/۳۲۱۰	۳/۳۲	۰/۰۰۰۹
	Time4	۰/۱۶۱۱	۰/۰۵۲۲	۰/۰۵۸۹	۰/۲۶۳۴	۳/۰۹	۰/۰۰۲۰
	Time5	۰/۱۰۲۶	۰/۰۳۲۰	۰/۰۳۹۸	۰/۱۶۵۳	۳/۲۰	۰/۰۰۱۴
	Time6	۰/۰۰۰۱	۰/۰۰۲۴	-۰/۰۰۴۶	۰/۰۰۴۸	۰/۰۳	۰/۹۷۲۸
	Sex	-۰/۰۱۷۲	۰/۰۴۱۶	-۰/۰۹۶۸	۰/۰۶۴۳	-۰/۴۱	۰/۶۷۹۴
	Base-line	۰/۱۴۰۷	۰/۰۲۴۷	۰/۰۹۲۴	۰/۱۸۹۰	۵/۷۱	<۰/۰۰۰۱
Intractions	Treatment 1 * time 1	-۰/۴۴۲۷	۰/۱۰۸۸	-۰/۶۵۵۹	-۰/۲۲۹۴	-۴/۰۷	<۰/۰۰۰۱
	Treatment 1 * time 2	-۰/۳۵۲۳	۰/۰۹۵۲	-۰/۵۳۸۸	-۰/۱۶۵۷	-۳/۷۰	۰/۰۰۰۲
	Treatment 1 * time 3	-۰/۳۰۳۳	۰/۰۷۹۵	-۰/۴۵۹۲	-۰/۱۴۷۵	-۳/۸۱	۰/۰۰۰۱
	Treatment 1 * time 4	-۰/۲۲۱۰	۰/۰۶۵۲	-۰/۳۴۸۹	-۰/۰۹۳۲	-۳/۳۹	۰/۰۰۰۷
	Treatment 1 * time 5	-۰/۱۴۴۸	۰/۰۴۰۸	-۰/۲۲۴۸	-۰/۰۶۴۸	-۳/۵۵	۰/۰۰۰۴
	Treatment 1 * time 6	-۰/۰۲۵۰	۰/۰۱۳۱	-۰/۵۰۷	۰/۰۰۰۶	-۱/۹۱	۰/۰۵۵۸
	Treatment 2 * time 1	-۰/۱۴۹۵	۰/۱۰۲۸	-۰/۳۵۱۰	۰/۰۵۲۱	-۱/۴۵	۰/۱۴۶۲
	Treatment 2 * time 2	-۰/۱۱۶۱	۰/۰۹۵۴	-۰/۳۰۳۱	۰/۰۷۰۹	-۱/۲۲	۰/۲۲۳۵
	Treatment 2 * time 3	-۰/۰۷۱۲	۰/۰۷۵۳	-۰/۲۱۸۹	۰/۰۷۶۴	-۰/۹۵	۰/۳۴۴۵
	Treatment 2 * time 4	-۰/۰۶۰۷	۰/۰۶۴۳	-۰/۱۸۶۷	۰/۰۶۵۴	-۰/۹۴	۰/۳۴۵۴
	Treatment 2 * time 5	-۰/۰۶۰۷	۰/۰۳۹۳	-۰/۱۳۷۷	۰/۰۱۶۳	-۱/۵۵	۰/۱۲۲۱
	Treatment 2 * time 6	۰/۰۲۱۵	۰/۰۱۶۲	-۰/۰۱۰۳	۰/۰۵۳۳	۱/۳۲	۰/۰۰۰۸

Treatment 1: Control group
Treatment 2: Consumers of fluoride mouth rinse

Table 1. Results of the GEE estimation for AR(1) structure.

Results of Table 1 show that three groups reach statistical significance (p¡0.05). Difference between the F-gum group, F-rinse group and control group reach statistical significance. Subgroup analyses show a statistically significant difference between genders. Also, the interaction between fallow up times and control group, and also between F-rinse group and last fallow up time (p¡0.05) are significant. Nevertheless, the difference between efficacy of F-gum and F-rinse is not significance. Furthermore, we study the contrast between the groups, known as the effect size. It was not pre-specified in the trial convention. The DMFT of 15 years-old Iranian students, according to WHO oral health country/area profile 36 was assumed as threshold and odds ratio was calculated (Table 2). In all follow-up sessions, the odds ratios are lower than one, indicating that F-gum has greater efficacy than F-rinse.

Follow up examination	Odds ratio	P-value	95% Confidence interval
1	0.000	0.562	(0.159, 2.191)
2	0.000	0.177	(0.046, 1.491)
3	0.350	0.291	(0.053, 1.854)
4	0.084	0.056	(0.000, 1.046)
5	0.109	0.119	(0.000, 1.498)
6	0.145	0.234	(0.000, 2.360)
7	0.153	0.249	(0.000, 2.460)

Table 2. Results of contrast between the efficacy of F-gum and F-rinse.

5 Recommendations

For practical use, some recommendations are needed to decide whether the GEE approach for longitudinal data analysis should be used. In general, the GEE method is used if the mean of responses conditional on covariates is correctly specified. Otherwise, misspecification may yield inconsistent estimates of the model parameters. Furthermore, the incorrect choice of correlation structure leads to inefficiency of the GEE estimation. However, this issue has not been investigated in detail and remains open. The recommendation is to consider various association structures in the GEE estimation process and select the best fitted model using suitable some selection criteria. In this paper,

we used GEE in analyzing a data set came from a dental clinical trial and the quasi-likelihood criterion to select the best form of the correlation matrix. We find that the AR(1) correlation structure is best fitted to the data. Furthermore, results show that in all follow-up sessions fluoride gum has greater efficacy than fluoride rinse.

References

[1] Ballinger G.A. (2004). Using Generalized Estimating Equations for Longitudinal Data. *Analysis Organizational Research Methods*, **7**(**2**), 127-150.

[2] Crouchley, R. and Davies R.B. (2001). A Comparison of GEE and Random Effects Models for Distinguishing Heterogeneity, Nonstationarity and State Dependence in a Collection of Short Binary Event Series. it Statistical Modelling: An International Journal, **1**(**4**), 271-285.

[3] Diggle, P.J., Heagerty, P., Liang, K.-Y. and Zeger, S.L. (2002). Analysis of Longitudinal Data (2nd ed.). Oxford, UK: Oxford University Press.

[4] Emond, M.J., Ritz, J. and Oakes, D. (1997). Bias in GEE Estimates from Misspecified Models for Longitudinal Data. Commun. Statist.-Theory Meth., 26(1), 15-32.

[5] Fitzmaurice, G.M. and Laird, N.M. (1995). Regression Models for a Bivariate Discrete and Continuous Outcome With Clustering. Journal of the American Statistical Association, 90, 845-852.

[6] Hardin, J.W. and Hilbe, J.M. (2003). Generalized Estimating Equations. Boca Raton, FL: Chapman and Hall/CRC Press.

[7] Ismail, A.I. (2004). Visual and Visuo-tactile Detection of Dental Caries. J. Dent. Res., 83(suppl_1), C: 56 -66.

[8] Liang, K.-Y., and Zeger, S.L. (1986). Longitudinal Data Analysis Using Generalized Linear Models. Biometrika, 73, 13-22.

[9] Pan,W. (2001). Akaike's Information Criterion in Generalized Estimating Equations. Biometrics, 57, 120-125.

[10] Pan, W. and Connett, J.E. (2002). Selecting the Working Correlation Structure in Generalized Estimating Equations with Application to the Lung Health Study. Statistica Sinica, 12, 475-490.

[11] SAS Institute (1999). SAS/STAT Software: Changes and Enhancements for Release 9.1, Cary, NC: SAS Institute, Inc.

[12] Wedderburn, R. W. M. (1974). Quasi-likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. Biometrika, 61, 439-447.

[13] Ziegler, A., Kastner, C. and Blettner, M. (1998). The Generalised Estimating Equations: An Annotated Bibliography. Biometrical Journal, 40(2),115-39.

[14] Zorn, C. J. W. (2001). Generalized Estimating Equation Models for Correlated Data: A review with Applications. American Journal of Political Science, 45, 470-490.

A Modified Process for Modeling Wireless Communication Networks

N. Modarresi, and S. Rezakhah

Department of Mathematics and Computer Sciences, Amirkabir University

Abstract: Extended Alternating Fractal Renewal Process (EAFRP) proposed for modeling the self-similar and impulsive traffic of high-speed networks. It yields traffic is long-range dependence in the both single and multi-user case. In this paper we propose a modification of the EAFRP model that is motivated by the distinctive two-slope appearance of the log-log complementary distribution(or survival) function(LLCD) of real traffic data. We here takes into account the fact that the network has a limit R on the total traffic rate. In addition each user's traffic rate is often independently limited to a value L which is significantly lower than the network's limit ($L < R$). In the rate-limited EAFRP model, reality is captured by modeling the transmission rate during the On states by a cut-off Pareto random variable. We show that this model achieves a closer approximation of the observed reality than the EAFRP model.

Keywords: Wireless network traffic; Selfsimilarity; Heavy-tailness; Alternating fractal renewal process.

1 Introduction

As the popularity of multimedia services grows at an unprecedented rate, the need of higher speed data access is becoming imperative. Extensive studies involving high-definition network measurements indicate that data traffic in high-speed communication networks exhibits self-similar and impulsive characteristics [1], [2], [3]. Self-similarity implies a non-trivial structure for traffic, which can be exploited for data analysis. Recently, there have been some works suggesting that wireless traffic may also exhibit self-similarity. It is true that recent advances in wireless networks can enable high speeds, and it is also true that both wireless and wireline users need to access the same multimedia applications.

A stochastic process is impulsive if its marginal distribution is heavy-tailed. The main characteristic of a random variable obeying a heavy-tailed distribution is that it exhibits extreme variability, i.e. it fluctuates far away from its mean value with non-negligible probability. Statistical modeling of traffic is very important in network engineering, and a substantial body of literature has been devoted to it. However, models developed for traditional teletraffic no longer apply to data traffic.

In high-speed networks, the packets are communicated in a packet train fashion; furthermore, the length of the packet train is heavy-tail distributed. This observation led to the well-known On/Off model [2], also called Alternating Fractal Renewal Process (AFRP) [4]. While the AFRP model provides insight on the causes of self-similarity of traffic, its Gaussian aggregated results is inconsistent with real traffic data, which depart greatly from Gaussianity. An extension of the AFRP, namely the EAFRP, that captures heavy-tailness as well as self-similarity of traffic has been proposed in [5]. Based on the On/Off model, a single source destination active pair alternates between two states: the On, during which, there is data flow between source and destination, along either way, and the Off, which is the quiet duration. Both the On and Off durations are heavy-tailed distributed. The self-similar characteristics of the AFRP have been attributed to the heavy-tail properties of the On/Off states durations. It was shown that the cumulative superposition of infinite AFRPs is fractional Brownian motion, the only Gaussian process with stationary increments that is self-similar [6].

In this paper, we propose a modification of the EAFRP model that enables a closer match to real traffic. The modification is motivated by the distinctive two-slope appearance of the log-log complementary distribution (or survival) function (LLCD) of real traffic data, and the nature of the true bounds on the user transmission rates in real networks. Limits on the senders and the receivers TCP

window sizes, TCP congestion avoidance strategies, and bandwidth bottlenecks within the end-systems are among many of the reasons that lead to an independent limit on each individual users transmission rate [7]. In reality, therefore, if R is the peak rate of the link on to which traffic from multiple users is multiplexed, the sum of the user transmission rates is bounded by R and each users transmission rate is bounded by an even smaller quantity, $L(L < R)$. In our modified EAFRP model, i.e. the rate-limited EAFRP, we capture this reality by modeling the transmission rate during the On states by a cut-off Pareto random variable, while the On/Off durations are distributed as in the EAFRP model. We show that the existence of these two rate limits, L and R, results in the distinctive two slope behavior of the LLCD of the overall traffic, a fact that has not been explained so far by existing models.

2 Mathematical preliminaries

In this chapter, we formulate technical backgrounds of long-range dependence, self-similarity, heavy-tail distributions and AFRP model.

Definition 1 ***Long-range dependence***

Let $\{X_k\}_{k\in Z}$ be a discrete-time stationary stochastic process with finite second-order statistics. Denote $\mu = E\{X_k\}$ and $\sigma^2 = E\{(X_k - \mu)^2\}$ then $\{X_k\}$ is a long-range dependence process, if

$$\lim_{l\to\infty} \frac{r(l)}{l^{2H-2}} = c \tag{1}$$

where the correlation of $\{X_k\}$ is denoted as

$$r(l) = \frac{E\{(X_k - \mu)(X_k + 1 - \mu)\}}{\sigma^2}$$

and the coefficient H is reffered to as the Hurst parameter and $1/2 < H < 1$; c is a positive constant.

For a long-range dependence process, the hyperbolic decay of $r(l) = cl^{-\beta}(0 < \beta < 1)$*, and* $\beta = 2 - 2H$*, results in a non-summable correlation i.e.*

$$\sum_{i=-\infty}^{\infty} r(l) = \infty \tag{2}$$

For processes which might not have second-order statistics, a structure measure different than the autocorrelation is needed. We use the quantity defined in [3] as

$$I(\rho_1, \rho_2; \tau) = -\ln E\{e^{j[\rho_1 x(t+\tau)+\rho_2 x(t)]}\} + \ln E\{e^{j\rho_1 x(t+\tau)}\} + \ln E\{e^{j\rho_2 x(t)}\} \tag{3}$$

We say that the stationary process $X(t)$ *is long-range dependent in a generalized sense if*

$$\lim_{\tau\to\infty} \frac{-I(1, -1; \tau)}{\tau^{\beta-1}} = c \tag{4}$$

Definition 2 ***Self-similarity***

The real-valued process $X(t)(t \in R)$*, is self-similar with index* $H > 0$*, if for all* $a > 0$ *and* $t \geq 0$*, all finite dimensional distributions of* $a^{-H}X(at)$ *are identical to the finite dimensional distributions of* $X(t)$*, i.e.*

$$a^{-H}X(at) \stackrel{d}{=} X(t) \tag{5}$$

where $\stackrel{d}{=}$ *represents equality for all finite dimensional distributions. The index* H *is referred to as the Hurst parameter of the self-similar process* $X(t)$*.*

Note that, the time-scaled long-range dependent process maintains similarity to the original process, thus indicating a relationship between long-range dependence and self-similarity. When we say traffic is self-similar, we mean it is long-range dependence in the sense that the autocorrelation function has hyperbolically decaying tail.

Definition 3 ***Heavy-tail distributions***

A random variable X is heavy-tailed distributed with index α if

$$P(X \geq x) \sim px^{-\alpha}L(x) \text{ and } P(X \leq -x) \sim qx^{-\alpha}L(x), \quad x \to \infty \quad (6)$$

where $0 < \alpha < 2; p, q \geq 0$ with $p+q = 1; L(x)$ is a slowly varying function such that $L(x)$ is positive for large x and $\lim_{x\to\infty} L(tx)/L(x) = 1$ for any positive b.

For heavy-tail distributions, $m-th$ order statistics exist if $m < \alpha$. Thus heavy-tail distributions have infinite second-order statistics and can have infinite mean if $\alpha < 1$.

An example of important subclass of distributios with regulary varying tail is the Pareto distribution, which has a survival function of

$$\overline{F}(x; \alpha, K) = P(X \geq x) = \begin{cases} (\frac{K}{x})^{\alpha} & x \geq K, \\ 1 & x < K, \end{cases}$$

where K is positive constant and $0 < \alpha < 2$ and especially the mean exists if $1 < \alpha < 2$. We will be using a variation of the Pareto distribution, namely the cut-off Pareto, defined in terms of density function equals:

$$f_L(x; \alpha, K) = f(x; \alpha, K)(1 - u(x - L)) + (\frac{K}{L})^{\alpha}\delta(x - L) \quad (7)$$

Where $f(.)$ and $F(.)$ denote the Pareto density function and Pareto distribution respectively, $u(.)$ is the unit step function and $\delta(.)$ is the Dirac function.

Definition 4 ***AFRP model***

The Altenating Fractal Renewal Process(AFRP), proposed in [2] for modeling of network traffic, is a process that alternates between two states, 0 or 1. The time $\{X_n, n \in Z\}$, spent in state 1, is a random variable with probability density function $f_1(t)$, and the time $\{Y_n, n \in Z\}$, spent in state 0, is a random variable with probability

density function $f_0(t)$, where X_n, Y_n are independent, and both Pareto distributed,

$$f_i(t) = O(t^{-(\alpha_i+1)}), \ \ where \ \ i = 0, 1 \ \ and \ \ \alpha_i \in (1, 2) \tag{8}$$

Generally speaking, $f_0(t) = f_1(t) = 0$ for $t < 0$, and the associated dwell mean times $\mu_1 := E[X_n]$ and $\mu_0 := E[Y_n]$ are finite. The expected value of the AFRP process $V(t)$ is $\mu_1/\mu_0 + \mu_1$. The power spectral density of AFRP equals [8]:

$$S(\omega) = E\{V(t)\}\delta(\omega/2\pi) + \frac{2\omega^{-2}}{\mu_0 + \mu_1} Re\{\frac{[1 - Q_0(-j\omega)][1 - Q_1(-j\omega)]}{1 - Q_0(-j\omega)Q_1(-j\omega)}\} \tag{9}$$

where $Q_0(-j\omega)$, $Q_1(-j\omega)$ are Fourier transforms of $f_0(t)$ and $f_1(t)$ respectively.

3 The proposed model

The EAFRP model was proposed in [5] as an extension to the AFRP [8]. It yields traffic that is impulsive and long-range dependent in the generalized codifference sense at both single and multi-user case. However, for mathematical tractability, it relies on the assumption of infinite bandwidth be available during the On state, an assumption that is not met in a real network. When used to synthesize traffic, although it matches well a significant portion of the log-log complementary distribution (LLCD), it does not capture the distinctive two-slope appearance of the LLCD. This traffic behavior, although never commented on, can also be seen in the figures of [9] ,[1] and other papers.

PROPOSITION 0.12 *Let us consider an On/Off process, $s(t)$, defined as:*

(A1) The On periods $\{X_n, n \in Z\}$ and the Off periods $\{Y_n, n \in Z\}$ are i.i.d independent of each other with distributions $\overline{F_1} = \overline{F}(x; \alpha_1, K_1)$ and $\overline{F_0} = \overline{F}(x; \alpha_0, K_0)$, with $\alpha_0, \alpha_1 > 1$, thus have finite means μ_1 and μ_0 respectively;

(A2) The rates during the On states are random variables cut-off Pareto distributed with probability density function $f_L(x;\alpha,K)$, and independent of X_n, Y_n.

Then $s(t)$ is distributed according to:

$$f_s(x)=\frac{\mu_0}{\mu_1+\mu_0}\delta(x)+\frac{\mu_1}{\mu_1+\mu_0}f_L(x;\alpha,K) \tag{10}$$

Proof: $s(t)$, can be expressed as $s(t)=A(t)V(t)$ where $V(t)$ is an AFRP, and $A(t)$ corresponds the transmission rate, distributed according to assumption (A2). The probability density function of $s(t)$ is $f_s(x)=P[V(t)=0]\delta(x)+P[V(t)=1]f_L(x)$ where $\delta(x)$ is the Dirac function, taking value of 1 at $x=0$ point only. Since $P[V(t)=1]=1-P[V(t)=0]=\frac{\mu_1}{\mu_1+\mu_0}$ for$x>0$, $f_s(x)$ is a scaled version of $f_L(x)$. Thus $s(t)$ is a cut-off Pareto, exhibiting a power-law survival function for values less that L.□

Let us now consider M independent i.i.d. On/Off process, $s_i(t)$; $i=1,\ldots,M$, each $s_i(t)$ constructed according to assumptions (A1),(A2). Let us form the process $S(t)$ as the superposition of the $s_i(t)$; $i=1,\ldots,M$, followed by a thresholding operation with threshold R. In the following we will provide some insight on the form of the pdf and the LLCD of the process $S(t)$.

For simplicity, let us first consider the case where $M=2$. The pdf of $S(t)$ will be:

$$f_S(x)=(f_s(x)*f_s(x))(1-u(x-R)) \tag{11}$$

where $f_s(x)$ denotes the pdf of each $s_i(t)$ given on (10), and '*' denotes convolution.

It is simple to extend the above result to any M. Still in that case the LLCD will exhibit a linear trend over the interval $[MK;L]$, followed by a Gaussian-type decay, i.e. the tail decays with the manner corresponding to a Gaussian process in LLCD. As expected by the

Central Limit Theorem, as M increases, the resulting process will become Gaussian. This is in agreement by the discussion above, since for M very large, the linear segment will occur over increasingly smaller range and will eventually disappear. Of course, the larger the L, the larger M it will take for the result to become Gaussian.
Since it is hard to derive the pdf of the superposition $S(t)$ in closed form, we approximate it over the interval $[MK; R]$ by the following mixture:

$$f_S(x) \sim (\frac{K_{A_2}}{R})^{\alpha_{A_2}} \delta(x-R) + f(x; A_1, K_1)[1-u(x-L)] \\ + f(x; A_2, K_2)[u(x-L) - u(x-R)] \qquad (12)$$

where $A_1 = \alpha$ if the $\{s_i(t), i = 1, \ldots, M\}$ are i.i.d otherwise, $A_1 = min\{\alpha_1, \ldots, \alpha_M\}$ and $K_1 = KM$ if the $\{s_i(t), i = 1, \ldots, M\}$ are i.i.d otherwise, $K_1 = \sum_{i=1}^{M} K_i$. α_2 can be any positive value and

$$K_{A_2} = \exp\{\frac{1}{\alpha_{A_2}} \ln[K_{A_1}^{\alpha_{A_1}} L^{(\alpha_{A_2} - \alpha_{A_1})}]\} \qquad (13)$$

PROPOSITION 0.13 *The process defined through (A1),(A2) , and also the supposition of M such processes are long-range dependent.*

Proof: Due to the cut-off nature of $f_L(x)$, the proposed model will have finite moments. Therefore, long-range dependence is here examined in terms of its covariance. The joint characteristic function of an On/Off process with arbitrarily distributed On durations can be found in [5]. Based on that expression, the covariance of $s(t)$ can be found as the second-order derivative of the characteristic function, $\phi_s(s_1, s_2; \tau)$, evaluated as 0.

The overall traffic $S(t)$ is the supperposition of M independent and identical distributed as proposed process $s_m(t)$, $(m = 1, 2, \ldots, M)$, i.e. $S(t) = \sum_{m=1}^{M} S_m(t)$ then,

$$\phi_S(s_1, s_2; \tau) = \prod_{m=1}^{M} \phi_{s_m}(s_1, s_2; \tau) \qquad (14)$$

The covariance function of $S(t)$ equals:

$$c_S(\tau) = -\frac{\partial^2 \phi_S(s_1, s_2; \tau)}{\partial s_1 \partial s_2}|_{s_1=0,s_2=0} = \sum_{m=1}^{M} \{-\frac{\partial^2 \phi_{S_m}(s_1, s_2; \tau)}{\partial s_1 \partial s_2}\}|_{s_1=0,s_2=0}$$

$$\times \prod_{n=1,n\neq m}^{M} \{\phi_{S_n}(s_1, s_2; \tau)\}|_{s_1=0,s_2=0}$$

It can be easily derived that:

$$c_S(\tau) \sim M c_{\rho_1}^{M-1} c_{\rho_2} \tau^{M(1-\alpha_i)} \tag{15}$$

where $\alpha_i = \min(\alpha_1, \alpha_0)$.

In the proof of Proposition 2, we assume that $ML < R$. If $ML > R$, there is a non-zero probability of traffic congestion at the multiplexing point. The total transmitted traffic rate, in the presence of congestion, is not a mere supposition of the processes that describe the transmission rates of individual users, and therefore, it is not clear that it will still be long-range dependent. While a detailed queuing-theoretic analysis is necessary to obtain the characteristics of the process that describes the total traffic in the presence of such congestion, initial results in [10] obtained through simulation suggest that at least the degree of long-range dependence in the total traffic will reduce in comparison to that of the process describing the individual user transmission rates.

To provide some insight on the queuing characteristics of traffic synthesized by the proposed model, we assume that the total network traffic can be represented by a single On/Off process defined as in (A1) and (A2). Let $Q(t)$, $t \geq 0$ denote the buffer content at time t, and let

$$Q(\infty) \stackrel{d}{=} \lim_{t\to\infty} Q(t) \tag{16}$$

where $\stackrel{d}{=}$ represent equality in distribution. It can be shown that stationary queue length $Q(\infty)$ is heavy-tail distributed with tail index $\alpha_1 - 1$.

4 Conclusion

We have presented a constructive model for high-speed network traffic that achieves a close approximation to real traffic than previously known constructive models. The modeling of traffic generated by a single user was performed along the lines of the EAFRP [5], with the primary difference being the introduction of a maximum rate limit (L).

The total traffic rate through the network also experiences a limit (R) in real networks. We have shown that the existence of these two limits, L and R, lead to a total traffic whose LLCD and autocorrelation match those of real traffic very closely. In particular, this provides insight for the first time into the two-slope appearance of the LLCD of real traffic. We have further shown that the supposition of the proposed rate-limited EAFRP processes is long-range dependent. Thus, our model preserves the long-range dependence of the total traffic in the traffic in the absence of congestion. The correlation structure of the total traffic when $ML > R$, i.e. when there is a non-zero probability of congestion, remains to be investigated.

The proposed model has finite variance and thus, as the number of users increases, the total traffic will eventually become Gaussian. This is consistent with what can be observed in data that are collected at gateways. However, we were able to show that as L increases, it will take a larger number of users for the traffic to become Gaussian. This implies that in modern networks, when L is large, the traffic will be non-Gaussian. In that case, the proposed model can be a useful tool in making certain design choices in the network infrastructure.

References

[1] M. E. Crovella, A. Bestavros, Self-similarity in world wide web traffic: Evidence and possible causes, IEEE/ACM Trans. Networking, Vol.5, No.6, December 1997.

[2] W. Willinger, M.S. Taqqu, R. Sherman, and D.V. Wilson, Self-similarity through high-variability: statistical analysis of Ethernets LAN traffic at the source level, IEEE/ACM Trans. Networking, Vol.5, No.1, February 1997.

[3] W. Willinger, V. Paxson and M. S. Taqqu, Self-similarity and heavy tails: structural modeling of network traffic, A Practical Guide to Heavy Tails : Statistical Techniques and Applications, R. J. Adler, R. Feldman, M. S. Taqqu(Ed.),1998.

[4] S.B. Lowen and M.C. Teich, Fractal renewal processes generate 1/f noise, Physical Review E, Vol.47, No., Feb 1993

[5] X. Yang, A.P. Petropulu, The extended alternating fractal renewal process for modeling traffic in high-speed communication networks, IEEE Trans. Sig. Proc., vol. 49, no. 7, July 2001.

[6] M.S. Taqqu, W. Willinger and R. Sherman, Proof of a fundamental result in self-similar traffic modeling, Computer Communication Review 27, pp.5-23, 1997.

[7] Y. Zhang and L. Breslau and V. Paxson and S. Shenker, On the Characteristics and Origins of Internet Flow Rates, Proceedings of ACM SIGCOMM, Aug. 2002, Pittsburgh, PA.

[8] S. B. Lowen and M. C. Teich, Fractal renewal processes generate 1/f noise, Physical Review E, Vol.47, No.2, Feb 1993.

[9] O. Cappe, Eric M., Jean-Christophe P., A. Petropulu and X. Yang, Long-Range Dependence and Heavy-Tail Modeling for Teletraffic Data, IEEE Singal Processing Magazine, Vol.19, May 2002, pp. 14-27.

[10] Y. Zhou and H. Sethu, A Simulation Study of the Impact of Switching Systems on the Self-Similar Properties of Traffic, Proceedings of the IEEE Workshop on Statistical Signal and Array Processing, Aug. 2000, Pocono Manor, PA.

Keywords: Multivariate Moving Average Process, Backward and Forward Moving Average Representations, Spectral Factorization.

1 Introduction

We shall consider a discrete time q-variate second order moving average of order one (MA(1)) process $(\mathbf{x}_n)_{n\in\mathbb{Z}}$, $\mathbb{Z}$ the set of integers, and provide its forward moving average representation.
In the univariate case the situation is rather straight forward. Forward moving average representation, whenever exists, assume correspondingly the same coefficients as the backward one; due to the facts that the spectral density is even and admits a spectral factorization with commutative spectral factors. In the multivariate case neither the spectral density matrix is necessarily even and nor the spectral factors are commutative. As we shall show in this article the situation in the multivariate case, in general, is different.

Indeed the forward MA coefficients come from the right Wiener and Masani spectral factorization, called here the right W-M spectral factorization. Soltani and Mohammadpour (2006) provided a method to establish the right factorization from the left factorization. There are extensive works on the analytical and numerical derivations of the left W-M spectral factorization; Wiener and Masani (1957, 1958), Tuel (1968), Miamee (1988), Bini, Fiorentino, Gemignani and Meini (2003) among others.

Forward MA representations, in the univariate case, have been applied for different purposes in time series and some other fields. Simultaneous consideration of the left and right W-M spectral factors has appeared to be necessary and promising in certain areas of multivariate time series. This issue is brought into sight by Findley (1991), following Hannan and Diestler (1988, page 270), in providing the multivariate generalization of the Baxter's inequality and rate of convergence of finite predictors. Cheng and Pourahmadi (1993) proved that certain smoothness properties of the spectral density are inherited by its spectral factors. The works of Pourahmadi (1992, 2001)

and Soltani and Yaghmaee (2001) illustrate benefits of links between forward and backward MA representations in interpolation using finite or infinite future values.

2 Preliminaries

Let $(\mathbf{x}_n)_{n\in\mathbb{Z}}$ be a q-variate second order process. Each $\mathbf{x}_n$ is a column vector whose components $x_n^{(1)}, \cdots, x_n^{(q)}$ are complex random variables, possessing the second moment, $E|x_n^{(j)}|^2 < \infty, \quad n \in \mathbb{Z}, \quad j = 1, \cdots, q$. Following Wiener and Masani (1957, 1958), the time domain of the $(\mathbf{x}_n)_{n\in\mathbb{Z}}$ consists of certain complex q-variate random vectors generated by $(\mathbf{x}_n)_{n\in\mathbb{Z}}$ equipped with the Gramian $(\mathbf{x}, \mathbf{y})$, which is indeed the covariance matrix, and the trace $((\mathbf{x}, \mathbf{y})) = \tau(\mathbf{x}, \mathbf{y})$, the trace of the covariance matrix.

Let $(\mathcal{L}_2, ((\ ,\ ))\ )$ denote the Hilbert space of all centered q-variate random vectors $\mathbf{x}$. Let $(\mathbf{x}_n)_{-\infty}^{+\infty}$ be a centered sequence in $\mathcal{L}_2$, then $(\mathbf{x}_n)_{-\infty}^{+\infty}$ is said to be stationary if for each m, n, $(\mathbf{x}_m, \mathbf{x}_n)$ depends only on $m-n$. We refer to a centered stationary sequence $(\mathbf{x}_n)_{-\infty}^{+\infty}$ in $\mathcal{L}_2$ as a centered q-variate second order stationary stochastic process (q-ssp), we drop the terms "centered" and "second order" whenever there is no ambiguity.

It is known that a q-ssp $(\mathbf{x}_n)_{n\in\mathbb{Z}}$ admits the following spectral representation

$$\mathbf{x}_n = \int_0^{2\pi} e^{-inx} d\mathbf{\Lambda}(x), \tag{2.1}$$

where

$$\mathbf{\Lambda}(A) = \mathbb{E}(A)\mathbf{x}_0, \quad A \subset [0, 2\pi),$$

and $\mathbb{E}(\theta)$ is the spectral resolution of the shift operator of the process $\mathbf{x}$, Wiener and Masani (1957, page 140). It is also well known that every q-ssp $(\mathbf{x}_n)$ possesses one and only one spectral distribution matrix function (spectral distribution matrix in short), i.e., a σ-additive (under the trace norm) matrix function $\mathbf{F}$ on Borel sets of $[0, 2\pi)$ for

which $\mathbf{F}(A) = (\mathbf{\Lambda}(A), \mathbf{\Lambda}(A))$ and $\mathbf{F}(A)$ is nonnegative definite for every $A \subset [0, 2\pi)$. If $\mathbf{F}$ is absolutely continuous with respect to the Lebesgue measure, then its derivative $\mathbf{F}'$ will be called the spectral density matrix function (spectral density matrix in short). For almost all $\theta \in [0, 2\pi)$, $\mathbf{F}'$ is Hemitian, i.e., $\mathbf{F}' = \mathbf{F}'^*$, and nonnegative definite.

We define a vector-valued function associated with a matrix and closely related to the Kronecker product. For a matrix $\mathbf{A} \in \mathcal{F}^{m\times n}$, write

$$\mathbf{A} = [\mathbf{A}_{*1}, \mathbf{A}_{*2}, \ldots, \mathbf{A}_{*n}],$$

where $\mathbf{A}_{*j} \in \mathcal{F}^m$, $j = 1, 2, \ldots, n$. Then the vector

$$\begin{bmatrix} \mathbf{A}_{*1} \\ \mathbf{A}_{*2} \\ \vdots \\ \mathbf{A}_{*n} \end{bmatrix} \in \mathcal{F}^{mn}$$

is said to be the *vec-function* of $\mathbf{A}$ and is written vec($\mathbf{A}$). It is the vector formed by stacking the column of $\mathbf{A}$ into one log vector. Note that the vec-function is linear and has the property that

$$\text{vec}(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}\mathbf{X},$$

where $\mathbf{A} \in \mathcal{F}^{m\times m}$, $\mathbf{X} \in \mathcal{F}^{n\times n}$, $\mathbf{B} \in \mathcal{F}^{m\times n}$ and $\otimes$ denotes the Kronecker product.

3 Forward Moving Average Representation

In this section we establish a new approach for obtaining the forward MA representation for multivariate MA(1) processes by the right factorization of the spectral density.

A q-ssp $(\mathbf{x}_n)$ is said to be moving average if

$$\mathbf{x}_n = \sum_{k=-\infty}^{\infty} \mathbf{A}_k \mathbf{u}_{n-k}, \quad \text{for every } n \in \mathbb{Z}, \tag{3.1}$$

where $(\mathbf{u}_n)_{n\in\mathbb{Z}}$ is a q-ssp for which $(\mathbf{u}_m, \mathbf{u}_n) = \delta_{mn}\mathbf{U},\ \mathbf{U} \neq \mathbf{0}$. Let us bring the following theorem from Wiener and Masani (1957, Theorem 7.7) concerning moving average processes. The theorem as appeared in the cited reference has parts (a) and (b); part (c) is added here for the purpose of this work.

Theorem 3.1. (a) The moving average process $(\mathbf{x}_n)$, given by (3.1) has an absolutely continuous spectral distribution $\mathbf{F}$ such that

$$\mathbf{F}'(e^{i\theta}) = \mathbf{\Phi}(e^{i\theta})\mathbf{\Phi}^*(e^{i\theta}), \quad \mathbf{\Phi}(e^{i\theta}) = \sum_{k=-\infty}^{\infty} \mathbf{A}_k \mathbf{U}^{\frac{1}{2}} e^{ik\theta}, \ \ a.e.\ \theta.$$

(b) If for this process, $\mathbf{A}_k = 0$ for $k < 0$, then

$$\mathbf{x}_n = \sum_{k=-\infty}^{n} \mathbf{A}_k \mathbf{u}_{n-k}, \ \ \mathbf{\Phi}(e^{i\theta}) = \sum_{k=0}^{\infty} \mathbf{A}_k \mathbf{U}^{\frac{1}{2}} e^{ik\theta},$$

and either $\triangle\mathbf{\Phi}_+$ ($\mathbf{\Phi}_+$ is the analytic extension of $\mathbf{\Phi}$ to the open unit disc) vanishes identically, or $\log\triangle\mathbf{F}'$ is absolutely integrable, w.r.t. the Lebesgue measure, on $\{|z| = 1\}$ and

$$\log\triangle(\mathbf{A}_0\mathbf{U}\mathbf{A}_0^*) = \frac{1}{2\pi}\int_0^{2\pi} \log\triangle\{\mathbf{F}'(e^{i\theta})\}d\theta,$$

where $\triangle$ stands for the determinant of matrix.
(c) If for this process, $\mathbf{A}_k = 0$ for $k > 0$, then

$$\mathbf{x}_n = \sum_{k=0}^{+\infty} \mathbf{A}_{-k}\mathbf{u}_{n+k}, \ \mathbf{\Phi}(e^{i\theta}) = \sum_{k=0}^{\infty} \mathbf{A}_{-k}\mathbf{U}^{\frac{1}{2}} e^{-ik\theta},$$

and the rest of the conclusion of (b) is fulfilled.

We refer to the process in (b) as "backward" moving average, and to the process in (c) as "forward" moving average. Clearly a non-trivial moving average can not be backward and forward with respect

to the same orthogonal process. But if it is, with respect to orthogonal processes $(\mathbf{u}_n)$ and $(\mathbf{v}_n)$ respectively, then it will follow from Theorem 3.1 that

$$\mathbf{F}'(e^{i\theta}) = \mathbf{\Phi}(e^{i\theta})\mathbf{\Phi}^*(e^{i\theta}), \quad \mathbf{\Phi}(e^{i\theta}) = \sum_{k=0}^{+\infty} \mathbf{\Phi}_k e^{ik\theta}, \ \ a.e.\ \theta, \tag{3.2}$$

and

$$\mathbf{F}'(e^{i\theta}) = \mathbf{\Psi}(e^{i\theta})\mathbf{\Psi}^*(e^{i\theta}), \quad \mathbf{\Psi}(e^{i\theta}) = \sum_{k=0}^{+\infty} \mathbf{\Psi}_k e^{-ik\theta}, \ \ a.e.\ \theta. \tag{3.3}$$

where $\mathbf{\Phi}_k = \mathbf{A}_k \mathbf{U}^{\frac{1}{2}}$ and $\mathbf{\Psi}_k = \mathbf{A}_{-k}\mathbf{V}^{\frac{1}{2}}$; $(\mathbf{v}_m, \mathbf{v}_n) = \delta_{mn}\mathbf{V}$, $\mathbf{V} \neq \mathbf{0}$.

Upon deriving the spectral factors $\mathbf{\Phi}$ and $\mathbf{\Psi}$, the orthonormal processes $\mathbf{u} = (\mathbf{u}_k)_{-\infty}^{+\infty}$ and $\mathbf{v} = (\mathbf{v}_k)_{-\infty}^{+\infty}$ are constructed through

$$\mathbf{u}_n = \int_0^{2\pi} e^{-inx} d\beta(x), \ \ \beta(A) = \int_A \mathbf{\Phi}^{-1}(\theta)\mathbb{E}(d\theta)\mathbf{x}_0, \ \ A \subset [0, 2\pi), \tag{3.4}$$

and

$$\mathbf{v}_n = \int_0^{2\pi} e^{-in\theta} d\gamma(\theta), \quad \gamma(A) = \int_A \mathbf{\Psi}^{-1}(\theta)\mathbb{E}(d\theta)\mathbf{x}_0, \quad A \subset [0, 2\pi), \tag{3.5}$$

where $\mathbb{E}(\theta)$ is the spectral resolution of the shift operator of the process $\mathbf{x}$, Wiener and Masani (1957, page 140). Furthermore we deduce that

$$(\mathbf{u}_n, \mathbf{v}_m) = \int_0^{2\pi} e^{-i(n-m)\theta} \mathbf{\Phi}^{-1}(\theta)\mathbf{F}'(\theta)\mathbf{\Psi}^{-1}(\theta)^* = (\mathbf{\Psi}^{-1}\mathbf{\Phi})^*_{n-m}. \tag{3.6}$$

Theorem 3.2. Assume $\mathbf{G}_{\mathbf{\Lambda}}$ is a matrix valued function

$$\mathbf{G}_{\mathbf{\Lambda}}(\theta) = \mathbf{I} - \mathbf{\Lambda} e^{i\theta} - \mathbf{\Lambda}^* e^{-i\theta}. \tag{3.7}$$

(a) $\mathbf{G_\Lambda}$ can be factorized as

$$\mathbf{G_\Lambda}(\theta) = \mathbf{\Phi_\Lambda}(\theta)\mathbf{\Phi^*_\Lambda}(\theta), \tag{3.8}$$

where

$$\mathbf{\Phi_\Lambda}(\theta) = (\mathbf{I} + \mathbf{A}\mathbf{A}^*)^{-\frac{1}{2}}\,(\mathbf{I} - \mathbf{A}e^{i\theta}), \tag{3.9}$$

and $\mathbf{A}$ is subject to

$$\mathbf{\Lambda} = (\mathbf{I} + \mathbf{A}\mathbf{A}^*)^{-\frac{1}{2}}\,\mathbf{A}\,(\mathbf{I} + \mathbf{A}\mathbf{A}^*)^{-\frac{1}{2}}\,. \tag{3.10}$$

In addition if

$$\triangle\,(\mathbf{I} + \mathbf{A}\mathbf{A}^*) = \exp\{-\frac{1}{2\pi}\int_0^{2\pi} \log \triangle(\mathbf{G_\Lambda}(\theta))d\theta, \tag{3.11}$$

then (3.8) is a left factorization for $\mathbf{G_\Lambda}$.

(b) If (3.8) is a left factorization for $\mathbf{G_\Lambda}$, then

$$\mathbf{G_\Lambda}(\theta) = \mathbf{\Psi_\Lambda}(\theta)\mathbf{\Psi^*_\Lambda}(\theta), \tag{3.12}$$

where

$$\mathbf{\Psi_\Lambda}(\theta) = (\mathbf{I} + \mathbf{B}\mathbf{B}^*)^{-\frac{1}{2}}\,(\mathbf{I} - \mathbf{B}e^{-i\theta}), \tag{3.13}$$

and $\mathbf{B}$ is subject to

$$\mathbf{\Lambda}^* = (\mathbf{I} + \mathbf{B}\mathbf{B}^*)^{-\frac{1}{2}}\,\mathbf{B}\,(\mathbf{I} + \mathbf{B}\mathbf{B}^*)^{-\frac{1}{2}}\,, \tag{3.14}$$

will be a right factorization for $\mathbf{G_\Lambda}$;
consequently $\triangle\,(\mathbf{I} + \mathbf{A}\mathbf{A}^*) = \triangle\,(\mathbf{I} + \mathbf{B}\mathbf{B}^*)$.

Proof. (a) The factorization (3.8) can be verified directly. This factorization is indeed a left factorization according to Wiener and Masani (1957, Theorem 7.13) if

$$\begin{aligned}\triangle\{\mathbf{\Phi_{\Lambda_+}}(0)\}^{-2} &= \triangle\left((\mathbf{I} + \mathbf{A}\mathbf{A}^*)^{-\frac{1}{2}}\right)^{-2} = \triangle\,(\mathbf{I} + \mathbf{A}\mathbf{A}^*) \\ &= \exp\{-\frac{1}{2\pi}\int_0^{2\pi} \log \triangle(\mathbf{G_\Lambda}(\theta))d\theta\}.\end{aligned}$$

(b) Note that $\mathbf{G}_{\mathbf{\Lambda}}(-\theta) = \mathbf{G}_{\mathbf{\Lambda}^*}(\theta)$, and apply Theorem 3.3 in Soltani and Mohammadpour (2005). Consequently (3.12) will be a right analytic factorization for $\mathbf{G}_{\mathbf{\Lambda}}$. Moreover since

$$\int_0^{2\pi} \log \triangle(\mathbf{G}_{\mathbf{\Lambda}}(-\theta))d\theta = \int_0^{2\pi} \log \triangle(\mathbf{G}_{\mathbf{\Lambda}}(\theta))d\theta,$$

and

$$\mathbf{\Psi}_{\mathbf{\Lambda}}(-\theta) \mid_{\theta=0} = \mathbf{\Psi}_{\mathbf{\Lambda}}(\theta) \mid_{\theta=0} = \mathbf{\Psi}_{\mathbf{\Lambda}}(0),$$

then from Soltani and Mohammadpour(2005, Theorem 3.3) and Wiener and Masani (1957, Theorem 7.13)

$$\triangle\{\mathbf{\Phi}_{\mathbf{\Lambda}_+}(0)\}^2 = \triangle\{\mathbf{\Psi}_{\mathbf{\Lambda}_+}(0)\}^2;$$

giving that $\triangle(\mathbf{I} + \mathbf{A}\mathbf{A}^*) = \triangle(\mathbf{I} + \mathbf{B}\mathbf{B}^*)$.□

Theorem 3.3. Let $(\mathbf{x}_n)_{n\in\mathbb{Z}}$ be a q-variate MA(1) process

$$\mathbf{x}_n = \mathbf{u}_n - \mathbf{\Theta}\mathbf{u}_{n-1},$$

where $(\mathbf{u}_n)_{n\in\mathbb{Z}}$ is the orthogonal processes defined by (3.4) and $\triangle(\mathbf{I} - \mathbf{\Theta}z) \neq 0$, for $|z| \leq 1$.
Then the right analytic factorization of $\mathbf{F}'(\theta)$ can be obtained as

$$\mathbf{F}'(\theta) = \mathbf{\Psi}(\theta)\mathbf{\Psi}^*(\theta), \tag{3.15}$$

where

$$\mathbf{\Psi}(\theta) = (\mathbf{I} + \mathbf{\Theta}\mathbf{\Theta}^*)^{\frac{1}{2}} (\mathbf{I} + \mathbf{B}\mathbf{B}^*)^{-\frac{1}{2}} (\mathbf{I} - \mathbf{B}e^{-i\theta}), \tag{3.16}$$

and the matrix $\mathbf{B}$ is subject to

$$(\mathbf{I} + \mathbf{\Theta}\mathbf{\Theta}^*)^{-\frac{1}{2}} \mathbf{\Theta}^* (\mathbf{I} + \mathbf{\Theta}\mathbf{\Theta}^*)^{-\frac{1}{2}} = (\mathbf{I} + \mathbf{B}\mathbf{B}^*)^{-\frac{1}{2}} \mathbf{B} (\mathbf{I} + \mathbf{B}\mathbf{B}^*)^{-\frac{1}{2}}. \tag{3.17}$$

Proof. The left analytic factorization of the spectral density $\mathbf{F}'(\theta)$,

$$\mathbf{F}'(\theta) = \mathbf{I} + \mathbf{\Theta}\mathbf{\Theta}^* - \mathbf{\Theta}e^{i\theta} - \mathbf{\Theta}^* e^{-i\theta}.$$

is

$$\mathbf{F}'(\theta) = \mathbf{\Phi}(\theta)\mathbf{\Phi}^*(\theta),$$

where

$$\mathbf{\Phi}(\theta) = (\mathbf{I} - \mathbf{\Theta}e^{i\theta}).$$

Now the spectral density $\mathbf{F}'(\theta)$ can be written as

$$\mathbf{F}'(\theta) = (\mathbf{I} + \mathbf{\Theta}\mathbf{\Theta}^*)^{\frac{1}{2}}\, \mathbf{G}(\theta)\, (\mathbf{I} + \mathbf{\Theta}\mathbf{\Theta}^*)^{\frac{1}{2}}\,,$$

where

$$\mathbf{G}(\theta) = \mathbf{I} - \mathbf{\Lambda}e^{i\theta} - \mathbf{\Lambda}^* e^{-i\theta},$$

and

$$\mathbf{\Lambda} = (\mathbf{I} + \mathbf{\Theta}\mathbf{\Theta}^*)^{-\frac{1}{2}}\, \mathbf{\Theta}\, (\mathbf{I} + \mathbf{\Theta}\mathbf{\Theta}^*)^{-\frac{1}{2}}\,.$$

Now by applying Theorem 3.2,

$$\mathbf{G}(-\theta) = \mathbf{\Psi}_{\mathbf{G}}(\theta)\mathbf{\Psi}^*_{\mathbf{G}}(\theta),$$

where

$$\mathbf{\Psi}_{\mathbf{G}}(\theta) = (\mathbf{I} + \mathbf{B}\mathbf{B}^*)^{-\frac{1}{2}}\, (\mathbf{I} - \mathbf{B}e^{i\theta}),$$

and

$$\mathbf{\Lambda}^* = (\mathbf{I} + \mathbf{B}\mathbf{B}^*)^{-\frac{1}{2}}\, \mathbf{B}\, (\mathbf{I} + \mathbf{B}\mathbf{B}^*)^{-\frac{1}{2}}\,.$$

Thus $\mathbf{B}$ is subject to

$$(\mathbf{I} + \mathbf{\Theta}\mathbf{\Theta}^*)^{-\frac{1}{2}}\, \mathbf{\Theta}^*\, (\mathbf{I} + \mathbf{\Theta}\mathbf{\Theta}^*)^{-\frac{1}{2}} = (\mathbf{I} + \mathbf{B}\mathbf{B}^*)^{-\frac{1}{2}}\, \mathbf{B}\, (\mathbf{I} + \mathbf{B}\mathbf{B}^*)^{-\frac{1}{2}}\,. \tag{3.18}$$

Therefore

$$\begin{aligned}
\mathbf{F}'(-\theta) &= (\mathbf{I} + \mathbf{\Theta}\mathbf{\Theta}^*)^{\frac{1}{2}}\, \mathbf{G}(-\theta)\, (\mathbf{I} + \mathbf{\Theta}\mathbf{\Theta}^*)^{\frac{1}{2}} \\
&= (\mathbf{I} + \mathbf{\Theta}\mathbf{\Theta}^*)^{\frac{1}{2}}\, (\mathbf{I} + \mathbf{B}\mathbf{B}^*)^{-\frac{1}{2}}\, (\mathbf{I} - \mathbf{B}e^{i\theta}) \\
&\quad (\mathbf{I} - \mathbf{B}^* e^{-i\theta})\, (\mathbf{I} + \mathbf{B}\mathbf{B}^*)^{-\frac{1}{2}}\, (\mathbf{I} + \mathbf{\Theta}\mathbf{\Theta}^*)^{\frac{1}{2}}\,,
\end{aligned}$$

and the right analytic factorization of $\mathbf{F}'(\theta)$ is

$$\mathbf{\Psi}(\theta) = (\mathbf{I} + \mathbf{\Theta}\mathbf{\Theta}^*)^{\frac{1}{2}}\, (\mathbf{I} + \mathbf{B}\mathbf{B}^*)^{-\frac{1}{2}}\, (\mathbf{I} - \mathbf{B}e^{-i\theta}).\square$$

Theorem 3.3. Let $\mathbf{X}$ and $\mathbf{Y}$ be the matrices that satisfies

$$\left[\ \mathbf{I}\otimes\mathbf{I}\quad(\mathbf{\Theta}^*\otimes\mathbf{\Theta}^*)\ \right]\left[\begin{array}{c}\mathrm{vec}(\mathbf{X})\\ \mathrm{vec}(\mathbf{Y})\end{array}\right]=\mathrm{vec}(\mathbf{I}+\mathbf{\Theta}\mathbf{\Theta}^*). \tag{3.19}$$

Then the matrix $\mathbf{B}$ in (3.16) is given by

$$\mathbf{B}=\mathbf{C}^{-1}\mathbf{\Theta}^*\mathbf{C}^{-1*}, \tag{3.20}$$

where $\mathbf{X}=\mathbf{C}\mathbf{C}^*$, $\mathbf{Y}=\mathbf{C}^{-1^*}\mathbf{C}^{-1}$ and $\mathbf{X}=\mathbf{Y}^{-1}$.

Proof. To find $\mathbf{B}$ satisfy (3.17), let us try $\mathbf{B}=\mathbf{C}^{-1}\mathbf{\Theta}^*\mathbf{C}^{-1*}$. Then from (3.15)

$$\mathbf{F}'(\theta)=\mathbf{\Psi}(\theta)\mathbf{\Psi}^*(\theta)$$

where

$$\mathbf{\Psi}(\theta)=\mathbf{C}(\mathbf{I}-\mathbf{C}^{-1}\mathbf{\Theta}^*\mathbf{C}^{-1*}e^{-i\theta}).$$

Thus

$$\begin{aligned}\mathbf{F}'(\theta) &= \mathbf{C}\left(\mathbf{I}+(\mathbf{C}^{-1}\mathbf{\Theta}^*\mathbf{C}^{-1*})(\mathbf{C}^{-1}\mathbf{\Theta}^*\mathbf{C}^{-1*})^*\right.\\ &\quad\left.-\mathbf{C}^{-1}\mathbf{\Theta}^*\mathbf{C}^{-1*}e^{-i\theta}-\mathbf{C}^{-1}\mathbf{\Theta}\mathbf{C}^{-1*}e^{i\theta}\right)\mathbf{C}^*\\ &= \mathbf{C}\mathbf{C}^*+\left(\mathbf{\Theta}^*\mathbf{C}^{-1*}\right)\left(\mathbf{C}^{-1}\mathbf{\Theta}\right)-\mathbf{\Theta}^*e^{-i\theta}-\mathbf{\Theta}e^{i\theta},\end{aligned}$$

and $\mathbf{C}$ is subject to

$$\mathbf{C}\mathbf{C}^*+\left(\mathbf{C}^{-1}\mathbf{\Theta}\right)^*\left(\mathbf{C}^{-1}\mathbf{\Theta}\right)=\mathbf{I}+\mathbf{\Theta}\mathbf{\Theta}^*. \tag{3.21}$$

But

$$\mathbf{C}\mathbf{C}^*+\left(\mathbf{C}^{-1}\mathbf{\Theta}\right)^*\left(\mathbf{C}^{-1}\mathbf{\Theta}\right)=\left(\mathbf{C}+\left(\mathbf{C}^{-1}\mathbf{\Theta}\right)^*\right)\left(\mathbf{C}+\left(\mathbf{C}^{-1}\mathbf{\Theta}\right)^*\right)^*-\mathbf{\Theta}-\mathbf{\Theta}^*,$$

so

$$\begin{aligned}\left(\mathbf{C}+\left(\mathbf{C}^{-1}\mathbf{\Theta}\right)^*\right)\left(\mathbf{C}+\left(\mathbf{C}^{-1}\mathbf{\Theta}\right)^*\right)^* &= \mathbf{I}+\mathbf{\Theta}\mathbf{\Theta}^*+\mathbf{\Theta}+\mathbf{\Theta}^*\\ &= (\mathbf{I}+\mathbf{\Theta})(\mathbf{I}+\mathbf{\Theta})^*. \end{aligned}\tag{3.22}$$

Now by applying Remark 3.1, (3.21) can be written as

$$(\mathbf{I}\otimes\mathbf{I})\,\mathrm{vec}(\mathbf{X})+(\mathbf{\Theta}^*\otimes\mathbf{\Theta}^*)\mathrm{vec}(\mathbf{Y})=\mathrm{vec}(\mathbf{I}+\mathbf{\Theta}\mathbf{\Theta}^*),$$

or

$$\left[\begin{array}{cc}\mathbf{I}\otimes\mathbf{I} & (\mathbf{\Theta}^*\otimes\mathbf{\Theta}^*)\end{array}\right]\left[\begin{array}{c}\mathrm{vec}(\mathbf{X})\\ \mathrm{vec}(\mathbf{Y})\end{array}\right]=\mathrm{vec}(\mathbf{I}+\mathbf{\Theta}\mathbf{\Theta}^*),\qquad(3.23)$$

where $\mathbf{X}=\mathbf{C}\mathbf{C}^*$ and $\mathbf{Y}=(\mathbf{C}\mathbf{C}^*)^{-1}$. By solving this equation and decomposing $\mathbf{Y}$ as $\mathbf{C}^{-1^*}\mathbf{C}^{-1}$, the matrix $\mathbf{C}^{-1}$ will be found that satisfies(3.21). Consequently $\mathbf{B}$ can be deduced from (3.20) that satisfies (3.18). □

4 Numerical Evaluations

The Theorem 3.3 provides a practical method to derive the right analytic factorization from the left analytic factorization. In this section we put this result into practise by solving (3.23) numerically.
A computer program is written in S-plus. The input of the program is the matrix $\mathbf{\Theta}$, the MA(1) coefficient, and the output is the matrix $\mathbf{B}$ in (3.20), from which the right analytic factorization can be easily deduced by (3.16).

Example 3.1. Consider the bivariate series defined by

$$\begin{aligned}x_n^{(1)} &= u_n^{(1)},\\ x_n^{(2)} &= \phi u_{n-1}^{(1)}+u_n^{(2)},\end{aligned}$$

where $\phi>0$ and $(\mathbf{u}_n)$ is an orthogonal sequence in $\mathcal{L}_2$ with $(\mathbf{u}_n,\mathbf{u}_n)=\sigma^2\mathbf{I}$, $n\in\mathbb{Z}$.
The bivariate $(\mathbf{x}_n)$ is indeed MA(1) process,

$$\mathbf{x}_n=\mathbf{u}_n+\mathbf{\Theta}\mathbf{u}_{n-1},$$

where $\mathbf{\Theta}=\left[\begin{array}{cc}0 & 0\\ \phi & 0\end{array}\right]$, $(\mathbf{u}_n)$ is an orthogonal sequence which admits the spectral representation (3.4).

So the spectral density of $\mathbf{x}_n$ is given by

$$\mathbf{F}'(\theta) = (\mathbf{I} + \mathbf{\Theta}e^{i\theta})(\mathbf{I} + \mathbf{\Theta}e^{i\theta})^*$$

and the left analytic factorization is

$$\mathbf{\Phi}(\theta) = \mathbf{I} + \mathbf{\Theta}e^{i\theta}.$$

Now the right factorization of the spectral density obtains by applying (3.22) and (3.16) as

$$\mathbf{\Psi}(\theta) = \begin{bmatrix} \frac{1}{\sqrt{1+\phi^2}} & 0 \\ 0 & \sqrt{1+\phi^2} \end{bmatrix} + \begin{bmatrix} 0 & \frac{\phi}{\sqrt{1+\phi^2}} \\ 0 & 0 \end{bmatrix} e^{-i\theta}.$$

Thus $(\mathbf{x}_n)$ admits the forward MA representation

$$\mathbf{x}_n = \mathbf{\Psi}_0\mathbf{v}_n + \mathbf{\Psi}_1\mathbf{v}_{n+1},$$

where $(\mathbf{v}_n)$ is an orthogonal sequence in $\mathcal{L}_2$, with $(\mathbf{v}_n, \mathbf{v}_n) = \sigma^2\mathbf{I}$, $n \in \mathbb{Z}$, admitting the spectral representation (3.5), and

$$\mathbf{\Psi}_0 = \begin{bmatrix} \frac{1}{\sqrt{1+\phi^2}} & 0 \\ 0 & \sqrt{1+\phi^2} \end{bmatrix}, \qquad \mathbf{\Psi}_1 = \begin{bmatrix} 0 & \frac{\phi}{\sqrt{1+\phi^2}} \\ 0 & 0 \end{bmatrix}.$$

Clearly $\mathbf{\Phi}_0 \neq \mathbf{\Psi}_0$, $\mathbf{\Phi}_1 \neq \mathbf{\Psi}_1$.

References

[1] Bini, D. A., Fiorentino, G. , Gemignani, L. and Meini, B. (2003) Effective fast algorithms for polynomial spectral factorization. *Numerical Algorithms* 34, 217-227.

[2] Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods*. Springer-Verlag, New York.

[3] Cheng, R. and Pourahmadi, M. (1993) Baxter's inequality and convergence of finite predictors of multivariate stochastic processes. *Probab. Theory Relat. Fields* 95, 115-124.

[4] Findley, D. (1991) Convergence of finite multistep predictors from incorrect models and its role in model selection. *Note di Matematica* XI, 145-155.

[5] Hannan, E. J. and Diestler, M. (1988) *The Statistical Theory of Linear Systems.* John Wiley.

[6] Miamee, A. G. (1988) On the Weiner-Masani algorithm for finding the generating function of multivariate stochastic processes. *The Annals of Probability* 16, 4, 1854-1859.

[7] Pourahmadi, M. (2001) *Foundation of Time Series Analysis and Prediction Theory.* John Wiley. New York.

[8] Pourahmadi, M. (1992) Alternating projections and interpolation of stationary processes. *J. Appl. Prob.* 29, 921-931.

[9] Rozanov, Yu. A. (1967) *Stationary Random Processes,* San Francisco: Holden-Day.

[10] Soltani, A. R. and Yaghmaee, F. (2001) An algorithm to interpolate noise in second order stationary sequences. *Journal of Applied Statistical Science* 10, 3, 223-235.

[11] Tuel, W. G. (1968) Computer algorithm for spectral factorization of rational matrice. *IBM Journal of Research and Development* 12, 2*, 163-170.*

[12] Wiener, N. and Masani, P. (1958) The prediction theory of multivariate stochastic processes. *Part II. Acta Math.* 99, 93-137.

[13] Wiener, N. and Masani, P. (1957) The prediction theory of multivariate stochastic processes. *Part I. Acta Math.* 98, 111-150.

Efficient Adaptive Sampling Designs for Assessing Marine Life Stacks

M. Mohammad Salehi*

Department of Mathematical Science, Isfahan University of Technology

Abstract: Marine life populations often form large, widely scattered schools with few species in between. For rare species of Marine life may form small groups that are hard to find. Applying standard sampling methods such as simple random sampling of plots to such a population could yield little information, with most of the plots being empty. Adaptive sampling designs can reasonably deal with this problem. They are based on the simple idea that when some fishes are located on a sample plot, more plots in the vicinity of the sample plot are added to the sample. The hope is to find more fishes. Adaptive allocation sampling is a class of adaptive sampling which has been used in marine life study in recent years. We first review successful adaptive sampling designs applied in marine studies. We then introduce a simple adaptive sampling design. In a simulation study, we show that this sampling design is an efficient sampling design for fish populations.

Keywords: Sampling design; Clump fish population; Abundance estimation.

1 Introduction

During last two decades, adaptive sampling designs have been drawn researchers attention in fishery industry. Adaptive samplings (Thompson and Seber 1996) are appealing because they mimic how biologists would like to collect data - at least more so than most statistical sampling techniques (Sparre and Venema 1992). When adaptively sampling, biologists search for a species of interest at predetermined locations, and if the species is found then searching continues nearby. Additional appeal of adaptive samplings is attributed to its statistical properties. In general, we may categorize adaptive samplings used in fisheries studies into classes 1) adaptive cluster samplings and 2)

Invited speaker

adaptive allocation samplings. Salehi (2001 [a] and [b]) reviewed and developed density estimators for adaptive cluster samplings which is mote applicable to fish populations. Most recent review is done by Seber and Salehi (2004) in the biostatistics encyclopedia on this subject. In this paper, we focus on those adaptive allocation samplings which are shown to be useful in Marine and fishery studies. We also introduce a new adaptive allocation sampling design, called Complete Adaptive Allocation Sampling (CAAS). Using a simulation study, we show that it can be an efficient sampling design for fish populations.

In section 2, we review development of an adaptive allocation sampling, called two-phase adaptive stratified sampling introduced by (Francis, 1984). In section 3, two-stage sequential sampling (Salehi and Smith, 2005), which is an adaptive allocation sampling, will be reviewed. In section 4, Complete allocation stratified sampling will be introduced and its properties will assess with simulating a rockfish population.

2 Two-phase adaptive stratified random sampling

Stratified random sampling is one the more commonly used methods for estimating density and other characteristics of a biological population within a defined geographical area (Cochran 1977:Chapter 5; Thompson 2002:Chapter 11). With this method the total area is divided into a number of discrete sub-areas or strata such that within each of these strata the variable Y of interest (e.g., the number of individuals per square meter) is expected to be relatively constant in comparison with the variation that exists over the entire area. A simple random sample is then taken from each stratum and the mean or total of Y is estimated for that part of the population. An unbiased estimator of the overall population mean of X then is obtained by averaging the means for the individual strata, weighting them by the areas involved. An unbiased estimator of the overall population total is simply the sum of the estimates for the different strata.

The motivation for using stratified sampling is that if Y has a

small variance within each of the strata then the stratum means and totals will be estimated accurately even with small samples. Hence the population means and total of Y also will be estimated accurately-Stratified sampling is therefore a tool for eliminating effects of some of the population variation in Y from the estimation of the population parameters.

If enough is known about the characteristics of a population, the sample allocation to strata can be optimized. Suppose that there are H strata, with N_h sample units in stratum h, and with the standard deviation of Y equal to σ_h for these units. Also, let n_h denote the sample size in stratum h, and $n = \sum_{h=1}^{H} n_h$ denote the total sample size in all strata. Then it is well known (Sparre and Venema, 1992; page 207) that the variance of the estimator of the population mean or total of Y is minimized for a fixed total sample size when the sample size in stratum h is set at

$$n_h = \frac{n_h \sigma_h}{\sum_{h=1}^{H} n_h \sigma_h}. \tag{1}$$

When estimating the density of an animal or plant population the stratification employed often will be based on habitat characteristics that are thought to be associated with the presence of the organism. At other times strata may simply be geographical areas because detailed habitat information is not available or the habitat associations of the organism are not well understood. In either case, the practical utility of equation (1) will be limited because only rough guesses will be available for the within-stratum standard deviations of Y.

It is often the case that all that can be done is to assume that the standard deviation will be low in areas with a low density of fishes and high in areas with a high density of fishes. The sample then is allocated to strata based on guessed relative mean densities and the assumption that standard deviations are, for example, proportional to the mean densities. In practice, therefore, the sample allocation to strata may be far from optimal, even with a good stratification of the geographical area with respect to the variable Y.

To overcome this problem in the context of fisheries stock assessment, Francis (1981) suggested a two-phase approach to sampling, with a similar scheme offered independently by Jolly and Hampton (1990). Francis proposed that a first-phase, stratified random sample should be taken using the best available information for the choice of strata and the allocation of the sample, based on equation (1). Then, using the results obtained from the first phase, a second sample should be taken to increase the number of sample units in those strata where this is expected to be most effective for reducing the variance of the estimator of the population mean or total. The idea, therefore, is to use the information obtained from the first-phase sample to compensate in the second phase for any shortcomings in the sample allocation.

Francis (1984) was concerned only with the estimation of one fish species in one fishery. However, his two-phase adaptive sampling method is easily generalized for use in situations where there are several populations to be estimated at several different locations. Manly et al. (2002) described the following example, involved the need to estimate abundance of three species of shellfish at 11 beaches around Auckland, New Zealand.

In this case a limited amount of sampling effort was available, and there was a need, as far as possible, to get good estimates of shellfish numbers for all three species on all 11 beaches. To achieve this objective, the first phase of sampling consisted of taking 75 transect samples from each of the 11 beaches with an appropriate stratification for each beach based on local knowledge. At the second phase another 145 transects were sampled. These were allocated out to the strata within beaches in such a way that the average of the coefficients of variation (CV) over 20% reduced as much as possible for the estimates required from the study (the population size for each of the shellfish species for each of the beaches where the species was present).

As proposed by Francis (1984) for one population at one location the first phase of the two-phase, adaptive stratified sampling design involves taking a conventional random stratified sample of size n_I from the population, with the sample sizes in the different strata chosen to

approximate the optimal allocation of equation (1), based on whatever prior knowledge is available. From the first phase data, the variances in the different strata can be estimated; these are assumed to be good approximations for the true variances. These variances are used to determine how best to allocate the second-phase sample of size n_{II} to the strata.

In choosing the fraction of the total sample to allocate to the first phase of the survey there must be a compromise between having large enough to give good estimates of variance and having n_I large enough to make effective use of the information from the first phase. As a rule of thumb, allocating 75 % of the units to the first phase and 25 % to the second phase seems reasonable (Francis 1984).

At the second phase the sample n_{II} units are allocated out one by one to the strata. The first unit is allocated to the stratum where its use will give the largest reduction in the variance of the estimated population total or mean. The second unit is allocated to the stratum where its use will give the largest reduction in the variance of the estimated total or mean, given the allocation of the first extra unit, and so on. The process of adding units continues until all n_{II} units have been allocated. No actual sampling is carried out during this second-phase allocation exercise, so the variances assumed for the strata remain equal to the estimates from the first-phase sample throughout the allocation process. Following the allocation, the extra second-phase sampling proceeds. The data then are analyzed as if they came from a conventional stratified random sample with a total size of $n_I = n_{II}$.

Treating the results of a two-phase stratified sample as a conventional stratified sample leads to a negative bias in the estimators of population mean and total. However/ this generally seems quite small in comparison with the standard error (Francis 1984, 1991; Jolly and Hampton 1990, 1991). As demonstrated by Brown (1999), the two-phase design generally has better properties than the alternative adaptive cluster sampling method that was proposed by Thompson (1990), except for very highly clustered populations. Francis (1984) considered

two algorithms for two-phase sampling. The first involved estimating within-stratum variances with their sample values in the usual way and the second involved assuming that the within-stratum variances are proportional to the squares of the means for these strata. The second algorithm seems to be appropriate for fisheries stock assessment, but only the first algorithm is considered by Manly et. al. (2002) because of its more general applicability for populations in which the assumption that the variance is proportional to the mean squared may not be reasonable.

With Francis' original sampling design, there was only one observation on each sample unit, where this might typically be the population density per unit area for a single species. Manly et.al. (2002) generalized it to the case in which each sample unit provides $K > 1$ observations, where K is number of species. For example, these observations might be the densities per unit area of the S species present in the region. The species then would represent several populations.

As a further generalization it can be assumed that the populations of interest occur at L different locations and that there is the need to estimate the mean or the total for each population at each location. For example, with the Auckland area shellfish study mentioned earlier, there were $L = 11$ beaches (the locations), at each of which there was one or more of the $K = 3$ species of shellfish (the populations). It was the individual beaches that were stratified with the same strata applying for each of the shellfish species.

Manly et al, (2002) carried out a simulation study to examine the proper- ties of the generalized two-phase adaptive stratified sampling design. This was based on real data from stratified surveys carried out in the past for the New Zealand Ministry of Fisheries. Ten model populations were set up using quadrat counts along transects from the past surveys, to represent 10, different locations. The quadrat counts were put in order based on their geographic locations within strata and then bootstrap resampled with replacement to produce new sets of data. There were two or three shellfish species recorded for each of the model populations namely pipis, cockles, and wedge shells. The

densities varied considerably among strata within locations and among locations. They concluded:

- The estimated CV varied considerably for the estimation of different abundances, but the mean estimated CV was not very sensitive to the optimization criterion used.
- The situation for the CVs actually obtained for estimation was similar to that for the estimated CVs in the sense that the different criteria produced about the same results, which suggested that the criterion used was not critical.
- Percentage biases were relatively small (less than 5%) in general unless the corresponding CVs were large.
- For CVs that were not too large (i.e., about 25% or less), the estimated CVs were nearly equal to the true CVs on average, but for situations where the true CV was large, there was a definite tendency for the estimated CV to be too low.
- In comparison to stratified sampling with proportional allocation, the two-phase sampling process reduced the highest CVs and increased the lowest CVs (which is what was expected to happen).

Two-phase adaptive stratified random sampling is a useful design for situations in which the density of a plant or animal population is likely to vary with defined strata based on the habitat or general geographical areas but it is not possible to know in advance of sampling where the high densities are likely to occur. Except with very clustered populations/ it is likely to give better estimation than the alternative adaptive sampling method of Thompson (1990), which is usually applied to rare populations. Two-phase adaptive stratified random sampling often will be easier to use than this alternative because there is no need to make the correct choices for sampling parameters like the trigger level for further sampling. The generalized version of two-phase adaptive sampling has proven to be particularly

useful for surveys in which one needs to estimate population sizes for multiple species at multiple locations at the same time.

3 Two Stage sequential sampling

A neighborhood-free adaptive allocation sampling design, called two-stage sequential sampling (TSS) was introduced by (Salehi and Smith, 2005). It has been shown to perform well on a variety of populations compared to conventional and neighborhood-based adaptive cluster sampling designs. In the TSS design, an initial sample of secondary units (n_1) is selected within a sample of primary units. A condition is evaluated independently within each primary unit. If the condition is met then an additional sample of secondary units (n_2) is selected, but sampling stops there regardless of observations in the n2 units. So, under the TSS design the final sample size is restricted to be no more than $n_1 + n_2$ in each primary unit.

They applied the TSS to three freshwater mussels populations on a section of riffle habitat in the Cacapon River near Capon Bridge, WV, USA. The river section was approximately rectangular and measured approximately 40 m wide (bank to bank) by 90 m long. Density of freshwater mussels is difficult to estimate well because of their tendency to be rare and clustered at some spatial scales (Strayer and Smith 2003). Smith et al. (2003) applied adaptive cluster sampling to low-density populations of freshwater mussels and found that adaptive cluster sampling increased observations of individuals and rare species. However, they also found that sampling of edge units greatly increased effort with little or no gain in efficiency in density estimates (See also Salehi; 1999). Thus, an adaptive sampling procedure that reduces or eliminates sampling of edge units would be of great interest for freshwater mussel population assessments. The three species, Elliptio complanata, E. fisheriana, and Lampsilis cariosa exhibited different spatial distributions. The E.complanata population was not rare, but it was relatively clustered. The E. fisheriana population was relatively rare and clustered. The L. cariosa population was rare, but

not clustered. Application of two-stage sequential sampling to these populations demonstrated design performance over a range of spatial distributions. Two-stage sequential sampling was more efficient for E. fisheriana population, which is a not only rare but also clustered population.

Two-stage sampling is a general case of stratified sampling. Whenever all Primary Sampling Units (PSU's) are selected in the first stage the two-stage sampling become stratified sampling (See Cachran, 1977). Salehi and Smith (2002) concluded when a rare and clustered population is partitioned into PSU's such that there is one cluster per PSU and the PSU is roughly the size of the cluster, some PSU's will be without the rare event and some will have the majority of the rare events. Hence, the variance within PSU's will be less than the variance between PSU's. For such a population, stratified sampling is an efficient design (two-stage sampling with all PSU's being selected). This was true for the E. fisheriana population where greatest efficiency was found when all PSU's were selected at the first-stage (Stratified sampling).

Sampling freshwater mussels is expensive and labor intensive especially in deep water where SCUBA diving is required (Strayer and Smith 2003). In rivers, SCUBA divers must work close to an anchored boat for safety, and repositioning the boat is time consuming. Thus, it is practical to sample in stages by first selecting a primary unit to position the boat and then selecting secondary units for collection of freshwater mussels. If we want to estimate population size of a rare species, then the amount of habitat is an appropriate way to gauge the partitioning of the study area into PSU's. If species distribution is known to be restricted to a small fraction of a site and auxiliary information is available, a stratified sampling could be designed to confine rare units to a few strata (Kalton and Anderson 1986, Christman 2000). However, if species distribution is unknown or relevant auxiliary information is not in the sampling frame, then two-stage sequential sampling can still be an efficient sampling design by partitioning the population in a haphazard fashion for rare and highly

clustered populations.

4 Complete Allocation Stratified Sampling

In this section, we develop a new adaptive allocation based on simple idea of "keep fishing whenever you catch a fish in a region". Consider a population of N nits is partitioned into H strata and there is N_h plots in stratum h $(h = 1, \cdots, H)$. Let y_{hi} is the numbers of species in unit i from stratum h. In step 1 a simple random sample of size is taken without replacement from stratum h, The selected plots are observed. If we observe any species in stratum h all units in this stratum are selected. Otherwise, no more plots are selected from stratum h. Since no specie is observed in latter strata we can ignore them without losing any information. Let π_h be the probability that whole stratum h is selected. Suppose that m_h is the number of nonempty units in stratum h. We now have

$$\pi_h = 1 - \frac{\binom{N_h - m_h}{n_h}}{\binom{N_h}{n_h}}$$

Therefore the Horvitz-Thompson estimator of the total number of species, say T in the area is

$$\hat{T}_{HT} = \sum_{h=1}^{\gamma} \frac{y_h^*}{\pi_h} \tag{2}$$

Where y_h^* is the sum of the y_{hi} for the h^{th} stratum, $\gamma$and is the number of strata for which at least one species is observed in the first wave of sampling.

Its variance is given by

$$Var(\hat{T}_{HT}) = \sum_{h=1}^{H} \frac{(1 - \pi_h) y_h^*}{\pi_h} \tag{3}$$

0	0	0	0	0	0	0	0	126	176	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	741	1585	814	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	714	6643	232	251	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	67	395	1963	303	101	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	59	303	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1132	3153	67	0	0	0	0	0	0	0	0	0	0	0
252	457	243	0	63	271	1949	5411	1179	29	0	0	0	0	0	0	0	0	0	0
1995	4594	1089	82	0	148	240	641	258	0	0	0	0	0	0	0	0	0	0	0
1629	3191	567	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5501	664
543	390	0	0	0	0	0	0	0	0	0	0	0	0	0	0	146	217	2773	1338
452	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5581	4166	2842
0	0	0	0	0	0	0	0	263	261	1593	156	0	0	0	0	0	7	635	639
0	0	0	0	0	0	0	0	135	3238	6065	812	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	67	658	2425	261	0	0	0	0	0	0	0	0

Figure 1: A simulated population of a high-density rockfish population. The numbers in the units is the number of rockfishes.

An unbiased variance estimator is given by

$$\widehat{Var}(\hat{T}_{HT}) = \sum_{h=1}^{\gamma} \frac{(1-\pi_h)y_h^*}{\pi_h^2} \tag{4}$$

To evaluate the precision of (3), we use a simulated population given in Figure (1) which is adopted from Su and Quinn II (2003). It was modeled for rockfish population in the Gulf of Alaska. Catch statistics and trawl surveys showed that many rockfish are aggregated in their distribution in this region. This fish population has $N = 400$ units. Figure (1) shows the abundance in each plot.

We partition the population into equal stratum sizes of$N_h = 2\times 2$, 2×4, 4×2, 2×5, 5×2, 4×5, $5\times 4$and calculate the variances and the relative efficiencies of $\hat{T}$ to simple random sample and stratified sampling with equal sample sizes. The relative efficiencies are given by

$$Eff_{SRS}(\hat{T}) = \frac{Var(\hat{T}_{SRS})}{Var(\hat{T})}, \quad Eff_{ST}(\hat{T}) = \frac{Var(\hat{T}_{ST})}{Var(\hat{T})},$$

Where $Var(\hat{T}_{SRS})$ and $Var(\hat{T}_{ST})$ are variances of estimators of simple random sample and stratified sampling with equal sample sizes.

Since the sample size for Complete Allocation Stratified Sampling (CAST) is random we should calculate effective sample sizes, say ν , which is expected sample size

$$E(n) = \sum_{i=1}^{N} \pi_i = \sum_{H=1}^{H} N_h \pi_h = \nu$$

We then compute the variance of simple random sample of size ν,

$$Var(\hat{T}_{SRS}) = N^2(1 - \frac{\nu}{N})\frac{s^2}{\nu}$$

and the variance of stratified sample of sizes of ν/H from each stratum,

$$Var(\hat{T}_{ST}) = \sum_{h=1}^{H} N_h^2(1 - \frac{\nu}{HN_h})\frac{s_h^2}{\nu}$$

.

The results are given in Table 1 (A), (B), (C), (D), (E), (F), (G), (H) for strata sizes of $N_h = 2 \times 2$, 2×4, 4×2, 2×5, 5×2, 4×5, 5×4 respectively.

The results show that we gain in efficiency for 46 out of 49 cases. The gain in efficiencies were up to stunning 3076% and 3293% over Stratified sample and Simple Random sample, respectively. This was for nh=10 and stratum size of 5×4 . The efficiencies increase as the n_h's increase. The gain in efficiencies for $n_h = 10$ and stratum size of are 950% and 1058% which is much smaller than that for stratum size . We can therefore say the gain in efficiency heavily depends on the shape of strata rather than their sizes.

References

[1] Christman, M. C. (2000). A review of quadrat-based sampling of rare, geographically clustered populations. Journal of Agricultural, Biological, and Environmental Statistics 5:168-201.

[2] Francis, R.I.C.C. (1984). An adaptive strategy for stratified random trawl surveys, New Zealand Journal of Marine and Freshwater Research 18, 59-71. Francis, R.I.C.C. (1991). Statistical properties of two-phase surveys: comment, Canadian Journal of Fisheries and Aquatic Science, 48, 1128.

[3] Jolly, G.M. and Hampton, I. (1990). A stratified random transect design for acoustic surveysof fish stocks, Canadian Journal of Fisheries and Aquatic Scienc,e 47, 1282-1291.

[4] Jolly, G.M. and Hampton, I. (1991). Reply to comment by R.I.C.C. Francis, Canadian Journal of Fisheries and Aquatic Science 48, 1228-1129.

[5] Manly, B.F.J. (2004). Using the bootstrap with two-phase adaptive stratified samples from multiple populations at multiple location. Environmental and Ecological Statistics. 11: 367 - 383 .

[6] Manly, B.F.J., Akroyd, J.A. R. Walshe. (2002) Two-phase stratified random survey on multiple populations at multiple location. New Zealan Journal of Marine and Fershwater Research 36: 581-591.

[7] Salehi M. M. (1999). Rao-Blackwell versions of the Horvitz-Thompson and Hansen-Hurwitz estimators in adaptive cluster sampling. Journal of Environmental and Ecological Statistics. 6, 183-195.

[8] Salehi M, M. (2001a). Application of Adaptive Sampling in Fishery, Part 1: Adaptive cluster sampling and its strip versions. Iranian Journal of Fisheries Sciences 3(1),55-76.

[9] Salehi M, M. (2001b). Application of Adaptive Sampling in Fishery Part 2: Truncated Adaptive Cluster Sampling Designs. Iranian Journal of Fisheries Sciences, 3(1), 77-84.

[10] Salehi M. M. and Smith (2005). Two-stage sequential sampling design: a neighborhood-free adaptive sampling procedure. Journal of agricultural, biological and environmental statistics , Volume 10, Number 1, Pages 84-103.

[11] Seber, GAF and Salehi M, M. (2004). Adaptive Sampling, Encyclopedia of Biostatistics Volume 1. 2nd edition page 59-65. Peter Armitage and Theodore Colton (ed). John Wiley & Sons, Ltd, Chichester.

[12] Smith, D. R., R. F. Villella, and D. P. Lemarie. (2001). Survey protocol for assessment of endangered freshwater mussels in the Allegheny River, Pennsylvania. Journal of the North American Benthological Society 20:118-132.

[13] Smith, D. R., R. F. Villella, and D. P. Lemarie. (2003). Application of adaptive cluster sampling to low-density populations of freshwater mussels. Environmental and Ecological Statistics 10(1):7-15.

[14] Strayer, D.L., S. Claypool, and S. Sprague. (1997). Assessing unionid populations with quadrats and timed searches. pp. 163-169 in: K.S. Cummings, A.C. Buchanan, C.A. Mayer, and T.J. Naimo (editors). Conservation and management of freshwater mussels II. Initiatives for the future. Upper Mississippi River Conservation Committee, Rock Island, IL.

[15] Strayer, D. L. and D. R. Smith. (2003). A guide to sampling freshwater mussel populations. American Fisheries Society, Monograph 8, Bethesda, Maryland.

[16] Su, Z., and T. J. Quinn II. (2003). Estimator bias and efficiency for adaptive cluster sampling with order statistics and a stopping rule. Environmental and Ecological Statistics 10:17-41. Thompson, S. K. 1990. Adaptive cluster sampling. Journal of the American Statistical Association 85:1050-1059.

[17] Thompson, S. K. and G. A. F. Seber. (1994). Detectability in conventional and adaptive sampling. Biometrics 50:712-724. Thompson, S. K. and G. A. F. Seber. 1996. Adaptive sampling. Wiley, New York. Wolf, P. (1988a). Status of the spawning biomass of Pacific sardine, 1987-1988. Calif. Dep. Fish. Game, Mar. Res. Div., Rep. To the legislature, 9 pp.

[18] Wolf, P. (1988b). Status of the spawning biomass of Pacific sardine, 1988-1989. Calif. Dep. Fish. Game, Mar. Res. Div., Rep. To the legislature, 8 pp.

Table 1: The Relative efficiencies of Complete Allocation Stratified Sampling (CAST) in respect to Simple Random Sample (SRS) and Stratified Sample for Rockfish simulated population given by Su and Quinn II (2003).

n_h	$E[n]$	$Var[\hat{\tau}_{HT}]$	$Var[\hat{\tau}_{ST}]$	$Var[\hat{\tau}_{SRS}]$	$Eff_{st}[\hat{\tau}]$	$Eff_{srs}[\hat{\tau}]$
(H)						
1	76.0000	1612697594	1104446489	1180102338	0.6848441	0.7317567
2	128.1053	600409513	549853621	587519223	0.9157976	0.9785308
3	165.7193	291307078	366249206	391337696	1.2572616	1.3433855
4	194.1548	155090665	274666617	293481595	1.7710068	1.8923228
5	216.5686	85437522	219427640	234458685	2.5682819	2.7442122
6	234.9510	47203966	181990597	194457161	3.8554091	4.1195090
7	250.6228	25612464	154410548	164987848	6.0287268	6.4417015
8	264.4896	13422648	132732444	141824769	9.8886929	10.5660797
9	277.1793	6692079	114795243	122658848	17.1538984	18.3289601
10	289.1248	3128242	99348743	106154246	31.7586513	33.9341553
(G)						
1	77.0000	1319424855	1032306937	1161181345	0.7823916	0.8800663
2	122.5263	442366200	557300553	626874607	1.2598172	1.4170943
3	152.6491	197409368	398764226	448546420	2.0199863	2.2721638
4	175.3437	102621640	315301248	354663826	3.0724636	3.4560335
5	194.4182	60656250	260222432	292708905	4.2901174	4.8257007
6	211.6249	39981526	219055370	246402500	5.4789147	6.1629089
7	227.7469	28468612	186127954	209364386	6.5380059	7.3542183
8	243.1338	21168590	158774685	178596303	7.5004847	8.4368540
9	257.9511	16019237	135518233	152436488	8.4597181	9.5158392
10	272.2908	12134882	115421374	129830713	9.5115366	10.6989679
(F)						
1	80.0000	889869022	966360565	1107256514	1.085958	1.244292
2	124.0000	278139841	537732895	616134673	1.933318	2.215197
3	155.0286	117709245	381753386	437413258	3.243189	3.716048
4	180.5451	55877671	293656069	336471300	5.255338	6.021570
5	203.0769	27117863	234269228	268425822	8.638927	9.898487
6	223.7183	12793773	190364082	218119279	14.879432	17.048863
7	243.1063	5739103	155915226	178647759	27.167178	31.128169
8	261.6615	2441268	127726867	146349520	52.319894	59.948166
(E)						
1	91.0000	588874065	822143812	939951272	1.396128	1.596184
2	149.1111	161559072	407382273	465757304	2.521569	2.882892
3	192.9167	53419988	259899894	297141732	4.865218	5.562370
4	229.3810	17364858	180094673	205900980	10.371215	11.857337
5	261.6667	5271549	127999757	146341227	24.281244	27.760576
(D)						
1	92.0000	626776100	778305582	926725561	1.241760	1.478559
2	150.8889	169428595	383816017	457008304	2.265356	2.697350
3	193.7500	59364421	247479655	294673105	4.168821	4.963800
4	228.9524	24816093	173683725	206804565	6.998834	8.333486
5	260.3968	12076002	124636965	148404771	10.321046	12.289231
(C)						
1	99.0000	439252500	686476038	841626795	1.562828	1.916043
2	160.2857	116793127	337670547	413987618	2.891185	3.544623
3	208.5714	38880235	207226849	254062282	5.329876	6.534484
4	250.8571	12827083	134236329	164575143	10.465070	12.830286
(B)						
1	102.0000	186477781.2	632637580	808731474	3.392563	4.336878
2	157.4286	26664315.8	333652962	426524855	12.513089	15.99609
3	202.1429	3822859.0	211949417	270945278	55.442646	70.87503
4	243.2000	766444.3	139611561	178472267	182.154875	232.8574
(A)						
1	143.0000	40739896	345405722	497491126	8.478316	12.21140
2	234.6667	7669009	135407107	195028136	17.656402	25.43068
3	319.0000	178257	48800782	70288227	273.766425	394.30837

Notes Due to Lorenz Curve and Lorenz Ordering in View of Weighted Distributions

G. R. Mohtashami Borzadaran[1] ,and Zahra Behdani[2]

[1]*Department of Statistics, Ferdowsi University of Mashhad*
[2]*Department of Statistics, University of Birjand*

Abstract: The Lorenz curve is an important tool for analysis the size distribution of income and wealth. An income distribution F is preferred to an income distribution G in the sense of Lorenz order (generalized Lorenz order), if its Lorenz curve (generalized Lorenz curve), is nowhere below the Lorenz curve (generalized Lorenz curve), of G. Let Ł denote the set of all non-negative random variables with finite and positive means, for $X \in$ Ł with distribution F_X and quantile function $F_X^{-1}(u) = \{x : F_X(u) \leq u\}, u \in [0,1]$, the Lorenz curve is given by $L_X(p) = \frac{1}{E(X)} \int_0^p F_X^{-1}(u)du, p \in [0,1]$. The most widely used alternative to the Lorenz order is the generalized Lorenz order that we need define, generalized Lorenz curve ($GL_X(.)$), where $GL_X(p) = E(X)L_X(p)$. The concepts such as Lorenz partial order, generalized Lorenz partial order, dilation order and second-order absolute Lorenz order are reviewed and discussed as seen in (Ramos and Sordo 2003). In Bartoszewicz and Skolimowska (2006) used representation in view of weighted distributions by the Lorenz curve via the idea of Jain et al (1989) that the length biased distribution is closely related to the Lorenz curve. In view of the weighted distributions :
(i) It is shown that how to derive and determine Characterization results related to Lorenz ordering, generalized Lorenz ordering, dilation order, and second-order absolute Lorenz order for the cases that weights are increasing or decreasing functions is obtained. For special cases such as, probability weighted moments, order statistics, proportional hazard rate, reversed proportional hazard rate, generalized version of records, upper and lower records, truncated distributions, renewal equilibrium distributions, hazard rate and reversed hazard rate are some simple weights that the above ordering aspects obtained in this paper.
(ii Increasing proportional likelihood ratio (IPLR) or decreasing proportional likelihood ratio (DPLR) properties are sufficient conditions for the Lorenz ordering

Corresponding author, email: gmb1334@yahoo.com

of truncated distributions and their connection with a few cases to weighted distribution is obtained. Preservation of IFRA (DFRA) and NBW (NWU) classes under weighting is discussed in Bartoszewicz and Skolimowska (2006), some of the results applied for special cases of weights. Also, their results applied in view of the generalized Lorenz curves.

Keywords: Weighted Distribution, Lorenz Curve, Lorenz Ordering, Generalized Lorenz Ordering, IFRA, DFRA, NBU, NWU, Probability
Weighted Moments, Dispersion Ordering.

1 Introduction

100 years ago, in June 1905, a short article titled " Methods of measuring the concentration of wealth"published in JASA, proposing a simple method. It is called the Lorenz curve for visualizing distribution of income or wealth with respect to inequality or concentration. After this, huge of publications is obtained and reveal that the Lorenz curve is an important tool for analysis the size of the income and wealth. Among the applications of the Lorenz curve we will concentrate on Lorenz ordering and generalized Lorenz order and its connection with economics. Arnold (1987) proposed a class of Lorenz curve ordered with respect to indexing parameter and Sarabia et al (1999) proposed a family of Lorenz curve that can be ordered in a large number of cases. Kleiber (2003) surveyed selected applications of the Lorenz curve and related stochastic order in economics and econometrics with a bias towards problems in statistical distribution theory. These include characterizations of income distribution in terms of families of inequality measures, Lorenz ordering of multi parameter distributions in terms of their parameters, probability inequalities for distributions of quadratic forms. Based on the Lorenz ordering, if two distribution functions have associated Lorenz curves which do not intersect, then they can be ordered without ambiguity in terms of welfare functions which are symmetric, increasing and quasi-concave. Shorrocks (1983) and Kakwani (1984) introduced gen-

eralized Lorenz curves and generalized Lorenz ordering for the case that the Lorenz curves are intersected. Thistle (1989) showed that a distribution is uniquely determined by its generalized Lorenz curve. It is well-known from Atkinson (1970), Shorrocks (1983) and Kakwani (1984) that the generalized Lorenz ordering allow important judgments concerning economics welfare.
In this paper, we study the relationship between the weighted distributions and the parent distributions in the context of Lorenz curve, generalized Lorenz curve, Lorenz ordering, generalized Lorenz ordering.These relationship depends on the nature of the weighted function and give rise to interesting connections.

2 Weighted Distributions

On considering weighted function $w(x,\beta)$ which is a non-negative function with parameter β represent a family of distributions with pdf, $g_w(x,\beta,\theta) = \frac{w(x,\beta)f(x,\theta)}{E[w(X,\beta)]}$, which is called a weighted version of distribution. Let X be a random variable with pdf $f(x)$ and $w(x)$ be a non-negative function with $E(w(X)) \neq 0 < \infty$, Rao (1965) defined the weighted distribution of X with the weight function $w(.)$ as a distribution having $g_w(x) = \frac{w(x)f(x)}{E(w(X))}$. The weighted distribution with $w(x) = x^k$, k positive integer, is called the length-sized biased of order k distribution. If $E(X) < \infty$, we have for distribution $\hat{f}(x) = \frac{xf(x)}{E(X)}, x > 0$, as the length biased or sized biased distribution associated with F such that $\hat{F}(x) = \frac{\int_0^x tf(t)dt}{E(X)}$. The results due to many aspects of statistical inference can be generalized based on various weights. Known aging properties of life distributions can be obtained via weighting distributions that can be seen in Jian et al (1989). Order statistics, record value, residual lifetime of a stationary renewal process, selection samples, hazard rate, reversed hazard rate, proportional hazard model, reversed proportional hazard model, Lorenz curve and probability weighted moments are some special cases of weighted families. Bartoszewicz and Skolimowska (2006) studied re-

lation of weighted distribution with classes of life distribution and used a representation of weighted distributions by Lorenz curve to obtain some results concerning their relation with life distributions.

- The weight $w(x) = e^{lx}x^i[F(x)]^j[\overline{F}(x)]^k$ is one that implies many famous weights. If $(i = j = k = 0)$, $(l = j = k = 0)$, $(i = l = 0, k = n - j)$, $(i = l = k = 0)$ and $(i = j = l = 0)$, then $w(.)$ is moment generating function, moments, order statistics, reversed proportional hazard and proportional hazard respectively.
- The weight $w(x) = [-\ln F(x)]^j[-\ln \overline{F}(x)]^k$ is another weight that $(j = 0)$ and $(k = 0)$ implies upper record and lower record respectively.
- Let $A(x) = E[w(X)/X > x]$, then, $\overline{G}(x) = \frac{1}{E(w(X))}\overline{F}(x)A(x)$.

3 Lorenz Curve for Weighted Distributions

Let Ł denote the set of all non-negative random variables with finite and positive means. For $X \in$ Ł with distribution F_X and quantile function $F_X^{-1}(u) = \{x : F_X(u) \leq u\}, u \in [0, 1]$, the Lorenz curve is given by

$$L_X(p) = \frac{1}{E(X)} \int_0^p F_X^{-1}(u)du, p \in [0, 1].$$

It is increasing, convex, continuous on $p \in [0, 1]$ with $L(0) = 0$ and $L(1) = 1$. Also, $L_X(p) \leq p, \overline{L}_X(p) = 1 - L_X(p)$ and any function possessing these properties is the Lorenz curve of a certain statistical distribution (Thomp son 1976). It is also worth noting that the Lorenz curve itself may be considered as an cdf on unit interval. The Zenga curve is given by $Z_X(p) = 1 - \frac{F^{-1}(p)}{F_1^{-1}(p)}, p \in [0, 1]$, where $F_1^{-1}(p) = \inf\{x : F_1(x) \geq p\}$ and $F_1(x) = \frac{1}{E(X)} \int_0^x tf(t)dt$. The concentration curve takes values from point $(0, 1)$ to $(1, 0)$but does not have the behaviour of the Lorenz curve. The Lorenz curve and Zenga curve provided partial ordering. Here, we are interested to Lorenz ordering but in the cases that the Lorenz curves intersected, we need to use

generalized lorenz order. It is defined in terms of generalized Lorenz curve , $GL_X(u)$ where

$$GL_X(p) = E(X)L_X(p) = \int_0^p F_X^{-1}(u)du, p \in [0,1].$$

Generalized Lorenz curves are non-decreasing, continuous and convex with $GL_X(0) = 0$ and $GL_X(1) = E(X)$. A distribution is uniquely determined by its generalized Lorenz curve.

- $A_X(p) = \int_0^p [F^{-1}(t) - E(X)]dt, 0 \leq p \leq 1$ is called the absolute Lorenz curve, and is used in economics to compare income distributions. It is decreasing for $0 \leq p \leq F(E(X))$ and increasing for $F(E(X)) < p \leq 1$ and it takes values $A_X(0) = A_X(1) = 0$, and is convex function with respect to p.

It is interesting to note that the length-biased distribution (as a weighted distributions) is rather closely related to the Lorenz curve which is used in economics to illustrate income distributions. Connections due to Lorenz curves and with weighted distributions is discussed in this part. For special weights some remarks are noticeable. Let $U = w(X)$ and L_U be its Lorenz curve, on assuming $U \sim h(x)$, and $X \sim f(x)$, Bartoszewicz and Skolimowska (2006) proved that, let w be a monotone left continuous and increasing [decreasing] function, then $G_w(x) = L_U(F(x))$ $[G_w(x) = \overline{L}_U(\overline{F}(x))]$.

Theorem 1 *Let w be a monotone left continuous and increasing [decreasing] function, then $G_w(x) = \frac{1}{E(w(X))} GL_U(F(x))$ [$G_w(x)$ $= \frac{1}{E(w(X))} \overline{GL}_U(\overline{F}(x))$], where GL_U is the generalized Lorenz curve.*

Proof : We have, $GL_U(p) = \int_0^{H^{-1}(p)} th(t)dt$ and $H(x) = P(X \leq w^{-1}(x)) = F(w^{-1}(x))$. So, $GL_U(F(x)) = E(w(X)) \int_0^x w(t)f(t)dt$, and easy verify the results.

- For $w(x) = e^{lx}x^i[F(x)]^j[\overline{F}(x)]^k$, when

$$\frac{1}{r(x)}[l + \frac{i}{x}] + j\frac{r(x)}{\widetilde{r}(x)} > (<)k, \forall x > 0, \qquad (1)$$

then $G_w(x) = L_U(F(x))$ $(G_w(x) = \overline{L}_U(\overline{F}(x)))$, where $r(.)$ and $\widetilde{r}(.)$ are hazard rate and reversed hazard rate respectively. So for $k \leq 0$ and various positive values l, i and j, $G_w(x) = L_U(F(x))$ and for weighted such as proportional hazard rate and reversed hazard rate, $G_w(x) = \overline{L}_U(\overline{F}(x))$.

- For $w(x) = [-\ln F(x)]^j[-\ln \overline{F}(x)]^k$, if $\frac{n}{F(x)} \ln \overline{F}(x) > (<) \frac{m}{\overline{F}(x)} \ln F(x)$, $\forall x > 0$, then, $w(.)$ is increasing (decreasing) respectively that implies $G_w(x) = L_U(F(x))$ $(G_w(x) = \overline{L}_U(\overline{F}(x)))$. Note that $m = 0$ (lower record) and $n = 0$ (upper record) are weights that are decreasing and increasing respectively. We can have the results for generalized Lorenz curve and weighted distribution also via the same arguments.

- On noting that the hazard rate of the weighted distribution and reversed hazard rate of the weighted distribution with increasing (decreasing) weights are equal to $\frac{w(x)r_F(x)}{E(w(X))} \frac{\overline{F}(x)}{\overline{L}_U(F(x))}$ $(\frac{w(x)\widetilde{r}_F(x)}{E(w(X))} \frac{\overline{F}(x)}{L_U(\overline{F}(x))})$ and $\frac{w(x)r_F(x)}{E(w(X))} \frac{F(x)}{L_U(F(x))}$ $(\frac{w(x)\widetilde{r}_F(x)}{E(w(X))} \frac{F(x)}{\overline{L}_U(\overline{F}(x))})$ respectively. We can find the increasing (decreasing) cases for the second terms in the above statements based on Lorenz curve and also, generalized Lorenz curve.

4 Lorenz Ordering for Weighted Distributions

The star shaped order was introduced by Barlow and Proschan (1975) to compare continuous life distributions. The convex order which requires the continuity of the distributions and have linked with star shaped order.

Definition 1 For $X, Y \in$ Ł we say that X has a star shaped distribution with respect to Y and write $X \leq_* Y$, if the ratio of inverse distribution functions $F_Y^{-1}(v)/F_X^{-1}(v)$ is an increasing function in $v \in (0, 1)$. For the continuous variates $X, Y \in$ Ł we say that X is convex ordered with respect to Y(in symbols: $X \leq_c Y$), if the function $F_Y^{-1}(F_X(x))$ is convex on the support of X.

Definition 2 The Lorenz partial order, $\leq_L$ on Ł is defined by

$$X \leq_L Y \Longleftrightarrow L_X(u) \geq L_Y(u), \forall u \in [0,1].$$

Lorenz partial order is invariant with respect to scale transformation. X is stochastically smaller than Y ($X \leq_{st} Y$) if $F(x) \geq G(x)$ for $\forall x > 0$ where F and G are distribution function of X and Y respectively.

For reasons of mathematical tractability, both the convex and the star shaped ordering can sometimes be used to verify the Lorenz ordering which they imply. The relations between the three partial orderings are presented in the next theorem.

Theorem 2 *Suppose that $X, Y \in Ł$ are continuous. If $X \leq_c Y$, then $X \leq_* Y$ and also, $X \geq_Z Y$. Moreover, $X \leq_* Y$ implies $X \leq_L Y$.*

Proof : The proof of the first part is apparent from Arnold (1987, pp. 77-78) or Barlow and Proschan (1975, p. 107), While the second part is shown in Arnold (1987, p. 78) or Moothathu (1991).

Theorem 3 *Let $g : \Re^+ \to \Re^+$ be continuous function satisfying : $g(x) > 0$ for all $x > 0$, $g(x)$ is non-decreasing on $[0, \infty)$ and $\frac{g(x)}{x}$ is non-decreasing on $(0, \infty)$. If $g(X) \in Ł$, then, $g(X) \leq_L X$.*

- For random variables X_1 and X_2 where $X_i \sim GB2(a_i, b_i, p_i, q_i), i = 1, 2$ Wilfling (1990, 1996) found that $X_1 \geq_L X_2 \Rightarrow a_1 p_1 \leq a_2 p_2, a_1 q_1 \leq a_2 q_2$ where $(a_1 \leq a_2, p_1 \leq p_2, q_1 \leq q_2)$.

- A complete characterization of the Lorenz ordering with in the generalized beta family of distribution includes some of the results due to Lorenz ordering and generalized Lorenz ordering for a big class of distributions.

- The Zenga curve provide a partial ordering between random variables. The Zenga partial order, $\leq_Z$ on Ł is defined by $X \leq_Z Y \Longleftrightarrow Z_X(u) \geq Z_Y(u), \forall u \in [0,1]$.

Definition 3 We say the random variable Y is more dispersed than X in the dilation sense if $E(\phi(X - E(X)) \leq E(\phi(Y - E(y))$ for all convex ϕ that is denoted by $X \leq_{dil} Y$ (Ramos 2003).

Theorem 4 *Let X and Y be random variables with respective finite means $E(X)$ and $E(Y)$ and let the corresponding be F and G, respectively. Then, $X \leq_{dil} Y$ if and only if $A_X(p) \geq A_Y(p), \forall p \in [0, 1]$.*

- Extension of the Theorem 4, when random variable Y be a weighted distribution of X with weight w, under most of them, does not lead to a simple and nice form.
- Let X be a non-negative continuous random variable with density f, it will be increasing proportional likelihood ratio (IPLR) if $\frac{f(\lambda x)}{f(x)}$ is increasing in x for any positive constant $\lambda < 1$. It will be said decreasing proportional likelihood ratio (DPLR) if $\frac{f(\lambda x)}{f(x)}$ is decreasing in x for any positive constant $\lambda < 1$, that is defined in Ramos and Sordo Diaz (2001). IPLR and DPLR properties are sufficient conditions for the Lorenz ordering of truncated distributions and their connection with a few cases to weighted distribution is specially weights that concentrated on them and some special cases of them is interesting.
- The equilibrium distribution corresponding to F and G are F_e and G_e defined by $F_e(x) = \frac{1}{\mu_F} \int_0^x \overline{F}(t)dt$ and $G_e(x)\frac{1}{\mu_G} \int_0^x \overline{G}(t)dt$. F is said to be more harmonic new better than used in expectation (HNBUE) than G ($F \leq_{HNBUE} G$) if and only if $\overline{F}_e(x\mu_F) \leq \overline{G}_e(x\mu_G)$. G is said to be more dispersived than F ($F \leq_{disp} G$) if $G^{-1}F(x) - x$ is increasing in x. Note that, if $\mu_F \leq \mu_G$, then, $F \leq_{HNBUE} G$ implies $F_e \leq_{disp} G_e$ which is discussed in Kochar (1989).

Theorem 5 *$F \leq_{HNBUE} G$ if and only if $X \leq_L Y$.*

Theorem 6 *Let $X, Y \in L$, $X \leq_{st} Y$, $U = w(X) \leq_L V = w(Y)$ and w be monotone left continuous. If w is increasing (decreasing), then,*

$\hat{X}_w \leq_{st} (\geq_{st}) \hat{Y}_w$ where $\hat{X}_w$ and $\hat{Y}_w$ are weighted version of the random variables X and Y with the weight w respectively.

Proof: We have via the assumption, $F(x) \geq G(x)$ and $L_U(p) \geq L_V(p), p \in (0,1)$. Thus, via increasing (decreasing) w concludes that $\hat{F}_w(x) = L_U(F(x))$
$\geq L_V(F(x)) \geq L_V(G(x)) = \hat{G}_w(x)$ which is $\hat{X}_w \leq_{st} (\geq_{st}) \hat{Y}_w$.

5 Generalized Lorenz Order for Weighted Distributions

The Lorenz curve is only a partial order, so what does one do if two Lorenz curves intersect? The most widely used alternative to the Lorenz order is the generalized Lorenz order due to Shorrocks (1983). If F and G have equal means, the Lorenz curve of distribution F is nowhere below the Lorenz curve of distribution G, then F is preferred to G. If the generalized Lorenz curve of F is nowhere below the generalized Lorenz curve of G, then, F is preferred to G if means μ_F and μ_G are different. So we have the following definition :

Definition 4 The generalized Lorenz partial order, $\leq_{GL}$ on Ł is defined by

$$X \leq_{GL} Y \Longleftrightarrow E(X)L_X(u) \leq E(Y)L_Y(u), \forall u \in [0,1].$$

Let F and G be two distribution function of random variable X and Y respectively, then $F(x) \leq G(x), \forall x \in \Re^+ \Rightarrow F \geq_{FSD} G$ and $\int_0^x F(t)dt \leq \int_0^x G(t)dt, \forall x \in \Re^+ \Leftrightarrow F \geq_{SSD} G$ where FSD and SSD are "First-order stochastic dominance" and "second-order stochastic dominance" respectively.

- It can be shown that how to determine the relation between the the parameters of the two generalized beta, generalized gamma families (including many income distributions as special cases) and some characterization notes in view of generalized Lorenz ordering in place of Lorenz ordering.

Definition 5 Let X and Y be two random variables with absolute Lorenz curves $A_X(t)$ and $A_Y(t)$ respectively. We say that X is smaller than Y in the second-order absolute Lorenz order if $\int_p^1 A_X(t)dt \geq \int_p^1 A_Y(t)dt, \forall p \in [0,1]$.

- Note that the dilation order implies second-order absolute Lorenz order.
- $F \geq_{FSD} G$ if and only if $GL_F(p) - GL_G(p)$ is increasing in p. For example, let $f(x) \sim G(\alpha, \lambda), \alpha, \lambda > 0$, we know that $F \geq_L G$ if and only if $\alpha \geq \beta$. $\lambda \geq \nu$ and $\frac{\alpha}{\lambda} \geq \frac{\beta}{\nu}$ implies $F \geq_{SSD} G \Leftrightarrow F \geq_{GL} G$. $\lambda \leq \nu$ and $\alpha \geq \beta$ implies $F \geq_{FSD} G$.

Theorem 7 *Let $X, Y \in Ł$, $X \leq_{st} Y$, $U = w(X) \leq_{GL} V = w(Y)$, $E(U) \leq E(V)$ and w be monotone left continuous. If w is increasing (decreasing), then, $\hat{X}_w \leq_{st} (\geq_{st}) \hat{Y}_w$ where $\hat{X}_w$ and $\hat{Y}_w$ are weighted version of the random variables X and Y with the weight w respectively.*

Proof: We have via the assumption, $F(x) \geq G(x)$ and $E(U)L_U(p) \geq E(V)L_V(p)$, $p \in (0,1)$. Thus, via increasing w concludes that $\hat{F}_w(x) = E(U)GL_U(F(x)) \geq E(V)\ GL_V(F(x)) \geq E(V)GL_V(G(x)) = \hat{G}_w(x)$ which is $\hat{X}_w \leq_{st} \hat{Y}_w$. $w(.)$ decreasing leads to $\hat{X}_w \geq_{st} \hat{Y}_w$ via the same arguments.

6 IFRA (DFRA) and NBU (NWU) Classes Connected to Lorenz Curve and Weighted Distributions

Various representation is obtained in Bartoszewicz and Skolimowska (2006), that is connected to Lorenz curve and weighted distribution. We will mention them here with some comments. Let F be an absolutely continuous with $F(0) = 0$ and S_F be an interval. F is increasing failure rate in average (IFRA) iff $xr_F(x) \geq \int_0^x r_F(t)dt, x \in S_F$. F is decreasing failure rate in average (DFRA) iff $xr_F(x) \leq$

$\int_0^x r_F(t)dt, x \in S_F$. A distribution F is said to be NBU (NWU) if $\overline{F}(x+y) \leq (\geq)\overline{F}(x)\overline{F}(y)$ for all $x, y, x+y \in [0,\infty)$. It is well known that $IFRA \subset NBU$ and $DFRA \subset NWU$.
The following theorems obtained in Bartoszewicz and Skolimowska (2006) in view of Lorenz curve and we mentioned them via the similar arguments in view of Generalized Lorenz curve. We have the following theorems in view of generalized lorenz curve on using F is IFRA (DFRA) and

$$h(t) = \frac{w(t)\hat{F}(t)}{E(U)[1-\hat{F}_w(t)]}, \tag{2}$$

is increasing (decreasing), then $\hat{F}_w$ is IFRA (DFRA).

Theorem 8 *Let $F(0) = 0$ and w be decreasing left continuous for which existence of the expected value of w. If $[GL_U(p)]^\alpha \leq (\geq)[E(U)]^{\alpha-1}GL_U(p^\alpha)$, for every $\alpha \in (0,1)$ and $p \in [0,1]$, and F is IFRA, then $\hat{F}_w$ is IFRA (DFRA).*

The next theorems are concerning the preservation of NBU and NWU classes in connection with generalized Lorenz curve on noting that if h that is defined in (2) is increasing (decreasing) and F is NBU (NWU), then $\hat{F}_w$ is NBU (NWU).

Theorem 9 *Let F be absolutely continuous, if F is IFRA (NBU) and $w(x)\overline{F}(x)$ is increasing, then $\hat{F}_w$ is IFRA (NBU). Let w be decreasing left continuous. If F is DFRA (NWU) and $\frac{w(x)}{GL_U\hat{F}(x)}$ is decreasing, then $\hat{F}_w$ is DFRA (NWU).*

Theorem 10 *Let $F(0) = 0$ and F be NBU (NWU). Let w be increasing (decreasing) left continuous for which existence of the expected value of w. If $GL_U(pq) \leq (\geq)[E(U)]GL_U(p)GL_U(q)$, for every $p, q \in [0,1]$. Then $\hat{F}_w$ is NBU (NWU).*

7 Conclusions

In this paper, notes due to Lorenz and generalized Lorenz curve is discussed in view of weighted distributions and their order is derived in view of Bartoszewicz and Skolimowska (2006) based on weighted version. Properties of these two order found via two general form weights that some special cases of them are very famous and important at least in reliability. Characterization results is also obtained via the above arguments.

References

[1] Arnold, B.C. (1987). Majorization and the Lorenz Order. Lecture Notes in Statistics 43, Springer-Verlag, Berlin - New York.

[2] Atkinson, A. B. (1970). On the measurement of inequality. J. of Economics Theory, 2, 244-263.

[3] Barlow, R.E. and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing. Holt, Rinehart and Winston, New York.

[4] Bartoszewicz, J. and Skolimoska, M. (2006). Preservation of classes of life distributions under weighting. *Stat. and Prob. Letters*, **76,** 587-596.

[5] Jain, K., Singh, H. and Bagai, I. (1989). Relation for reliability measures of weighted distributions. *Comm. Statistics, Theory and Methods,* **18,** 4393-4412.

[6] Kakwani, N. (1984). Welfare ranking of income distributions. Advances in Econometrics, 3, 191213.

[7] Kleiber, C. and Kotz, S. (2003). *Statistical size Distribution in Economics and Actuarial Sciences.* John Wiley.

[8] Kochar, S. C. (1989). On extension of DMRL and related partial ordering of life distribution. *Comm. Statistics, Stochastics Models,* **5,** 235-245.

[9] Lorenz, M. O. (1905). Methods for measuring concentration of wealth. Journal of the American Statistical Association 9, 209219.

[10] Moothathu, T.S.K. (1991). On the sufficient condition for two non-intersecting Lorenz curves. The Indian Journal of statistics. Volume, 53. series, B. pt, 2. pp, 268-274.

[11] Ramos, H. M. and Sordo, M. A. (2001). The proportional likelihood ratio order and applications. *QUESTIIO,* **61,** 123-131.

[12] Ramos, H. M. and Sordo, M. A. (2003). Dispersion measures and dispersive orderings. *Stat. and Prob. Letters,* **25, 2,** 211-223.

[13] Rao, C. R. (1965) . On discrete distributions arising out of methods of ascertainment, in *Classical and Contagious Discrete Distributions,* G. Patil Ed., Pergamon Press and Statistical Publishing Society, Calcutta, 320-332.

[14] Sarabia, J. Castillo, E. and Slottje, D. (1999). An orderd family of Lorenz curves. Journal of Econometrics, 91. 43-60.

[15] Shorrocks, A. F. (1983). Ranking income distributions, Economica 50, 3-17.

[16] Thistle, P. D. (1989). Duality between generalized Lorenz curves and distribution functions. Economic Studies Quarterly, 40, 183-187.

[17] Thompson, W. A., Jr. (1976). Fishermans luck. Biometrics, 32, 265271.

[18] Wilfling, B.(1996). A sufficient condition for Lorenz ordering. The Indian Journal of statistics. Volume, 58. series, B. pt, 1. pp, 62-69.

[19] Wilfling, B. (1990). Concepts of Lorenz-ordering with a special regard to the Lorenz ordering of order statistics. Diploma Thesis, Fak. Statist., Univ. Dortmund (Germany).

Methods for Estimating Special Variations in Rotating Panel Surveys

Mohammad-Reza Namazi-Rad

School of Mathematics and Applied Statistics , University of Wollongong, NSW 2522, Australia

Abstract: In this paper, after a short introduction to rotating panel surveys, two different methods for estimating the mean differences between two time epochs are introduced in these especial form of sample surveys. The first estimation method uses a linear additive model with independent errors while the the dynamic version of the Rao-Graham estimator will be used in the second method, as a composite estimator. These methods of estimation will be compared to each other through a simulation study based on hypothetical data. The direct estimation procedure will be added in this study to be compared with the other estimation methods. The results show that the estimation under the linear additive model with independent errors is the best one in all cases and have the least mean square error. As correlation increases, the efficiency rises for all three methods of estimation in this empirical study.

Keywords: Rotating Panel survey, Linear Additive Model, Composite Estimation, Simulation.

1 Introduction

Sampling design is a key device for efficient estimation and other forms of inference about a large population when the resources available do not permit collecting the relevant information from every member of the population. In many statistical systems, some surveys are known as repeated surveys. Such surveys -which are repeated continuously over regular time periods,- can reveal trends and fluctuations over time. These surveys are often performed in one of the following types:

- **Survey of the independent Units:** This method chooses separate samples from the population under consideration in each time section.

- **Panel Survey:** In this method, a "fixed" sample is pursued over the time. In other terms, we refer to a fixed sample in different time sections.

- **Rotating Panel survey:** This survey method is a combination form of the other methods which has been mentioned above. In this method, the sample individuals neither change entirely over the time nor do they stay entirely the same; however a part of the sample is repeated according to a pattern known as the "rotation pattern".

Many types of rotation designs are actually used in many surveys, and estimating the variation is one the most important factors to evaluate the survey plan and the errors in the final results. In the rotating panel surveys, there are some overlapping units in each time period and some new ones, simultaneously. It is not easy to compute the variance for different trend estimations because of these overlappings. In this paper, two procedures for estimating this key factor are presented. In section 2, we present a linear additive model with independent errors which has been introduced by Iachan and Jones (1987). The Rao-Graham(1964) composite estimator has been reviewed in section 3. Finally, an empirical analysis has been proposed to compare the estimation methods, numerically.

2 Linear Additive Model with Independent Errors

In this section, a rotating panel survey design will be presented. This design will be used in the following estimation procedures. A convenient representation of this design is given by the incidence matrix N, which has elements n_{ij}; $i = 1, \ldots, t$ & $j = 1, \ldots, b$. In this matrix, b is the number of blocks and t is the numbers of sample periods. The element $n_{ij} = 1$, if group j is interviewed in period i, and is zero, otherwise. To define the mentioned matrix, it is important to consider some construction as below:

1. Setting of the first r entries of the first row equal to 1 and the rest to 0. Note that, "r" is the number of units which will be interviewed in each period of rotating panel survey.

$$\begin{cases} n_{1j} = 1; & i= 1, \ldots, r \\ n_{1j} = 0; & i= r+1, \ldots, b \end{cases}$$

2. Constructing the second row of mentioned matrix by shifting all entries in the first row, s places right in a cyclical manner. Note that, "s" is the shift parameter from one sample period to another.

3. Constructing the following rows in exactly the same way from their directly proceeding rows.

4. Eventually, row 1 will be occur in the *(t+1)*th row. This last row is discarded to leave an incidence matrix with t rows and b columns.

There is an example of this matrix in table (1) for a rotating panel survey with parameters *b=9, r=5, s=3, & t=3.*

Table (1): Incidence matrix for *b=9, r=5, s=3, & t=3.*

1	1	1	1	1	0	0	0	0
0	0	0	1	1	1	1	1	0
1	1	0	0	0	0	1	1	1

The class considered here is a subclass of cyclic designs (see John et al. (1972)). In some statistical articles, it is common to use a linear additive model to study the parameters in a rotation design. Iachan and Jones (1987) introduced a linear model for characteristic, Y. A simple analysis of variance which can be used to describe the response Y in the mentioned linear model is given by:

$$\underline{Y} = A \begin{pmatrix} \mu \\ \underline{\alpha} \\ \underline{\beta} \end{pmatrix} + \underline{\varepsilon} \qquad (1)$$

where the vector $\underline{Y}_{rt}$, consists the different values of Y_{ij}; $i= 1, \ldots, t$ & $j= 1, \ldots, b$. and the vector $\underline{\varepsilon}$, consists the different values of ε_{ij}. Matrix A and the vectors $\underline{\alpha}$ and $\underline{\beta}$ is given by:

$$\underline{\alpha} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_t \end{pmatrix} \qquad \underline{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_t \end{pmatrix} \qquad A = \begin{pmatrix} \underline{1}_r & \underline{1}_r & \underline{0} & \ldots & \underline{0} & D^{(1)} \\ \underline{1}_r & \underline{0} & \underline{1}_r & \ldots & \underline{0} & D^{(2)} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \underline{1}_r & \underline{0} & \underline{0} & \ldots & \underline{1}_r & D^{(t)} \end{pmatrix} \tag{2}$$

where,
Y_{ij} : the observed value of characteristic for group j in period i,
μ : the overall mean,
α_i : the effect of period i,
β_j : the effect of group j, and
ε_{ij} : the independent random residuals with mean zero and variance σ^2.

It will be noted that, $\underline{1}_r$ is a vector with r elements equal to 1 and $D^{(k)}$ is a $r \times b$ matrix with the elements of $d_{ij}^{(k)}$ which are defined as below.

$$\begin{cases} d_{jj}^{(k)} = 1 & if \quad j = (k-1)s + i \leq b \;\; ; \;\; i= 1, \ldots, r \\ d_{(j-b)(j-b)}^{(k)} = 1 & if \quad j = (k-1)s + i > b \;\; ; \;\; i= 1, \ldots, r \\ d_{ij}^{(k)} = 0 & for\ other\ values\ of\ i\ and\ j \end{cases} \tag{3}$$

Using least square method for estimation, we have:

$$\begin{pmatrix} \hat{\mu} \\ \hat{\underline{\alpha}} \\ \hat{\underline{\beta}} \end{pmatrix} = (A'A)^{-} A'\underline{Y} \tag{4}$$

As $(A'A)$ is not an invertible matrix, Moore-Penrose inverse is used in the formula and the symbol $(A'A)^{-}$ shows this Moore-Penrose inverse for $(A'A)$. Assuming $\hat{\alpha}_i - \hat{\alpha}_{i-d}$, as an estimator for $\alpha_i - \alpha_{i-d}$; $d = 1, \ldots, [\frac{t}{2}]$, the variance for this estimator is given by:

$$Var(\hat{\alpha}_i - \hat{\alpha}_{i-d}) = \nu_d \, \sigma^2 \tag{5}$$

where ν_d should be computed from the covariance matrix of the vector containing the estimated parameters.

$$Var \begin{pmatrix} \hat{\mu} \\ \underline{\hat{\alpha}} \\ \underline{\hat{\beta}} \end{pmatrix} = \sigma^2 (A'A)^- \tag{6}$$

It will be noted that the mentioned estimators in (4) are the best unbiased linear estimators. Iachan and Jones (1987), computed the value of ν_d for some kinds of rotation design and sorted their results in a referable table.

3 Composite Estimator

It is well known that, while looking for an optimal estimator of a parameter in repeated surveys with the same time spacing, using the observations not only from the present edition of the survey (occasion) but also from previous occasions may significantly improve the quality of estimation. A typical solution is provided by the composite estimator introduced by Rao and Graham (1964) which, by definition makes use only of the last composite estimator and the observations from the present occasion and the last most recent occasion from the past. More precisely Rao-Graham Estimator (RGE) on hth occasion, $\hat{\mu}_h$, is given by:

$$\hat{\mu}_h = Q(\hat{\mu}_{h-1} + \bar{Y}^{(h)}_{h-1,h} - \bar{Y}^{(h-1)}_{h-1,h}) + (1-Q)\bar{Y}_h \tag{7}$$

where $\hat{\mu}_{h-1}$ is the RGE on $(h-1)$th occasion, $\bar{Y}^{(h)}_{h-1,h}$ is the sample mean for the units common to both $(h-1)$th and hth occasions calculated for the hth occasion, $\bar{Y}^{(h-1)}_{h-1,h}$ is the sample mean for the units common to both $(h-1)$th and hth occasions calculated for the $(h-1)$th occasion, and $Q \,\epsilon\, [0,1)$ is the sample mean for all the units on hth occasion and is a numerical parameter which does not depend on h.

Finding the best optimal value for Q, needs some complicated numerical computations. Rao and Graham presented a table for these

optimal values. Although we can find a good value for Q in this table, it is not simple to estimate the variance for changes between two periods. Additionally, the composite estimator, as proposed by Rao and Graham, suffers from certain disadvantages. It is designed for a stable situation in the sense that its basic parameter is kept constant on all occasions, and it is restricted only to a certain family of rotation designs. Ciepiela and *et al.* (2005) proposed a dynamic version of the Rao-Graham estimator without any restrictions on the rotation pattern. Mathematically, the algorithm which they have developed, is much simpler than the classical composite estimator. The dynamic version of RGE has the form:

$$\hat{\mu}_h^* = Q_h(\hat{\mu}_{h-1}^* + \bar{Y}_{h-1,h}^{(h)} - \bar{Y}_{h-1,h}^{(h-1)}) + (1 - Q_h)\bar{Y}_h \quad ; \quad h = 2, 3, \ldots \tag{8}$$

while $\hat{\mu}_1^* = \bar{Y}_1$, where all the symbols are introduced in (7) except of Q_h which plays the role of the former Q. Let us point out that they do not impose any priori restrictions on the range of (Q_h) (observe that in Rao and Graham (1964) the restriction on the range of Q made it possible to pass to the limit with $h \longrightarrow \infty$ in the expression for the variance of $\hat{\mu}_h$).

Ciepiela *et al.* (2005) offered a new way to choose in a dynamic way, i.e. on each occasion $h \geq 2$, the value Q_h which minimizes the variance of $\hat{\mu}_h^*$. In this way, Y_{ij} is assumed to be the observed value for the jth unit in the ith occasion of sampling in the following formulas:

$$E(Y_{ij}) = \mu_i \quad ; \quad i = 1, 2, \ldots \ \ \& \ \ j = 1, 2, \ldots \tag{9}$$

$$\begin{cases} Corr(Y_{ij}, Y_{lk}) = 0 & if \quad j \neq k \ \ ; \quad i, j, k, l = 1, 2, \ldots \\ Corr(Y_{ij}, Y_{(i+k)j}) = \rho^k & if \quad k = 0, 1, \ldots \ \ ; \ \ j = 1, 2, \ldots \end{cases} \tag{10}$$

$$Var(Y_{ij}) = \sigma^2 \quad ; \quad i = 1, 2, \ldots, j = 1, 2, \ldots \tag{11}$$

The rotation scheme is described by the rotation matrix $R = (r_{ij})$, where $r_{ij} = 1$, if the jth unit is in the sample on the ith occasion, otherwise $r_{ij} = 0$. There is absolutely no restriction on the rotation pattern. By n_i we denote the sample size on the ith occasion, and m_i

denotes the size of overlap between samples on the $(i-1)$th and ith occasions.

Denote also for $k = 2, 3, \ldots$

$$\begin{cases} Q_i \times Q_{i-1} \times \ldots \times Q_k & for \quad i = 1, 2, \ldots, k \\ 1 & for \quad i = k+1 \end{cases} \tag{12}$$

Then, they defined the weights which would be responsible for the form of quadratic functions (F_k) to be minimized.

$$W_{ij}^{1} = \tfrac{1}{n_1} \quad for \quad i = 1, 2, \ldots \;\; \& \;\; j = 1, 2, \ldots$$

and for any $i = 1, \ldots, k > 1$, and any $j = 1, 2, \ldots$

$$W_{ij}^{(k)} = r_{ij}[D_{ik}(\frac{r_{(i-1)j}}{m_i} - \frac{1}{n_i}) + D_{(i+1)k}(\frac{1}{n_i} - \frac{r_{(i+1)j}}{m_{i+1}})] \quad for \;\; i = 1, \ldots, k \;\; \& \;\; j = 1, 2, \ldots \tag{13}$$

where in the last expression they adopted the rule that $r_{(k+1)j} = r_{0j} = 0$.

According to the main result of Ciepiela *et al.* (2005) researches, in the model described above, the optimal value of Q_h which minimizes the variance of $\hat{\mu}_h^*$, $h \geq 1$, is given by:

$$Q_h = \frac{C_h - B_h}{A_h - 2B_h + C_h} \;\; ; \;\; h \geq 1 \tag{14}$$

with $C_1 = B_1$ & $A_h - 2B_h + C_h > 0$, where for $h \geq 2$:

$$A_h = \sigma^2 \sum_j \Big[\sum_{i=1}^{h-1} (W_{ij}^{(h-1)})^2 r_{ij} + 2 \sum_{1 \leq i_1 \leq i_2 \leq h-1} W_{i_1 j}^{(h-1)} W_{i_2 j}^{(h-1)} r_{i_1 j} r_{i_2 j} \rho^{i_2 - i_1} \Big] +$$

$$2\frac{(1-\rho)\sigma^2}{m_h} \Big[1 - \sum_j \sum_{i=1}^{h-1} W_{ij}^{(h-1)} r_{(h-1)j} r_{hj} r_{ij} \rho^{h-1-i} \Big]$$

$$B_h = \frac{\sigma^2}{n_h} \Big[1 - \rho + \sum_j \sum_{i=1}^{h-1} W_{ij}^{(h-1)} r_{ij} r_{hj} \rho^{h-i} \Big]$$

$$C_h = \frac{\sigma^2}{n_h} \tag{15}$$

then,

$$\hat{\mu}_h^* = \sum_{i=1}^{h} \sum_j W_{ij}^{(h)} r_{ij} Y_{ij} \tag{16}$$

Now we can use an unbiased estimator $\hat{\mu}_{(h+k)}^* - \hat{\mu}_h^*$ for estimating the changes in the mean value between the hth and $(h+k)$th occasions $(\mu_{(h+k)} - \mu_h)$. Ciepiela and *et al.* (2005) also estimated the variance for this estimator. Note that, the variance for the composite estimator in the hth occasion is given by:

$$Var(\hat{\mu}_h^*) = Q_h^2 A_h + 2Q_2(1 - Q_h) B_h + (1 - Q_h)^2 C_h \tag{17}$$

If the real values for variance of the hth occasion (σ^2) and the correlation between two different occasions (ρ) will be unknown in dynamic RGE, we should estimate them as below:

$$\hat{\sigma}^2 = \frac{1}{t} \sum_{h=1}^{t} S_h^2 \qquad \& \qquad \hat{\rho} = \frac{1}{t} \sum_{h=1}^{t} \hat{\rho}_h \tag{18}$$

where S_h^2 is the sample variance for the hth occasion and $\hat{\rho}_h$ is the correlation between the joint sample units for the hth and $(h+1)$th occasions.

4 Simulation

An experimental appraisal of the two methods of estimation which are presented above will be illustrated in this chapter. The direct estimation procedure will be added to this study in order to be compared with the other mentioned estimation methods. To do this, we suppose that the purpose of the survey is to estimate the variations of the characteristic "X" in a society whose population size is 200,000. It will be noted that the mentioned variations are studied through two sequential time sections in this empirical study.

In the population, the ith individual has two values X_{i1} and X_{i2}; the former shows the value of variable X for the first time section and the latter shows it for the second time section. To justify repeated survey, the distribution of the variable X is considered to be different in two sequential time sections and the data from these sections are assumed to be correlated. Now, suppose that the aim of a survey is to estimate $\bar{X}_1 - \bar{X}_2$ whose unbiased estimator is $\hat{\bar{X}}_1 - \hat{\bar{X}}_2$. Note that, $\hat{\bar{X}}_1$ and $\hat{\bar{X}}_2$ denote the sample means.

To also examine the effect of different overlap rates in the assessment, sampling in this case is performed in 5 different conditions (overlaps of 25%, 20%, 30%, 40%, & 50%). For example, in the rotating panel survey with an overlap of 10%, only 10% of the values of the samples in the second time section are chosen from the samples of the first section. Since the efficiency of the three estimation methods is expected to be influenced by the correlation between the time sections, our simulation is repeated with four different correlations, 0.25, 0.4, 0.6 and 0.8, in order to examine the effect of different correlation coefficients on the efficiency. The statistical population behaves as it follows in these four conditions:

$$\begin{cases} \textbf{1:} & X_1 \sim N(1000, 100) & X_2 \sim N(1950, 570) & r_{X_1,X_2} = 0.2 \\ \textbf{2:} & X_1 \sim N(1000, 100) & X_2 \sim N(1650, 295) & r_{X_1,X_2} = 0.4 \\ \textbf{3:} & X_1 \sim N(1000, 100) & X_2 \sim N(2000, 250) & r_{X_1,X_2} = 0.6 \\ \textbf{4:} & X_1 \sim N(1000, 100) & X_2 \sim N(1600, 190) & r_{X_1,X_2} = 0.8 \end{cases}$$

The mean squared error is used in different survey methods to quantify the amount by which the estimators differ from the true value of the quantity being estimated. The purpose is to estimate the variations in the average means of the two sections.

$$MSE = \frac{\sum_{i=1}^{1000} (\hat{\theta}_i - \theta)^2}{1000} \tag{19}$$

where:

$$\hat{\theta}_i = \hat{\bar{X}}_{2i} - \hat{\bar{X}}_{1i}$$

$$\theta = \bar{X}_2 - \bar{X}_1 \begin{cases} 950 & forcase \ \textbf{(1)} \\ 650 & forcase \ \textbf{(2)} \\ 1000 & forcase \ \textbf{(3)} \\ 600 & forcase \ \textbf{(4)} \end{cases}$$

In order to perform the experimental assessment, 1000 samples is selected in each case of sampling. Table (2), (3) & (4) present the mean squared error values for all cases based on the equations above.

Table (2): The value of MSE for all cases of simulation in direct estimation

	Correlation			
	0.2	0.4	0.6	0.8
10% overlapping	566	158	118	71
20% overlapping	555	153	115	67
30% overlapping	517	147	111	60
40% overlapping	516	146	103	55
50% overlapping	512	145	98	49

Table (3): The value of MSE for all cases of simulation in the composite estimation

	Correlation			
	0.2	0.4	0.6	0.8
10% overlapping	490	134	99	58
20% overlapping	472	128	95	55
30% overlapping	436	119	89	48
40% overlapping	425	117	82	43
50% overlapping	418	115	77	38

Table (4): The value of MSE for all cases of simulation in the estimation based on the linear additive model

	Correlation			
	0.2	0.4	0.6	0.8
10% overlapping	483	128	91	51
20% overlapping	473	122	88	47
30% overlapping	456	108	80	42
40% overlapping	445	98	71	37
50% overlapping	439	91	60	28

As it is presented in the tables above, the estimation based on the linear additive model is the most efficient method in all cases and have the least mean square error. The composite estimators are at the next step, and finally the direct estimation method has the largest mean square errors, in all cases. Additionally, as the correlation increases, the efficiency rises for all three methods of estimation.

References

[1] Ciepiela, P., Gniado, K., Weso lowski, J., and Wojtys M. (2005). Dynamic Rao-Graham estimator for an arbitrary rotation scheme, The III Conference on Survey Methods is Social and Economic Sciences, Katowice.

[2] John, J. A., Wolock, F. M., and David, H. A. (1972). *Cyclic Designs.* U.S. National Bureau of Standards, Applied Mathematics Series, 62.

[3] Iachan, R. and Jones, B. (1987). Classification and Properties of Rotation Sampling Designs. *Journal of Official Statistics.* **3**, 169-181.

[4] Rao, J. N. K., and Graham, J. E. (1964). Rotation designs for sampling on repeated occasions. *J. Am. Stat. Assoc.* **59**, 492-509.

Runge–Kutta Methods For Numerical Solution Of Stochastic Differential Equations

M. Namjoo[1], A. R. Soheili[2]

[1]*Department of Mathematics, Vali–Asr University of Rafsanjan.*
[2]*Department of Mathematics, University of Sistan and Baluchestan.*

Abstract: In This paper we construct stochastic Runge–Kutta (SRK) methods with strong order 1 for strong solutions of Stratonovich stochastic differential equations (SDEs). Three SRK methods are constructed in this paper. They are an explicit two–stage method, an explicit three–stage method with minimum principal error coefficients and an implicit three–stage method with minimum principal error coefficients. Numerical results with two test problems of our methods and Burrage method and Platen method will be compared.

Keywords: Stochastic differential equations; Strong approximation; Runge–Kutta methods.

1 Introduction

Consider the scalar autonomous Itô SDE given by

$$dy(t) = g_0(y(t))\ dt + g_1(y(t))\ dW(t)\ , \quad y(t_0) = y_0, \quad t \in [t_0, T], \ (1)$$

where g_0 is called the drift coefficient, g_1 is called the diffusion coefficient, and $W(t)$ is a standard Wiener process. In [5] we have analysed an estimate of the absolute error for SDE (1) by path–wise approximations. Moreover order conditions for coefficients of a class of SRK methods with strong order 1 for the SDE problem (1) is obtained (see [6]), especially explicit two–stage and three–stage SRK methods of this class with minimum principal error coefficients are constructed. For the SDE problem (1) the associated Stratonovich SDE given by

$$dy(t) = \bar{g}_0(y(t))\ dt + g_1(y(t))\ odW(t)\ , \quad y(t_0) = y_0, \qquad (2)$$

where

$$\bar{g}_0(y) = g_0(y) - \frac{1}{2}\ g_1'(y)\ g_1(y).$$

In other words two different SDE (1) and (2), under different rules of calculus, have the same solution. There are many different methods to solve these kinds of differential equations (see, for example, [1]). In this paper we will present three new classes of methods for solving (2). The organization of this paper is as follows: In the next section the rooted tree analysis of strong schemes for SDEs with a scalar Wiener process is introduced. In section three the new classes of explicit and implicit SRK methods for SDEs is constructed. Also we show that 1–norm of principal error coefficients our three–stage SRK methods are less than the 1–norm of principal error coefficients Platen and Burrage methods. In continuation the fixed–point iteration algorithm is used to improve of our implicit method. Finally some numerical results which show the efficiency of our methods will be presented in the last section.

2 Stochastic Runge–Kutta methods

Consider the scalar autonomous Stratonovich SDE given by

$$dy(t) = g_0(y(t))\, dt + g_1(y(t))\, o dW(t)\,, \quad y(t_0) = y_0, \quad t \in [t_0, T]. \tag{3}$$

General form of s–stage SRK methods for solving SDE (3) given by

$$\begin{cases} Y_i & = y_n + \sum_{j=1}^{s} Z_{ij}^{(0)}\ g_0(Y_j) + \sum_{j=1}^{s} Z_{ij}^{(1)}\ g_1(Y_j), \quad i = 1, 2, \ldots, s \\ y_{n+1} & = y_n + \sum_{j=1}^{s} z_j^{(0)}\ g_0(Y_j) + \sum_{j=1}^{s} z_j^{(1)}\ g_1(Y_j), \end{cases} \tag{4}$$

which can be represented in tableau form :

$$\begin{array}{c|c|c} & Z^{(0)} & Z^{(1)} \\ \hline & {z^{(0)}}^T & {z^{(1)}}^T \end{array},$$

where $Z^{(k)} = (Z_{ij}^{(k)})_{s\times s}$ and ${z^{(k)}}^T = (z_1^{(k)}, \ldots, z_s^{(k)})$ for $k = 0, 1$. This class of methods was introduced by Burrage and Burrage (see [1]).

Let τ_k $(k = 0, 1)$ be the tree with one vertex with colour k. Then other trees can be built up recursively by defining a new tree t which is formed by joining trees $t_1, \ldots, t_m$ to a new root τ_k as $t = [t_1, \ldots, t_m]_k$. Let T denote the set of all rooted trees with colour k $(k = 0, 1)$ then the Stratonovich Taylor expansion for the exact solution of the SDE given by (3) is (see [2])

$$y(t) = \sum_{t \in T} \frac{\gamma(t)}{\rho(t)!} J(t)\alpha(t)F(t)(y_0), \tag{5}$$

where $J(t)$ and $F(t)$ are used to represents the J–integral and the elementary differential associated with tree t and $\rho(t)$ and $\sigma(t)$ are used to represent the number of vertices and the symmetry of t. Here $\alpha(t)$ represents the number of possible different monotonic labellings associated with tree t. Let $a(t) = z^{(k)}\Phi(t)$ where Φ defined recursively by

$$\Phi(t) = \begin{cases} e, & t = \tau_k, \\ (Z^{(k)}\Phi(t_1)) * (Z^{(k)}\Phi(t_2)) * \ldots * (Z^{(k)}\Phi(t_m)), & t = [t_1, \ldots, t_m]_k. \end{cases}$$

with $*$ denoting the component–by component product (see [2]), then Stratonovich Taylor series expansion the numerical method by (4), given by (see [2])

$$Y(t) = \sum_{t \in T} \frac{\gamma(t)}{\rho(t)!} a(t)\alpha(t)F(t)(y_0). \tag{6}$$

From (5) and (6) local truncation error over one step with an exact initial value can be written

$$L_1 \equiv y(t) - Y(t) = \sum_{t \in T^\star} e(t)\ F(t)(y_0),$$

where for tree t term $e(t) = \dfrac{\gamma(t)}{\rho(t)!}(J(t) - a(t))\alpha(t)$ is the coefficient of local truncation error.

3 SRK methods with strong order 1.0

For solving the Stratonovich SDE (3), a class of SRK methods given by (4) can also be characterized by

$$Z^{(0)} = hA, \quad {z^{(0)}}^T = h\alpha^T, \quad Z^{(1)} = J_1 B, \quad {z^{(1)}}^T = J_1\gamma^T \tag{7}$$

where $A = (a_{ij})$ and $B = (b_{ij})$ are $s \times s$ real matrices and $\alpha^T = (\alpha_1, \ldots, \alpha_s)$ and $\gamma^T = (\gamma_1, \gamma_2, \ldots, \gamma_s)$ are row s–dimensional vectors. If the matrices A and B are strictly lower triangular, then the method (7) is said to be explicit, otherwise it is implicit. The convergence Theorem in [2] shows that the SRK method of the form (7) will converge to the exact solution of SDE (3) with strong order 1.0 if the local truncation error satisfies

$$(E(y(t) - Y(t))^2)^{\frac{1}{2}} = O(h^{1.5}), \quad E(y(t) - Y(t)) = O(h^2). \tag{8}$$

Hence a SRK method of the form (7) will satisfy the mean–square condition in (8) if

$$E(h - h\alpha^T e)^2 = 0, \quad E(J_1 - J_1\gamma^T e)^2 = 0, \quad E(J_{11} - {J_1}^2\gamma^T Be)^2 = 0 \tag{9}$$

Note that conditions (9) arise from trees τ_0, τ_1 and $[\tau_1]_1$ (see [1]). These conditions are equivalent to

$$\alpha^T e = 1, \quad \gamma^T e = 1, \quad \gamma^T Be = \frac{1}{2} \tag{10}$$

On the other hand, the method will satisfy mean condition in (8) if:

$$\begin{cases} E(J_{10} - J_1 h\alpha^T Be) = 0 \\ E(J_{01} - J_1 h\gamma^T Ae) = 0 \\ E(J_{111} - {J_1}^3\gamma^T B(Be)) = 0 \\ E(J_{111} - \frac{1}{2}{J_1}^3\gamma^T (Be)^2) = 0. \end{cases} \tag{11}$$

Note that conditions (11) arise from trees $[\tau_1]_0$, $[\tau_0]_1$, $[[\tau_1]_1]_1$ and $[\tau_1, \tau_1]_1$ (see [1] for further details). It can be shown that the mean conditions (11) are all satisfied and hence a SRK method of the form given in (7) will be of strong order 1.0 if the order conditions (10) are satisfied. In order to construct a class of explicit SRK methods of the form (7) with $s = 2$, we try to find a method of coefficients:

$$\begin{array}{c|cc|cc} 0 & 0 & 0 & 0 \\ ha_{21} & 0 & J_1 b_{21} & 0 \\ \hline h\alpha_1 & h\alpha_2 & J_1\gamma_1 & J_1\gamma_2 \end{array} \tag{12}$$

which complies with the conditions in (10). We have six unknowns and there are three equations to be satisfied. We choose the deterministic part coefficients of (12) the modified Euler method (see [3]). This ensure that our method works well in the case of small stochastic influence. From (10) it is seen that we can assume $B = A$ and $\gamma = \alpha$, consequently we have the following method that is called 'EM1' and is presented by the tableau:

$$\begin{array}{c|cc|cc} 0 & 0 & 0 & 0 \\ \frac{1}{2}h & 0 & \frac{1}{2}J_1 & 0 \\ \hline 0 & h & 0 & J_1 \end{array}$$

As there are six parameters to be determined, additional conditions can be considered, for example, the conditions for minimum principal error coefficients, namely the terms corresponding to $h^{1.5}$ have minimum coefficients. The mean square of the principal error coefficients are give by

$$\begin{cases} (\frac{1}{3} - \alpha^T Be + (\alpha^T Be)^2)h^3, \\ (\frac{1}{3} - \gamma^T Ae + (\gamma^T Ae)^2)h^3, \\ (\frac{1}{36} - \frac{1}{3}\gamma^T B(Be) + (\gamma^T B(Be))^2)15h^3, \\ (\frac{1}{9} - \frac{2}{3}\gamma^T (Be)^2 + (\gamma^T (Be)^2)^2)\frac{15}{4}h^3, \end{cases}$$

which can be derived from (11). These principal error coefficients are minimized if (see [1])

$$\alpha^T Be = \frac{1}{2}, \quad ,\gamma^T Ae = \frac{1}{2}, \quad \gamma^T B(Be) = \frac{1}{6}, \quad \gamma^T (Be)^2 = \frac{1}{3}, \quad (13)$$

and so principal error coefficients respectively are given by

$$\frac{1}{12}h^3,\ \frac{1}{12}h^3,\ 0,\ 0.$$

Note that the 'EM1' method has principal error coefficients

$$\frac{1}{12}h^3,\ \frac{1}{12}h^3,\ \frac{5}{12}h^3,\ \frac{5}{192}h^3$$

and the Burrage method (see [1]) is presented by the tableau:

$$\begin{array}{c|cc|cc} 0 & 0 & 0 & 0 & 0 \\ \frac{2}{3}h & 0 & \frac{2}{3}J_1 & 0 \\ \hline & \frac{1}{4}h & \frac{3}{4}h & \frac{1}{4}J_1 & \frac{3}{4}J_1 \end{array}$$

which has minimum principal error coefficients

$$\frac{1}{12}h^3,\ \frac{1}{12}h^3,\ \frac{5}{12}h^3, 0.$$

Also the Platen method (see [1]) is

$$\begin{array}{c|cc|cc} 0 & 0 & 0 & 0 \\ h & 0 & J_1 & 0 \\ \hline & h & 0 & \frac{1}{2}J_1 & \frac{1}{2}J_1 \end{array}$$

and has principal error coefficients

$$\frac{1}{3}h^3,\ \frac{1}{3}h^3,\ \frac{5}{12}h^3, \frac{1}{36}h^3.$$

With the restriction of $s = 2$, it was seen that $\gamma^T B(Be) = 0$, but if $s = 3$ then $\gamma^T B(Be) = 0$ is not zero and in order to have the minimum principal error it must be take its minimum value which is $\frac{1}{6}$. In order to construct a class of explicit SRK methods of the form (7) with $s = 3$, we consider the matrices A and B and the row vectors α^T and γ^T with the following forms:

$$A = \begin{pmatrix} 0 & 0 & 0 \\ a_{21} & 0 & 0 \\ a_{31} & a_{32} & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 & 0 \\ b_{21} & 0 & 0 \\ b_{31} & b_{32} & 0 \end{pmatrix},$$
$$\alpha^T = \begin{pmatrix} \alpha_1 , & \alpha_2 , & \alpha_3 \end{pmatrix}, \gamma^T = \begin{pmatrix} \gamma_1 , & \gamma_2 , & \gamma_3 \end{pmatrix}.$$

Hence by equations (10) and (13), we have the following system of seven equations with twelve unknowns:

$$\begin{cases} \alpha_1 + \alpha_2 + \alpha_3 = 1 \\ \gamma_1 + \gamma_2 + \gamma_3 = 1 \\ \gamma_2 b_{21} + \gamma_3 (b_{31} + b_{32}) = \frac{1}{2} \\ \alpha_2 b_{21} + \alpha_3 (b_{31} + b_{32}) = \frac{1}{2} \\ \gamma_2 a_{21} + \gamma_3 (a_{31} + a_{32}) = \frac{1}{2} \\ \gamma_3 b_{32} b_{21} = \frac{1}{6} \\ \gamma_2 {b_{21}}^2 + \gamma_3 (b_{31} + b_{32})^2 = \frac{1}{3}. \end{cases} \tag{14}$$

In order to reduce the free parameters, we choose the deterministic components of SRK method in (4) the three–stage explicit Runge–Kutta method of order 3 (see [3]). From equations (14) it is seen that we can assume $B = A$ and $\gamma = \alpha$, consequently we have the following three–stage explicit SRK method with minimum principal error coefficients, that is presented by the tableau:

$$\begin{array}{c|ccc|ccc} 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{2}{3}h & 0 & 0 & \frac{2}{3}J_1 & 0 & 0 \\ -h & h & 0 & -J_1 & J_1 & 0 \\ \hline & 0 & \frac{3}{4}h & \frac{1}{4}h & 0 & \frac{3}{4}J_1 & \frac{1}{4}J_1 \end{array}$$

where is called 'EM2', and has principal error coefficients:

$$\frac{1}{12}h^3,\ \frac{1}{12}h^3,\ 0,\ 0.$$

Now we will generalize the above explicit SRK method to implicit method for SDEs. In the implicit case with $s = 3$, we will use the matrices:

$$A = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix},\quad B = \begin{pmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{pmatrix},$$

which using conditions (10) and (13) we will have the following system of seven equations with eighteen unknowns:

$$\begin{cases} \alpha_1 + \alpha_2 + \alpha_3 = 1 \\ \gamma_1 + \gamma_2 + \gamma_3 = 1 \\ \gamma_1 b_{11} + \gamma_2(b_{21} + b_{22}) + \gamma_3(b_{31} + b_{32} + b_{33}) = \dfrac{1}{2} \\ \alpha_1 b_{11} + \alpha_2(b_{21} + b_{22}) + \alpha_3(b_{31} + b_{32} + b_{33}) = \dfrac{1}{2} \\ \gamma_1 a_{11} + \gamma_2(a_{21} + a_{22}) + \gamma_3(a_{31} + a_{32} + a_{33}) = \dfrac{1}{2} \\ \gamma_1 {b_{11}}^2 + \gamma_2(b_{21}b_{11} + b_{22}(b_{21} + b_{22})) + \gamma_3(b_{31}b_{11} + b_{32}(b_{21} + b_{22}) \\ \qquad\qquad + b_{33}(b_{31} + b_{32} + b_{33})) = \frac{1}{6} \\ \gamma_1 b_{11}^2 + \gamma_2(b_{21} + b_{22})^2 + \gamma_3(b_{31} + b_{32} + b_{33})^2 = \dfrac{1}{3}. \end{cases}$$

Again we choose the deterministic part coefficients of (4) the Lobatto III method that is a three–stage implicit Runge–Kutta method of order

4 (see [3]). Consequently with assuming $B = A$ and $\gamma = \alpha$ a family of three–stage implicit SRK methods with minimum principal error coefficients can be presented by the tableau:

$$\begin{array}{c|ccc|ccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4}h & \frac{1}{4}h & 0 & \frac{1}{4}J_1 & \frac{1}{4}J_1 & 0 \\ 0 & h & 0 & 0 & J_1 & 0 \\ \hline & \frac{1}{6}h & \frac{2}{3}h & \frac{1}{6}h & \frac{1}{6}J_1 & \frac{2}{3}J_1 & \frac{1}{6}J_1 \end{array}$$

where is referred to 'IM', and has principal error coefficients:

$$\frac{1}{12}h^3,\ \frac{1}{12}h^3,\ 0,\ 0.$$

If we use the 1–norm to estimate the contribution of all error terms to the principal error term then Table 1 presents these values for methods '$Platen$' , '$EM1$' , '$Burrage$' , '$EM2$' and 'IM'. We ob-

	$Platen$	$EM1$	$Burrage$	$EM2$	IM
$\|principal\ \ error\|_1$	1.1111	0.6094	0.5833	0.1667	0.1667

Table 1: Norm of Principal Error Coefficients

serve, in Table 1, that the 1–norm principal error coefficients '$EM2$' and 'IM' methods are less than the 1–norm of principal error coefficients '$Platen$' , '$EM1$' and '$Burrage$' methods. Also the difference between the 1–norm of principal error coefficients the '$EM1$' and '$Burrage$' methods is very small, and this amount is better than the '$Platen$' method. In order to improve the results of employing the 'IM' method at each step, the stage–variable Y_2 will be solved by the fixed–point iteration scheme with starting value for this variable comes from the '$EM2$' method. Since $J_1 \sim N(0, h)$, so $J_1 = \sqrt{h}R_n$ where $R_n \sim N(0, 1)$. Hence for the stage–variable Y_2 in the 'IM' method let

$$G(Y_2) \equiv y_n + \frac{1}{4}h(g_0(y_n) + g_0(Y_2)) + \frac{1}{4}\sqrt{h}R_n(g_1(y_n) + g_1(Y_2)),$$

and therefore the fixed–point iteration for solving Y_2 is given by

$$Y_2{}^{[s+1]} = G(Y_2{}^{[s]}), \quad s = 0, 1, 2, \ldots$$

with stopping criteria

$$|Y_2^{[s+1]} - Y_2^{[s]}| < \epsilon \ , \tag{15}$$

where ϵ is a positive known tolerance number. Consequently the stage–variable Y_3 is given by

$$Y_3 = y_n + hg_0(Y_2{}^{[s+1]}) + \sqrt{h}R_n g_1(Y_2{}^{[s+1]}),$$

such that $Y_2^{[s+1]}$ satisfy condition (15). Finally y_{n+1} for the 'IM' method will be evaluated by

$$\begin{aligned} y_{n+1} = y_n + h(\frac{1}{6}g_0(y_n) + \frac{2}{3}g_0(Y_2{}^{[s+1]}) + \frac{1}{6}g_0(Y_3)) \\ +\sqrt{h}R_n(\frac{1}{6}g_1(y_n) + \frac{2}{3}g_1(Y_2{}^{[s+1]}) + \frac{1}{6}g_1(Y_3)), \end{aligned}$$

where $Y_2^{[s+1]}$ satisfies condition (15).

4 Numerical results and conclusion

In this section, numerical results from the implementation of five methods are presented. These methods are '$Platen$' , '$EM1$' , '$Burrage$' , '$EM2$' and 'IM'. The above methods will be implemented in fixed stepsize mode on two different problems. In order to simulate the Gaussian variable J_1 with distribution $N(0, h)$ we have taken pseudo–random numbers generated by "randn" in $MATLAB$ 7.0. When these methods are simulated, the same sequence of random numbers for the Wiener increment J_1 are used for the stepsize under consideration. For each problem it is necessary to simulate many trajectories of the Wiener process and we take, 1000, where K stands for the number of different realizations of the Wiener process. The implementation

determines the average error for each stepsize at the end point of the interval $[0, T]$ is defined by

$$AE = \frac{1}{K} \sum_{i=1}^{K} \mid y_N^{(i)} - y^{(i)}(t_N) \mid,$$

where $y_N^{(i)}$ is the approximation solution and $y^{(i)}(t_N)$ is the exact solution of SDE at $t_N = T$ in the i–th path of the Wiener process. The results appear in Tables 2, 3, 4 and 5. In all Tables the column 6 determines the average error for 'IM' method, while at each step starting value for the stage–variable Y_2 come from the '$EM2$' method with $\epsilon = 0.0001$.

Test Problem 1. Consider

$$\begin{cases} dy = -a^2 y(1-y^2)dt + a(1-y^2)dW(t) = a(1-y^2)odW(t), & t \in [0,1] \\ y(0) = 0. \end{cases}$$

The exact solution of this equation is (see [4])

$$y(t) = \tan h(aW(t) + \arctan h(y_0)).$$

This problem is purely stochastic and was solved numerically twice, with two choices of parameters $a = 2$ and $a = 1$.

h	$Platen$	$Burrage$	$EM1$	$EM2$	IM
$\frac{1}{25}$	0.23379	0.17543	0.14732	0.90772e–1	0.49411e–2
$\frac{1}{50}$	0.11910	0.87367e–1	0.72644e–1	0.27840e–1	0.10981e–2
$\frac{1}{100}$	0.61476e–1	0.45824e–1	0.38650e–1	0.10742e–1	0.34410e–3
$\frac{1}{200}$	0.33242e–1	0.25185e–1	0.21597e–1	0.46055e–2	0.11119e–3
$\frac{1}{400}$	0.15803e–1	0.11847e–1	0.10035e–1	0.19898e–2	0.39186e–4

Table 2: Global errors for test Problem 1, with $a = 2$, $K = 1000$ and $\epsilon = 0.0001$.

Test Problem 2. Consider

h	$Platen$	$Burrage$	$EM1$	$EM2$	IM
$\frac{1}{25}$	0.24542e–1	0.13005e–1	0.95001e–2	0.55887e–2	0.11868e–3
$\frac{1}{50}$	0.11982e–1	0.64091e–2	0.48965e–2	0.21210e–2	0.31872e–4
$\frac{1}{100}$	0.63255e–2	0.34932e–2	0.26472e–2	0.96237e–3	0.92418e–5
$\frac{1}{200}$	0.32726e–2	0.17777e–2	0.13249e–2	0.45166e–3	0.33780e–5
$\frac{1}{400}$	0.15283e–2	0.81432e–3	0.60395e–3	0.22187e–3	0.32333e–5

Table 3: Global errors for test Problem 1, with $a = 1,\ K = 1000$ and $\epsilon = 0.0001$.

$$\begin{cases} dy = -(\alpha + \beta^2 y)(1 - y^2)dt + \beta(1 - y^2)dW(t) \\ \qquad = -\alpha(1 - y^2)dt + \beta(1 - y^2)odW(t), \quad t \in [0, 1] \\ y(0) = 0. \end{cases}$$

For this problem, the solution is known to be (see [4])

$$y(t) = \frac{(1 + y_0)\exp(-2\alpha t + 2\beta W(t)) + y_0 - 1}{(1 + y_0)\exp(-2\alpha t + 2\beta W(t)) + 1 - y_0}.$$

This problem be solved with two different values of parameters where $\alpha = 1$ and $\beta = 2,\ 0.01$. In the first case, one has that the stochastic part is significant, whereas it is much smaller in the second case.

h	$Platen$	$Burrage$	$EM1$	$EM2$	IM
$\frac{1}{25}$	0.19910	0.15513	0.13032	0.75505e–1	0.40515e–2
$\frac{1}{50}$	0.10342	0.78391e–1	0.65485e–1	0.24255e–1	0.10537e–2
$\frac{1}{100}$	0.54342e–1	0.42437e–1	0.36003e–1	0.94268e–2	0.31865e–3
$\frac{1}{200}$	0.29818e–1	0.23167e–1	0.19813e–1	0.41433e–2	0.98992e–4
$\frac{1}{400}$	0.13516e–1	0.10575e–1	0.90643e–2	0.17987e–2	0.40232e–4

Table 4: Global errors for test Problem 2, with $\alpha = 1,\ \beta = 2,\ K = 1000$ and $\epsilon = 0.0001$.

h	$Platen$	$Burrage$	$EM1$	$EM2$	IM
$\frac{1}{25}$	0.73817e–2	0.11148e–3	0.66508e–4	0.23513e–5	0.23346e–6
$\frac{1}{50}$	0.36668e–2	0.27635e–4	0.16562e–4	0.28906e–6	0.28717e–7
$\frac{1}{100}$	0.18273e–2	0.69352e–5	0.41577e–5	0.36640e–7	0.36196e–8
$\frac{1}{200}$	0.91215e–3	0.17761e–5	0.10659e–5	0.48004e–8	0.47420e–9
$\frac{1}{400}$	0.45571e–3	0.46845e–6	0.28133e–6	0.66410e–9	0.48845e–10

Table 5: Global errors for test Problem 2, with $\alpha = 1,\ \beta = 0.01,\ K = 1000$ and $\epsilon = 0.0001$.

With comparing the results in Tables 2, 3, 4 and 5, we conclude that the 'IM' and '$EM2$' methods are more accurate than the '$Platen$' , '$Burrage$' and '$EM1$' methods, as the error values for 'IM' method is less than '$EM2$' method. Moreover for two–stage SRK methods the '$EM1$' method is more effective '$Platen$' and '$Burrage$' methods. On the other hand for problems in which the deterministic term dominate (test problem 2 with $\beta = 0.01$) the improvement the '$Burrage$' , '$EM1$' , '$EM2$' and 'IM' methods becomes noticeable as the step-size is reduced. This is because the deterministic component of the '$Burrage$' and '$EM1$' methods are the second order Runge–Kutta methods, while the deterministic component '$EM2$' and 'IM' methods are the third and fourth order Runge–Kutta methods, respectively.

References

[1] Burrage K. and Burrage P. M. , *High strong explicit Runge–Kutta methods for stochastic ordinary differential equations*, Applied Numer. Mathematics, **22** , 81–101, (1996).

[2] Burrage K. and Burrage P. M. , *Order conditions of stochastic Runge–Kutta methods by B–series*, SIAM J. Numer. Anal. **38**, 1626–1646, (2000).

[3] Butcher J. C. , *Numerical Methods for Ordinary Differential Equations*, Wiley, U.K., (2003).

[4] Kloeden P.E. , Platen E , *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, (1995).

[5] Namjoo M. , Soheili A. R. , *An Estimate of the Error for Strong Solutions of Stochastic Differential Equations*, Int. Math. Forum, **2**, no. **6** , 251–262, (2007).

[6] Soheili A. R. and Namjoo M., *Strong Runge–Kutta methods With order one for Numerical Solution of Itô Stochastic Differential Equations*, Applied Mathematics Research eXpress, Article ID abm003, Oxford University Press, 17 pages, (2007).

On using wavelet thresholding for independent component analysis

Vahid Nassiri, and Mina Aminghafari

Department of Statistics, Faculty of Mathematics and Computer Science, Amirkabir University of Technology (Tehran Polytechnic)

Abstract: A novel wavelet-based algorithm for the noisy independent component analysis (ICA) problem is proposed. We show that in special situations the classical solutions of the noisy ICA are failed or may be inadequate. This new approach combines separate results to achieve better performance and overcome the shortcomings of the classical noisy ICA solution.

Keywords: ICA, BSS, wavelets, thresholding, α-stable family of distributions.

1 Introduction

Independent component analysis (ICA) is a recently developed and powerful multivariate statistical method which is a solution for blind source separation (BSS) problem. A simple description of BSS problem is a cocktail party. Consider two people are speaking simultaneously in a room. There are two microphones in different locations which record two signals. Obviously each of these signals is a mixture of the signals emitted by each person. In a linear approach, it is common to express this as linear equations as follows:

$$X_1 = a_{11}S_1 + a_{12}S_2$$

$$X_2 = a_{21}S_1 + a_{22}S_2$$

in matrix notation one can writes $X = AS$, where A is a square invertible mixing matrix. Now the problem is finding S, just with knowing X. This is why it is called blind; Since we have nothing more than an observed sample. If A is known, the problem is a simple linear system of equations, but in real world A is unknown. In order to solve this problem by ICA, we assume that S_is $(i = 1, \ldots, k)$ are

independent. This assumption is logical in many cases. ICA tries to find such an A which S_is are as independent as possible.

Wavelets are useful mathematical tools which have been widely used in statistics recent years. For a complete review of statistical applications of wavelets see e.g. Abramovich et al. (2000).

There are several methods have constructed separately which use wavelets or solve the ICA problem, see e.g. Hyvärinen et al. (2001) or Mallat (1999). In this paper we try to combine them such that, by some aspects, the result is better than the other similar techniques.

This paper is organized as follows. In Section 2 we look more precisely at ICA problem and review some of its classical solutions. In Section 3 some preliminary remarks are formulated about two technical issues: the wavelets and α-stable family of distributions. Section 4 dedicates to study the shortcomings of the techniques which are discussed in the previous sections. In this section some solutions for overcoming these shortcomings will be proposed. In Section 5 we introduce our new algorithm which combines separate techniques to achieve better results. Finally in Section 6 some conclusions are provided.

2 Independent component analysis and its classical solutions

The main problem of ICA is discussed in previous section. In fact the key factor to solve the ICA problem is using the statistical properties of the sources which we wish to find them blindly. A logical statistical property which can be considered is the *independence*. The ICA model is considered as follows:

$$X = AS \tag{1}$$

and one wants to find such a $W = A^{-1}$ which WX are as independent as possible. The main classical approaches to solve ICA are 1.Maximizing the non-Gaussianity, 2.Minimizing the mutual information and 3.Using the InfoMax principle.

A known restriction in ICA is that the source signals cannot come

from normal distribution. The reason is obvious, since in many situations the assumption is that A in (1) is orthogonal (similarly in PCA) and one knows uncorrelatedness and independence are the same in the Gaussian distribution. Comon (1994) used this statistical fact to find W. In this approach some measures of non-Gaussianity such as kurtosis are applied.

The two last methods give the same results. Mutual information is a measure of the information that members of a set of random variables have on the other random variables in the set. Therefore, when the mutual information be minimized the members of a set of random variables has the least information on each other and this means independence. Instead of mutual information, an equivalence quantity which can be used is the entropy.

Assume that $H(X)$ denotes the entropy of X.

$$H(X) = -\int_{-\infty}^{\infty} \log\{f_X(x)\} f_X(x) dx = \mathrm{E}(\log f_X(x))$$

Using entropy, $I(X_1, \ldots, X_n)$, the mutual information between n random variables $(X_i, i = 1, \ldots, n)$ defined as follows:

$$I(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i) - H(X_1, \ldots, X_n)$$

Therefore, it is obvious that minimizing the mutual information is the same as maximizing the entropy $H(X_1, \ldots, X_n)$.

Now we study the entropy of a linear transformation. Consider an invertible transformation of the random vector $(X_1, \ldots, X_n)$.

$$Y = MX$$

where M is a linear filter. It is easy to show that:

$$H(Y) = H(X) + \log|\det M| \tag{2}$$

Now remember the ICA problem. As we mentioned maximum entropy means independence. Consider M as the de-mixing matrix in

(1). Therefore, by maximizing (2) with respect to W (i.e. the demixing matrix) the independent sources can be achieved. There are some standard numerical methods exist for this purpose. One of the most famous ones is the gradient descend. Bell and Sejnowsky (1995) used the InfoMax principle.

3 Wavelets and α-stable family of distributions

The definitions of wavelets is different according to how the multiresolution analysis defined. For the sake of convenience we set our definitions here. In addition we need α-stable distributions in our work. Therefore, definitions about this family will present in this section.

3.1 Wavelets

Let ϕ be a function of $L^2(\mathbb{R})$ such that the family of translates $\{\phi(.-k), k \in \mathbb{Z}\}$ is an orthogonal system; let $V_j \subset L^2(R)$ be the subspace spanned by $\{\phi_{jk} = 2^{j/2}\phi(2^j.-k), k \in \mathbb{Z}\}$. By definition, the sequence of spaces (V_j), $j \in \mathbb{Z}$, is called a multiresolution analysis (MRA) of $L^2(R)$ if $V_j \subset V_{j+1}$ and $\bigcup_{j\geq 0} V_j$ is dense in $L^2(\mathbb{R})$. ϕ is called the father wavelet or the scaling function.

Let $(V_j)_{j\in\mathbb{Z}}$ be a multiresolution analysis of $L^2(R)$, with V_j spanned by $\{\phi_{jk} = 2^{j/2}\phi(2^j.-k), k \in \mathbb{Z}\}$. Define W_j as the complement of V_j in V_{j+1}, and let the family $\{\psi_{jk}, k \in \mathbb{Z}\}$ be a basis for W_j, with $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$. Let $\alpha_{jk}(f) = \langle f, \phi_{jk}\rangle$ and $\beta_{jk}(f) = \langle f, \psi_{jk}\rangle$. Where $\langle a, b\rangle$ denotes the inner product of a and b. A function $f \in L^2(\mathbb{R})$ admits a wavelet expansion on $(V_j)_{j\in\mathbb{Z}}$ if the series

$$\sum_k \alpha_{j_0k}(f)\phi_{j_0k} + \sum_{j=j_0}^{\infty}\sum_k \beta_{jk}(f)\psi_{jk}$$

is convergent to f in $L^2(\mathbb{R})$ for every $j_0 \in \mathbb{N}$. ψ is called the mother wavelet.

The definition of a multiresolution analysis on $L^2(\mathbb{R}^d)$ follows the same pattern. i.e. an MRA in dimension one induces an associated

MRA in dimension d, using the following tensorial product, see Meyer (1997).

Define V_j^d as the tensorial product of d copies of V_j. The increasing sequence $(V_j^d)_{j\in\mathbb{Z}}$ defines a multiresolution analysis of $L^2(\mathbb{R}^d)$): for $(i^1,\ldots,i^d)\in\{0,1\}^d$ and $(i^1,\ldots,i^d)\neq(0,\ldots,0)$, define:

$$\Psi(x)_{i^1,\ldots,i^d}=\prod_{l=1}^{n}\psi^{(i^l)}(x^l)$$

with $\psi^{(0)}=\phi$, $\psi^{(1)}=\psi$. So that ϕ appears at least once in the product $\Phi(x)$. For $(i^1,\ldots,i^d)=(0,\ldots,0)$, define $\Phi(x)=\prod_{l=1}^{d}\phi(x^l)$. Define W_j^d as the orthogonal complement of V_j^d in V_{j+1}^d.

A function f admits a wavelet expansion with respect to the basis (Φ,Ψ) if the series

$$\sum_{k\in\mathbb{Z}^d}\alpha_{j_0k}(f)\Phi j_0k+\sum_{j=j_0}^{\infty}\sum_{k\in\mathbb{Z}^d}\beta_{jk}(f)\Psi_{jk}$$

is convergent to f in $L^2(\mathbb{R}^d)$.

In connection with function approximation, wavelets can be viewed as falling in the category of orthogonal series methods, or also in the category of kernel-based methods. The approximation at level j of a function f that admits a multiresolution expansion is the orthogonal projet P_jf of f onto $V_j\subset L^2(\mathbb{R})$ defined by:

$$(P_jf)(x)=\sum_{k\in\mathbb{Z}^d}\alpha_{jk}\Phi_{jk}(x)$$

3.2 α-stable distributions

Stable distributions have a long history in the subject of probability. They form a subset of the class of so-called *infinite divisible* distributions, a class of characteristic functions at the heart of generalized central limit theory. Roughly speaking, the stable distributions are

those which are closed under the formation of linear combinations, i.e. linear mixtures of stable distributions result in distribution of the same type.

Closed form expressions do not exist in general for stable density and distribution function. Therefore, they are usually defined via the characteristic function.

Definition 1 *A real random variable X has stable distribution with characteristic exponent α, if its characteristic function can be written as follows:*

$$\Phi(t) = \exp\{-I(t) + i\mu t\}$$

where

$$I(t) = \begin{cases} \gamma^\alpha |t|^\alpha \Big(1 - i\beta \mathrm{sgn}(t) \tan(\frac{\pi\alpha}{2})\Big), & \alpha \neq 1, \\ \gamma\ |t| \Big(1 + i\beta \frac{2}{\pi} \mathrm{sgn}(t) \ln|t|\Big), & \alpha = 1, \end{cases} \tag{3}$$

and $\mathrm{sgn}(u) = -1,\ 0,$ *or* 1 *if* $u <, =,$ *or* > 0*, respectively.*

Now we define a stable random vector.

Definition 2 *A real random vector $\boldsymbol{X}$ has stable distribution with characteristic exponent α, if a finite measure Γ exists on the unit sphere S_M of $\mathbb{R}^M$ such that the characteristic function can be written as follows:*

$$\Phi(\boldsymbol{t}) = \exp\{-I(\boldsymbol{t}) + i\boldsymbol{t}^T\mu\}$$

where

$$I(\boldsymbol{t}) = \int_{S_M} \psi_\alpha(\boldsymbol{t}^T\boldsymbol{s}) d\Gamma(\boldsymbol{s})$$

and

$$\psi_\alpha(u) = \begin{cases} |u|^\alpha \Big(1 - i\mathrm{sgn}(u) \tan(\frac{\pi\alpha}{2})\Big), & \alpha \neq 1, \\ |u| \Big(1 + i\frac{2}{\pi} \mathrm{sgn}(u) \ln|u|\Big), & \alpha = 1, \end{cases} \tag{4}$$

The measure Γ is called the spectral measure *of the random vector $\boldsymbol{X}$, and μ is called the shift vector.*

The spectral measure is a real function and the spectral representation (Γ, μ) is unique. For a reference on stable distributions, see e.g. Samorodnitsky and Taqqu (1994).

4 Two main shortcomings of the classical ICA solutions and how to overcome

In this section we describe two main shortcomings of the classical ICA solutions and propose some methods which can overcome these shortcomings.

4.1 Noisy ICA: classical and proposed overcoming methods

Consider (1) again. In the real world what always exists is some kind of noise. There are many resources can be considered for the noise, such as measuring devices, situation or experimenter conditions and more. Existing of the noise is true also for the ICA model, so one should use noisy version of ICA, as follows:

$$X = AS + \epsilon \tag{5}$$

where ϵ is an additive noise and $\mathrm{Var}(\epsilon) = \Sigma_\epsilon$.

There are various methods have proposed to extracting the original signal from the noise. A useful and commonly recent used technique is using wavelets. Projecting on V_j (by father wavelet or ϕ) and W_j (by mother wavelet or ψ) enables one to analyze the time domain and spectral domain simultaneously. This ability with using wavelet thresholding are great for de-noising task, since low-pass and high-pass filters allow one to extract original signal and the noise at an appropriate level. Before this interesting property of wavelets, the main reason which make them very useful is their local behavior.

Some authors used wavelets to solve the noisy ICA problem. But their method ignores many valuable information which lies in the dependence structure of X_is, where Σ_ϵ is not diagonal. Since they use wavelet denoising on each column (X_i) separately. If the data are uncorrelated, this method may be suitable. But, in the case of ICA, the

variables are surely dependent and the target is finding the independent components. Therefore, this procedure for denoising failed to use for ICA.

Aminghafari et al. (2006) proposed a new algorithm which consider the correlation between variables. Precisely they introduced a multivariate denoising method which is using wavelet transform and principal component analysis (PCA) together.

The general denoising procedure introduced by Aminghafari et al. (2006) is as follows:

Step 1 Perform the wavelet transform at level J of each column of X.

Step 2 Define $\hat{\Sigma}_\epsilon$ the estimator of the noise covariance matrix as $\hat{\Sigma}_\epsilon = \text{MCD}(D_1)$ and compute V such that $\hat{\Sigma}_\epsilon = V\Lambda V^T$ where V^T is the transpose of V and $\Lambda = \text{diag}(\lambda_i)$. Apply to each detail after change of basis (namely D_jV), the p-univariate thresholding strategies using the threshold $t_i = \sqrt{2\lambda_i \log(n)}$ for the i-th column of D_jV.

Step 3 Reconstruct the denoised matric from the simplified detail and approximation matrices, by changing of basis using V^T and inverting the wavelet transform.

Step 4 Perform a final PCA of the denoised matrix obtained at step 4.

The MCD which is denoted in Step 2 is the minimum covariance determinant estimator which introduced by Rousseeuw (1984). It is a generalized version of the robust minimum variance estimators. Note that in multivariate statistics an index of variance is the determinant of the covariance matrix. MCD looks for the h-subset (typically $h = 0.75n$) of data with the smallest determinant of its covariance matrix. More recently, Rosseeuw and Van Driessen (1999) introduced a fast version of MCD which allows fast computation of this estimator.

There are useful remarks can be find in Aminghafari et al. (2006) about how to select the appropriate principal components or the de-

composition level J.

4.2 Stable rules are breaking the rules

In many ICA solutions the approach in nonparametric, i.e. there is no condition on the distribution of the source signals. But, some times these conditions are necessary.

The α-stable distributions were discussed in previous section. Now we present a main theorem which stands at the end of some theorems and results and shows that for the case of stable sources many of the classical ICA solutions are failed to use. For more on this topic and proof of the theorems see Der (2003).

Theorem 1 *Let S be a vector of n independent random variables drawn from a stable symmetric class ($\beta = 0$). Let $X = AS$, where A is a real matrix. If $h(x_1, \ldots, x_n)$ is a function of only the marginal distributions of X, then $h(X(A))$ is a constant.*

Above theorem shows that any contrast extremising any property of the marginal distributions of the mixtures (e.g. negentropy or kurtosis) is constant for stable symmetric sources.

According to the Theorem 1 where the sources are stable, we need an alternative method which find the ICs directly via the joint and marginal distributions of mixtures.

By the assumption of ICA, f, the density function of source signals S, is completely factorizable in a product of marginal distributions $f^1, \ldots, f^d$. Using the model in the Equation (1) the density function of mixed signal can be obtained as:

$$f_A(x) = |\det(A^{-1})| f(A^{-1}x) = |\det B| f^1(b_1 x) \ldots f^d(b_d x)$$

where b^l is the l-th row of the matrix $B = A^{-1}$.

A more logical and simple contrast for independence introduced by Rosenblatt (1975) as follows:

$$C(f_A) = ||f_A - f_A^{*1} \ldots f_A^{*d}||_2^2 \tag{6}$$

with f_A^{*l} is the marginal of the multivariate density f_A in dimension l. He used this contrast to construct a test for independence.

Consider the following theorem.

Theorem 2 *Let $f \in \mathbb{R}^d$ and its marginal distributions $f^1, \ldots, f^d \in \mathbb{R}$. $C(f) = 0$ if and only if $f = f^1 \ldots f^d$.*

Proof: Let $C(f) = 0$ then the joint density is equal to product of its marginal distributions, i.e. $f = f^1 \ldots f^d$. The sufficiency can be proved by the same way.

Using the contrast in (6) one should estimate $d + 1$ density functions: one multivariate density and d univariate densities. After this step by using optimization methods, one should find such an A which minimizes the contrast in (6). We denote the estimated B by W, i.e. W is a logical candidate for the inverse of the A.

Barbedor (2007) used wavelet density estimators to find W. In the next section we Introduce an algorithm which extends his method for the noisy ICA.

5 New wavelet-based ICA solution

Wavelets are recently developed tools for probability density functions estimating. Consider the noise-free ICA model in (1). By Theorem 2, it suffices to find such a W which minimize the contrast in (6). For this purpose first the density functions should be estimated.

For density estimation using wavelets a perfect and complete reference is the work of Donoho et al. (1996).

The linear wavelet estimator of the Rosenblatt's contrast is as follows:

$$\hat{C}_j = \sum_{k^1,\ldots,k^d} (\alpha_{jk} - \alpha_{jk^1,\ldots,jd^d})^2$$

where the parameters α_{jk} and $\alpha_{jk^1,\ldots,jd^d}$ can be estimated as follows:

$$\hat{\alpha}_{jk^1,\ldots,jd^d} = \frac{1}{n} \sum_{i=1}^{n} \phi_{jk^1}(X_i^1) \ldots \phi_{jk^d}(X_i^d)$$

$$\hat{\alpha}_{jk^l} = \frac{1}{n}\sum_{i=1}^{n} \phi_{jk^l}(X_i^l)$$

The following lemma allows us to center the design matrix, as it is usual in the ICA methods.

Lemma 1 *Using centered X (i.e. $X - E[X]$) does not change the matrix A.*

Proof: $X = AS \Rightarrow X - \mathrm{E}[X] = AS - \mathrm{E}[AS] = AS - A\mathrm{E}[S] = A(S - \mathrm{E}[S])$.

Now we present the new algorithm for solving noisy ICA by wavelet thresholding.

Step 1 Center the design matrix X. Call it $\tilde{X} = X - \mathrm{E}[X]$.

Step 2 Find the denoised $\tilde{X}$ based on the 4-steps procedure described in (4.1). Call it $\check{X}$.

Step 3 Choosing an initial W, e.g. the identity matrix.

Step 4 Finding $W\check{X}$ and computing the contrast in (6) for this matrix.

Step 5 Iterating steps 4 and 5 based on an appropriate minimizing procedure till the best W obtained. Call it $\hat{W}$.

Step 6 Finding $\hat{W}\check{X}$ as an estimation for the independent components.

6 Conclusions

In this article first we show that the denoising techniques which are used in noisy version of ICA are very poor, since of the surely dependence between observed signals. Also it has shown that for the case of stable distributions the classical ICA solutions are failed to use. In order to overcome these shortcomings we combine separately developed

techniques and obtained useful results to achieve a better solution for the ICA problem. This new algorithm can be applied wherever.

For future research we plan to use better thresholding rules and find some applicable ways for choosing the best resolution level J.

References

[1] Abramovich, F., Bailey, T. C., Sapatina, T. (2000). Wavelet analysis and its statistical applications. *The Statistician*, **49**, 1-29.

[2] Aminghafari, M., Cheze, N., Poggi, J.M., (2006). Multivariate denoising using wavelets and principal component analysis. *Computational Statistics and Data Analysis*, **50**, 2381-2398.

[3] Barbedor, P. (2007). Wavelet analysis and its statistical applications. *Test*, DOI 10.1007/s11749-007-0073-7.

[4] Bell, A.J., Sejnowsky, T.J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, **7**, 1129-1159.

[5] Comon, P. (1994). Independent component analysis, a new concept? *signal Processing*, **36**, 287-314.

[6] Der, R.(2003). *Stable Symmetric Distributions and Their Role In The Signal Separation Problem*. Technical repost in telecommunications & signal processing labratoary. McGill University, Montreal.

[7] Donoho, D.L., Johnstone, I.M, Kerkyacharian, G., Picard, D. (1996). Density estimation by wavelet thresholding. *Annals of Statistics*, **24**, 508-539.

[8] Hyvärinen, A., Karhunen, J., Oja, E.(2001). *Independent Component Analysis*. John Wiley, New York.

[9] Mallat, S.(1999). *A Wavelet Tour of Signal Processing*. Academic Press, London.

[10] Meyer, Y.(1997). *Ondelettes et Opérateurs*. Hermann, Paris.

[11] Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test for independence. *Annals of Statistics*, **3**, 1-14.

[12] Rousseeuw, P. (1984). Least median of squares regression. *Journal of American Statstical Association*, **79**, 871-880.

[13] Rousseeuw, P., Van Driessen, K., (1999). A fast algorithm for the minimum covariance deteminant estimator. *Technometrics*, **41**, 212-223.

[14] Samorodnitsky, G., Taqqu, M. S., (1994). *Stable Non-Gaussian Random Processes*. Chapman & Hall, New York.

Periodically correlated time series and their spectra

Alireza Nematollahi*

Department of Statistics, Shiraz University

Abstract: In this article, we briefly review the spectral analysis problem, and introduce a technique for the spectral estimation in the both context of univariate and multivariate stationary time series. Then, we consider periodically correlated (cyclostationary) time series, their spectra and also the estimation of these spectra using the mentioned technique. The method of estimation is illustrated with simulated and real time series.

Keywords: Spectral density function; Multivariate time series; stationary vector time series.

1 Introduction

Spectral analysis considers the problem of determining the distribution of total power over frequency, and the spectral estimation problem tries to estimate this distribution from a finite record of a data sequence, by means of either nonparametric or parametric techniques. Spectral analysis have widespread applications in various fields, including: economics, meteorology, astronomy, geology, radar and sonar systems, medicine, seismology, control systems and hydrology. In this paper, we briefly review the spectral analysis problem, and introduce a technique in the spectral estimation in both context of univariate and multivariate stationary time series. Then, we consider periodically correlated time series, their spectra and also the estimation of these spectra. We use the well known relation between the spectral density matrix of a periodically correlated time series and a stationary vector time series (Gladyshev, 1961). The method of estimation is illustrated with simulated and real time series.

Invited speaker

2 The spectrum of a stationary time series

Let $\{X_t\}$ be a discrete parameter, zero mean and real stationary time series and let $E|X_t|^2 < \infty$, $t \in Z$, where Z stands for all integers. Suppose $R(\tau) = E(X_\tau X_{t+\tau}), \tau \in Z$, is the autocovariance function of X_t satisfying $\sum_{\tau=-\infty}^{\infty} |R(\tau)| < \infty$. The power spectral density function of $\{X_t\}$ is defined by

$$h(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} R(\tau) e^{-i\omega\tau}, \ -\infty < \omega < \infty.$$

In the multivariate case, we suppose $\{\mathbf{X}_t, t \in Z\}$ is a zero mean vector, discrete-parameter and second order T-dimensional stationary series with $X_t(j)$, $j = 0, ..., T-1$, as its jth element and $\mathbf{R}(\tau) = E(\mathbf{X}_{t+\tau}\mathbf{X}_t')$, $\tau \in Z$, be the autocovariance matrix of $\mathbf{X}_t$. We assume that the series has an absolutely continuous spectrum and let $\mathbf{h}(\omega) = [h_{jk}(\omega)]_{j,k=0,1,...,T-1}$, denote its spectral density matrix, i.e.,

$$\mathbf{h}(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \mathbf{R}(\tau) e^{-i\omega\tau}, \ 0 \le \omega \le 2\pi.$$

3 The methods of estimation of the spectrum

A common and natural spectral estimator is based on a function called periodogram of data set $\{X_1, X_2, ..., X_n\}$, which is defined by

$$I_n(\omega) := \frac{1}{n} \left| \sum_{t=1}^{n} X_t e^{-it\omega} \right|^2.$$

It can be shown that $I_n(\omega)$ is asymptotically unbiased estimator of $2\pi h(\omega)$. But a critical disadvantage of the periodogram as an estimate of the power spectrum $h(\omega)$ is that for each $\omega \in [0, \pi]$, and $\varepsilon > 0$,

$$P\left(|I_n(\omega) - 2\pi h(\omega)| > \varepsilon\right) \to p > 0,$$

as $n \to \infty$. This means $I_n(\omega)$ is not a consistent estimator of $2\pi h(\omega)$. So for a process with a continuous spectrum, the periodogram provides a poor estimate and needs to be modified (smoothing periodogram).

In some cases, the classical methods (using periodogram) for estimating the spectrum (specially when the sample size is small) have not enough precision and they are not able to distinguish two strong peaks from each other. Two particular estimates (which are called high resolution estimators) for solving this problem are given by Capon (1969) and Pisarenko (1972), and these are widely used in signal processing problems.

Subba Rao and Gabr (1989) considered the function,

$$h_n(\omega) = \frac{1}{2\pi n} \sum_{t=1}^{n} \sum_{s=1}^{n} R(t-s) \exp(-i(t-s)\omega),$$

which is asymptotically equivalent to estimating $h(\omega)$ (and called $h_n(\omega)$ the "truncated spectral density function"). They showed that $h_n(\omega)$ can be written as

$$h_n(\omega_l) = \frac{1}{4\pi} \left\{ \lambda_{n,0} A_{n,0}(\omega_l) + \sum_{j=1}^{(n-1)/2} (\lambda_{n,2j} + \lambda_{n,2j-1}) A_{n,2j}(\omega_l) \right\},$$

where $A_{n,0}(\omega_l) = \frac{2}{n} 2\pi F_{n-1}(\omega_l)$, $A_{n,2j}(\omega_l) = A_{n,2j-1}(\omega_l) = \frac{1}{n}\{2\pi F_{n-1}(\omega_j + \omega_l) + 2\pi F_{n-1}(\omega_j - \omega_l)\}$ and $\lambda_{n,0}, \lambda_{n,1}, ..., \lambda_{n,n}$ are the eigenvalues of the Toeplitz matrix $\mathbf{R}_n$ given by

$$\mathbf{R}_n = \begin{pmatrix} R(0) & R(1) & ... & R(n-1) \\ R(-1) & R(0) & ... & R(n-2) \\ . & . & . & . \\ R(-(n-1)) & R(-(n-2)) & ... & R(0) \end{pmatrix},$$

and $F_n(\theta)$ is the Fejér kernel (e.g. Priestley, 1989).

Note that $h_n(\omega_l)$ is a smoothed version of the eigenvalues of $\mathbf{R}_n$, and the smoothing function (lag window) is the Fejér kernel, it is linear in the eigenvalues $\lambda_{n,j}$ and that $A_{n,j}(\omega_l)$ does not depend on the process. We can consider the function

$$h_{n,P}(\omega_l) = g\left\{\sum_{j=0}^{n-1} G(\lambda_{n,j})A_{n,j}(\omega_l)\right\},$$

as an approximation to $h_n(\omega)$, where $G(.)$ is a strictly monotone continuous function over the interval $(0,\infty)$ and let $g(.)$ be an inverse function, i.e. $g(G(x)) = x$.

In fact, this is the way that Pisarenko (1972) derived his estimate, and $h_{n,P}(\omega_l)$ is the theoretical form of the general Pisarenko estimate. Now substitute $G(x) = x^{-1}$ in $h_{n,P}(\omega_l)$, then we obtain the theoretical form of Capon's estimator (Capon, 1969),

$$h_{n,Cap}(\omega_l) = \frac{1}{\pi}\left\{\sum_{j=0}^{n-1} \lambda_{n,j}^{-1}A_{n,j}(\omega_l)\right\}^{-1}.$$

In the multivariate case, let $\{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n\}$ be a sample of size n from $\{\mathbf{X}_t\}$ and $\mathbf{I}_n(\omega) = [I_{n,jk}(\omega)]_{j,k=0,...,T-1}$ be the periodogram matrix. Consider the truncated spectral density matrix

$$\mathbf{h}_n(\omega) = \frac{1}{2\pi n}\sum_{t=1}^{n}\sum_{s=1}^{n}\mathbf{R}(t-s)e^{-i(t-s)\omega}.$$

It can be shown that $\mathbf{I}_n(\omega)$ is an unbiased estimator of $\mathbf{h}_n(\omega)$ and asymptotically unbiased estimator of $\mathbf{h}(\omega)$, (Priestley, 1989), but of course, it is not consistent (Brillinger, 1975). In order to find a consistent estimator of $\mathbf{h}(\omega)$, the usual procedure is to smooth the periodogram by a suitable kernel.

Nematollahi and Subba Rao (2005) showed that $\mathbf{h}_n(\omega)$ can be written in terms of the eigenvalues of the variance covariance matrix, then

they estimated $\mathbf{h}(\omega)$ using the eigenvalue decomposition of the sample variance covariance matrix. More precisely, define a $nT \times 1$ vector $\mathbb{X}_n = (\mathbf{X}'_n, \mathbf{X}'_{n-1}, ..., \mathbf{X}'_1)'$ and let $\mathbf{\Gamma}_n = E\mathbb{X}_n\mathbb{X}'_n$ be its variance-covariance matrix. We have

$$\mathbf{\Gamma}_n = \begin{pmatrix} \mathbf{R}(0) & \mathbf{R}(1) & ... & \mathbf{R}(n-1) \\ \mathbf{R}(-1) & \mathbf{R}(0) & ... & \mathbf{R}(n-2) \\ . & . & . & . \\ \mathbf{R}(-(n-1)) & \mathbf{R}(-(n-2)) & ... & \mathbf{R}(0) \end{pmatrix}.$$

We note that $\mathbf{\Gamma}_n$ is a block-Toeplitz matrix. Individual matrix elements are not, in general, symmetric ($\mathbf{R}(\tau) \neq \mathbf{R}'(\tau)$), although $\mathbf{R}(-\tau) = \mathbf{R}'(\tau)$.

They derived the following expression for estimation of spectral density matrix $\mathbf{h}(\omega)$

$$\mathbf{h}_n(\omega_l) = \frac{1}{2\pi n} \sum_{j=0}^{n-1} \mathbf{B}^*_n(\omega_j, \omega_l) \mathbf{\Lambda}_n(\omega_j) \mathbf{B}_n(\omega_j, \omega_l),$$

where $\mathbf{B}_n(\omega_j, \omega_l) = \sum_{s=1}^{n} \mathbf{W}^s_n(\omega_j) \exp(is\omega_l)$ and $\mathbf{W}^s_n(\omega_j)$ is the s-th block of $\mathbf{W}_n(\omega_j)$. Here $\mathbf{\Lambda}_n(\omega_j)$ is an "eigenvalue-matrix" of $\mathbf{\Gamma}_n$ and $\mathbf{W}_n(\omega_j)$ is an "eigenvector-matrix" associated with $\mathbf{\Lambda}_n(\omega_j)$. We thus have

$$\mathbf{h}_n(\omega_l) = \frac{1}{4\pi} \sum_{j=0}^{n-1} \mathbf{A}_n(\omega_j, \omega_l),$$

where

$$\begin{aligned} \mathbf{A}_n(\omega_j, \omega_l) &= \frac{2}{n} \mathbf{B}^*_n(\omega_j, \omega_l) \mathbf{\Lambda}_n(\omega_j) \mathbf{B}_n(\omega_j, \omega_l) \\ &= \frac{2}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} \mathbf{W}^{t*}_n(\omega_j) \mathbf{\Lambda}_n(\omega_j) \mathbf{W}^s_n(\omega_j) e^{-i(t-s)\omega_l}. \end{aligned}$$

In finding the above estimator, the asymptotic equivalence of $\mathbf{R}_n$ with a related circular symmetric matrix is used which have been proved in the multivariate case by Nematollahi and Shishebor (2005) and is a multivariate generalization of the well-known result of Gray (1972). Nematollahi and Subba Rao (2005) showed that an equivalence form for $\mathbf{h}_n(\omega_l)$ is

$$\mathbf{h}_n(\omega_l) = \frac{1}{4\pi}(\frac{2}{n}2\pi F_{n-1}(\omega_l)\mathbf{\Lambda}_n(\omega_0) + 2\sum_{j=1}^{\frac{n-1}{2}}\frac{1}{n}\mathbf{\Lambda}_n(\omega_{2j})[2\pi F_{n-1}(\omega_j+\omega_l)+2\pi F_{n-1}(\omega_j-\omega_l)]).$$

The above approximate relation tells us that the spectral density function $\mathbf{h}_n(\omega_l)$ is in fact a smooth function of $\mathbf{\Lambda}_n(\omega_j)$ and the smoothing function is the well known Fejér kernel. We observe that $\mathbf{h}_n(\omega_l)$ is linear in $\mathbf{\Lambda}_n(\omega_j)$ and that $\mathbf{A}_n(\omega_j,\omega_l)$ does not depend on the time series $\{\mathbf{X}_t\}$. Similar to the univariate case, this suggests that these eigenvalue-matrices can be replaced by any nonlinear function of $\mathbf{\Lambda}_n(\omega_j)$ and by suitably defining an inverse function, we can, in the limiting form, recover the original spectrum. Consider the function

$$\mathbf{h}_{n,P}(\omega_l) = g\left\{\sum_{j=0}^{n-1}\frac{2}{n}\mathbf{B}_n^*(\omega_j,\omega_l)G(\mathbf{\Lambda}_n(\omega_j))\mathbf{B}_n(\omega_j,\omega_l)\right\},$$

as an approximation to $\mathbf{h}_n(\omega_l)$, where $g(G(x)) = x$. In fact, this is a generalization of the way that Pisarenko (1972) derived his estimate in the univariate case, and $\mathbf{h}_{n,P}(\omega_l)$ is the generalization of the theoretical form of the Pisarenko's estimator to the multivariate case.

A multivariate generalization and also an explicit expression for the high resolution spectral density matrix (a generalization of Capon's estimator) of the vector series $\mathbf{X}_t$ are also given by Nematollahi and Subba Rao (2005). They derived an appropriate minimum variance (MV) spectral estimator with form

$$\mathbf{h}_{n,cap}(\omega) = \frac{1}{\pi}[\frac{2}{n}\mathbf{L}_n^*(\omega)\Gamma_n^{-1}\mathbf{L}_n(\omega)]^{-1},$$

where $\mathbf{L}_n(\omega_l) = (\mathbf{I}, e^{i\omega_l}\mathbf{I}, ..., e^{ni\omega_l}\mathbf{I})'$.

It can be shown that the theoretical form of the generalized Capon's estimator can also be written as

$$\mathbf{h}_{n,cap}(\omega_l) = \frac{1}{\pi}\left\{\sum_{j=0}^{n-1}\widetilde{\mathbf{A}_n}(\omega_j, \omega_l)\right\}^{-1},$$

where

$$\widetilde{\mathbf{A}_n}(\omega_j, \omega_l) = \frac{2}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\mathbf{W}_n^{t*}(\omega_j)\mathbf{\Lambda}_n^{-1}(\omega_j)\mathbf{W}_n^{s}(\omega_j)e^{-i(t-s)\omega_l}.$$

This is a multivariate generalization of the *minimum variance spectral (MVS) estimator* due to Capon (1969).

4 Periodically correlated processes and their applications

Conventional time series analysis is heavily dependent on the assumption of stationarity. But this assumption is unsatisfactory for many physical process of interest. In fact, many seasonal time series can not be filtered or standardized to achieve second-order stationarity, because the correlation structure of the series depends on the season. For instance, in a river, where high runoff periods occur in the spring and low flows coupled with irrigation diversions occur in the summer, the stream-flow correlations between spring months may be different from the correlations between summer months.

Periodically correlated (PC) processes are a class of processes which are in general nonstationary but exhibit many of the properties of

stationary processes. Gladyshev (1961) defines a periodically correlated random sequence as a process which has periodic structure in the usual wide sense.

More precisely, let $X = \{X_t, t \in Z\}$ be a discrete parameter and zero mean time series and let $E|X_t|^2 < \infty$, $t \in Z$, where Z stands for all integers. The time series $\{X_t\}$ is said to be periodically correlated (PC) with period T, if there is a positive integer T, such that the covariance function $R(t,s) = Cov(X_t, X_s)$ is periodic with period T, i.e.

$$R(t,s) = R(t+T, s+T),$$

for all t, s, and moreover T is the smallest integer for which the above two relations hold. When $T = 1$, a PC series is equivalent to a stationary series.

Example 1 Figure 1 shows 1000 observations from a realization of PCAR(1),

$$X_t = \phi(t)X_{t-1} + Z_t,$$

where

$$\phi(t) = \begin{cases} 0.6 & \text{if } t \text{ odd} \\ 0.2 & \text{if } t \text{ even,} \end{cases}$$

and

$$Z(t) \sim \begin{cases} N(0,3) & \text{if } t \text{ odd} \\ N(0,4) & \text{if } t \text{ even.} \end{cases}$$

Example 2 Figure 2 shows 1000 observations from a realization of PCAR(2),

$$X_t = \phi_1(t)X_{t-1} + \phi_2(t)X_{t-2} + Z_t,$$

where $\phi_1(t)$ is the same as $\phi(t)$ in Example 4.1 and

$$\phi_2(t) = \begin{cases} 0.3 & \text{if } t \text{ odd} \\ 0.7 & \text{if } t \text{ even.} \end{cases}$$

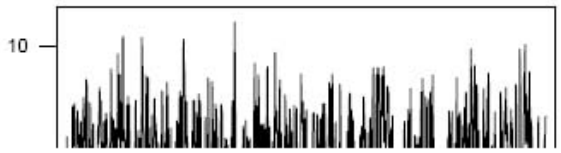


Figure 1: 1000 observations from a realization of PCAR(1)

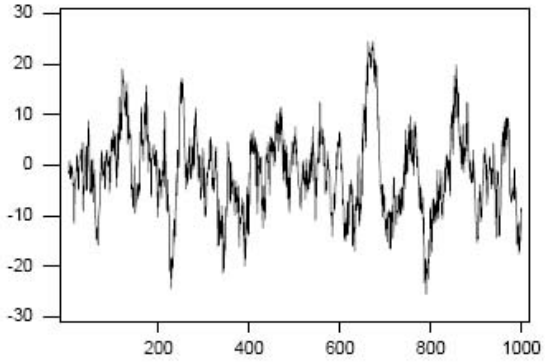


Figure 2: 1000 observations from a realization of PCAR(2)

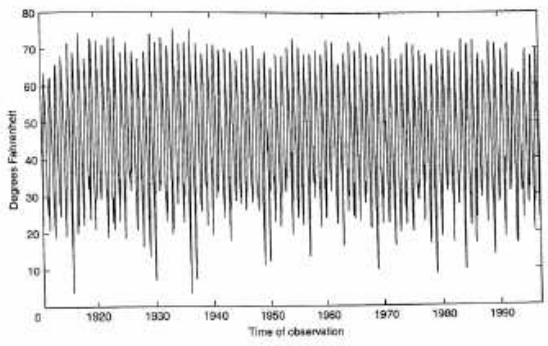


Figure 3: Average monthly temperatures at Big Timber, MT during 1911-1996

Example 3 (Lund and Symour (2001)) Figure 3 shows 86 years of monthly average temperatures at Big Timber, MT during the period 1911-1996. A seasonal cycle clearly appears in the first moment, with winter temperatures being cooler than summer temperatures. A periodic cycle in the variance is also evident: the summer picks are less jagged (from year to year)than their counterparts at winter troughs. Indeed, at most stations in the temperate zone, summer temperatures are less variable than winter temperatures.

Figures 4 and 5 plot the monthly sample mean and sample standard deviations of this time series and confirm these indications graphically.

Physical phenomena that involve periodicities give rise to random data for which appropriate probabilistic models exhibit periodically time-variant parameters. For example: In the mechanical-vibration monitoring and diagnosis for machinery, periodicity arises from rotation, revolution , and reciprocation of gears, belts, chains, shifts, propellers, bearing, pistons, and so on. In atmospheric science (e.g., for weather forecasting), periodicity arises from seasons, caused primarily by rotation and revolution of the earth. In radioastronomy,

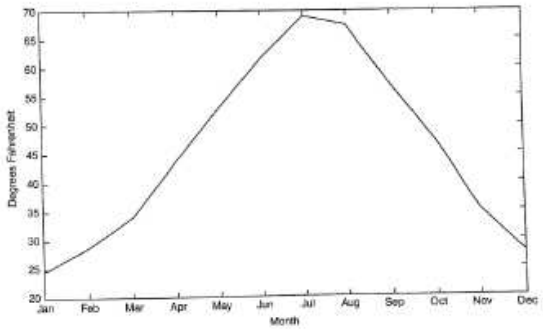


Figure 4: Monthly sample mean for Big Timber, MT

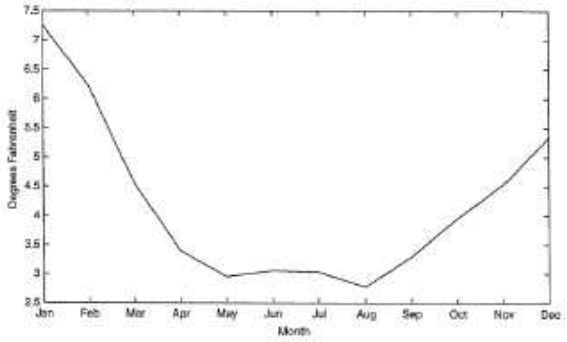


Figure 5: Monthly sample standard deviations for Big Timber, MT

periodicity arises from revolution of the moon, rotation and pulsation of the sun, rotation of Jupiter and revolution of its satellite Io, and so on, and can cause strong periodicities in time series (e.g., pulsar signals). In biology, periodicity in the form of biorhythms arises from both internal and external sources (e.g., circadian rhythms). In communications, telemetry, radar, and sonar, periodicity arises from sampling, scanning, modulating, multiplexing, and coding operations, and it can also be caused by rotating reflectors such as helicopter blades and air- and watercraft propellers.
The notion of PC processes seems to have begun with E.R. Bennett (1958) who observed their presence in a communication theoric context and called them cyclostationary. L.I. Gudzenko (1959) seems to have initiated the subject of nonparametric spectral analysis for PC process. A short time later Gladyshev (1961) published the first analysis of spectral properties and representations based on the connection between PC sequences and stationary vector sequences.

Works of Glaydsev initiated a line of research in the theory of second order processes and since the early 1960's a cosiderable amount of research has been executed in the area of periodic modelling. Breslford (1967) obtained a spectral like representation of mixed summation and integral form for PC sequences. Brelsford (1967) and Jones and Brelsford (1967) presented methods for estimation of the periodic coefficients in periodic autoregression models (now called PAR or PCAR sequences). The basic properties of these models, including the asymptotic theory of autocorrelation based estimators, are established by Pagano (1978). The PAR has been considered also by Troutman (1979). Andel (1989) investigated periodic autoregression with exogenous variables. Mcleod (1993) derived a test for detecting periodic correlation in the residuals of fitted seasonal autoregressive integrated moving-average (ARIMA) models by work in river flow forecasting. Bloomfield, Hurd and Lund (1994) fitted an autoregressive moving-average (ARMA) model with periodically varying coefficients (PARMA) to the stratospheric ozone data. Boshnakov (1994, 1997)

investigated the stochastic difference equation, which is used to define the PAR models and introduced a Markovian dual process for using to study the properties of PAR models. Boswijk and Franses (1996) analyzed the presence and consequences of a unit root in PAR models for univariate quarterly time series. Lund and Basawa(1999, 2000) overviewed general modeling and analysis problems with PC processes and also PAR and PARMA models. Hrafnelsson and Newton (2000), have obtained the simultaneous confidence bands for the spectral density of a vector autoregressive process, by using pagano's results. The books by Gardner (1994) and Hurd and Miamee (2007) are also rich in demonstrating various applications of periodically correlated processes in sciences and engineering.

Because of the periodic nature of the nonstationarity, PC processes have been called periodically nonstationary (Markelov, 1966), cyclostationary (Bennett, 1958; Gardner and Frank, 1975), periodically stationary (Papoulis, 1965), periodic nonstationary (Ogura, 1971) and processes with periodic structure (Brelsford, 1967).

Hurd and Gerr (1991) proposed two tools for determining the presence of periodic correlation, the coherent and the incoherent statistics. Martin (1999) detected the periodic autocorrelation in time series via zero-crossings. Spectral domain techniques in detecting and modeling periodic correlation are given by Broszkiewicz-Suwaj (2003) and Broszkiewicz-Suwaj et al. (2004).

Maximum likelihood estimation for periodic autoregressive moving average models is given by Vecchia, (1985). An algorithm for the exact likelihood of periodic autoregressive moving average models is obtained by Li and Hui (1988). An algorithm for the computation of the theoretical autocovariances of a periodic autoregressive process is shown by Boteva and Boshnakov (1992). Parameter estimation for periodic ARMA models is given by Adams and Goodwin (1995) and for recursive computation of the parameters of periodic autoregressive

moving average processes refer to Boshnakov (1995). For modeling and Inference for Periodically Correlated Time Series refer to Lund and Basawa (1999). Innovations algorithm for periodically stationary time series is giveb by Anderson, Meerschaert and Vecchia (1999). Recursive prediction and likelihood evaluation for periodic ARMA models are obtained by Lund and Basawa (2000). Shao and Lund (2004) considered the computation and characterization of autocorrelations and partial autocorrelations in periodic ARMA models. Covariance generating functions and spectral densities of periodically correlated autoregressive processes are also considered by Shishebor, Nematollahi and Soltani (2006).

A limiting property of sample autocovariances of periodically correlated processes with application to period determination is studied by Tian (1988). Maryan and Lavorsky (2002) considered the estimate of period of Gaussian periodically correlated stochastic process and cyclo-period estimation for discrete-time cyclo-stationary signals is given by Wang, Chen and Huang (2006).

5 The spectra of a periodically correlated time series and their estimation

Let $\{X_t\}$ be a discrete parameter, zero mean and real periodically correlated (PC) time series with period T. Since $\{X_t\}$ is a nonstationary time series, it does not possess a spectral density function in the conventional sense, but Gladyshev (1961) has shown that we may associate with $\{X_t\}$ a Hermitian nonnegative definite $T \times T$ matrix $\mathbf{f}(\omega)$, which we may call the "spectral matrix," defined as follows. The (j,k)th element of $\mathbf{f}(\omega)$, $f_{jk}(\omega)$ is given by

$$f_{jk}(\omega) = \frac{1}{T} f_{k-j}((\omega - 2\pi j)/T),\ j,k = 0,...,T-1,\ 0 \leq \omega \leq 2\pi,$$

where $f_k(\omega)$ satisfies the relation

$$f_k(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} B_k(\tau) \exp(-i\tau\omega),$$

with $B_k(\tau) = B_{k+T}(\tau)$ the kth order coefficient of the Fourier series expansion of the periodic function (with respect to t) $R_t(\tau) := E(X_{t+\tau}X_t)$, i.e.

$$B_k(\tau) = \frac{1}{T} \sum_{t=0}^{T-1} R_t(\tau) \exp(-\frac{2\pi ikt}{T}).$$

When $k < 0$ and $\omega < 0$ or $\omega > 2\pi$ the functions $\{f_k(\omega)\}$ are completely determined by the identities $f_k(\omega) = f_{k+T}(\omega)$ and $f_k(\omega + 2\pi) = f_k(\omega)$, $f_k(0) = 0$ for all k. These $f_k(\omega)$ are in a sense spectra of PC processes.

It is well known that the PC time series X_t must be harmonizable in the sense of Loeve (1963), i.e. can be represented as

$$X_t = \int_0^{2\pi} \exp(it\omega) d\mathcal{Z}(\omega),$$

where $\{\mathcal{Z}(\omega),\ 0 \leq \omega < 2\pi\}$ is a zero mean complex-valued random process (or random measure). The covariance structure of $\{\mathcal{Z}(\omega),\ 0 \leq \omega < 2\pi$ is described by the complex-valued bivariate measure μ, restricted to $[0, 2\pi) \times [0, 2\pi)$ with increments $\mu(d\omega_1, d\omega_2)$ $= E[d\mathcal{Z}(\omega_1)\overline{d\mathcal{Z}(\omega_2)}]$. The spectrum μ is concentrated on the $2T - 1$ parallel diagonal lines $\omega_2 = \omega_1 + 2\pi k/T$, $k = 0, \pm 1, ..., \pm(T - 1)$ restricted to $[0, 2\pi) \times [0, 2\pi)$. In this case, we see $\mathcal{Z}(\omega)$ has PC increments, i.e. $\mu(d\omega_1, d\omega_2) = 0$, unless $\omega_2 = \omega_1 + 2\pi k/T$, $k = 0, \pm 1, ..., \pm(T - 1)$.

When $T = 1$, then $\mathcal{Z}(\omega)$ has orthogonal increments, i.e. $\mu(d\omega_1, d\omega_2) = 0$, unless $\omega_2 = \omega_1$, i.e., μ is concentrated on the main diagonal $\omega_2 = \omega_1$, that is a well known result related to the stationary case.

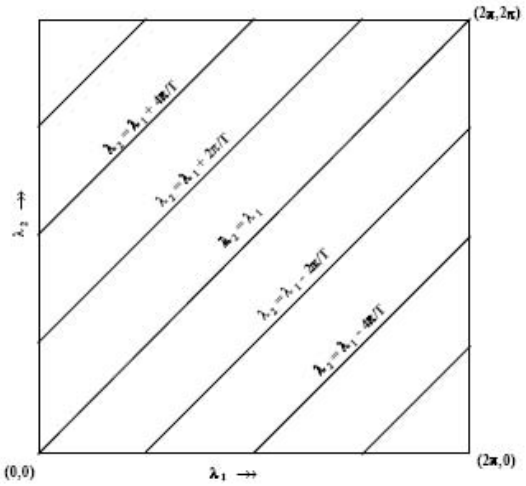


Figure 6: The periodic diagonal nature of the support set of the spectra for $T = 4$

In the general case, the main diagonal component of μ is $\mu(d\omega, d\omega) = E|d\mathcal{Z}(\omega)|^2 = f_0(\omega)d\omega$. The entire mass of the spectra in $f_0(\omega)$, is concentrated on the main diagonal $\omega_2 = \omega_1$, (like the spectrum of a stationary time series), and is a real valued and nonnegative function, and the masses of the other spectra, which can be complex valued, namely f_{k-j}, $j \neq k$, have their masses concentrated on the off diagonals $\omega_1 = \omega_2 + 2\pi k/T$, $k = \pm 1, \pm 2, .., \pm(T-1)$. The periodic diagonal nature of this support set for $T = 4$ is shown in Figure 6.

Following Gladyshev (1961), we may construct an alternative, but equivalent definition of a PC process as follows. We say that the series $\{X_t\}$ is periodically correlated with period T if and only if the T-dimensional vector series $\mathbf{Y}_t = (X_{tT}, X_{tT+1}, ..., X_{tT+T-1})'$ is stationary in the wide sense. It can then be shown that (Gladyshev,

1961) that

$$\mathbf{f}(\omega) = \frac{1}{T}\mathbf{U}(\omega)\mathbf{h}(\omega)\mathbf{U}^{-1}(\omega),$$

where $\mathbf{f}(\omega)$ is the spectral density matrix of $\{X_t\}$, $\mathbf{h}(\omega)$ the spectral density matrix of the stationary vector series $\{(X_{tT}, X_{tT+1}, ..., X_{tT+T-1})'\}$ and $\mathbf{U}(\omega)$ is a unitary matrix (i.e., $\mathbf{U}(\omega)\mathbf{U}^*(\omega) = \mathbf{I}$) with elements $U_{jk}(\omega) = T^{-1/2}\exp(\frac{2\pi ijk - ik\omega}{T})$.

The above relation suggests that the estimation of $\mathbf{f}(\omega)$ can be accomplished through the estimation of $\mathbf{h}(\omega)$. One can estimate $\mathbf{h}(\omega)$ using either parametric estimation (say via linear vector ARMA models) or using the kernel approach (Brillinger, 1975, Chapter 7 and Hannan, 1970). Here our object is to exploit the special structure of the block-Toeplitz matrix associated with multidimensional series $\{\mathbf{Y}_t\}$ to derive spectral estimates, using the techniques introduced in Section 3. This also enable us to derive a high resolution estimate of $\mathbf{h}(\omega)$ and hence of the spectral matrix, $\mathbf{f}(\omega)$ of PC time series.

Let $\{\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n\}$ be a sample of size n from T-dimensional stationary vector series $\{\mathbf{Y}_t\}$. We assume that $E\mathbf{Y}_t = 0$. Let $n = Mm$, where M and m are integers. Divide the data into M groups, where each group consists of m observations, and let the observations in the l-th group ($l = 1, ..., M$) be denoted by the $mT \times 1$ vector $\widetilde{\mathbf{Y}}_l$, where

$$\widetilde{\mathbf{Y}}_l = (\mathbf{Y}'_{lm}, \mathbf{Y}'_{lm-1}, ..., \mathbf{Y}'_{(l-1)m+1})' \ , \ l = 1, ..., M.$$

We estimate the $mT \times mT$ block-Toeplitz covariance matrix $\mathbf{\Gamma}_m$ of order m by

$$\widehat{\mathbf{\Gamma}}_m = \frac{1}{M}\sum_{j=1}^{M} \widetilde{\mathbf{Y}}_j \widetilde{\mathbf{Y}}'_j.$$

Let $\widehat{\mathbf{\Lambda}}_m(\omega_j)$, $j = 0, ..., m-1$ be the eigenvalue-matrices of $\widehat{\mathbf{\Gamma}}_m$ and assume that m is odd. According to our findings in Section 3, we

consider

$$\widehat{\mathbf{h}}_m(\omega_l) = \frac{1}{4\pi}\sum_{j=0}^{m-1}\widehat{\mathbf{A}}_m(\omega_j,\omega_l),$$

$$\widehat{\mathbf{h}}_{m,cap}(\omega_l) = \frac{1}{\pi}\left\{\sum_{j=0}^{m-1}\widetilde{\mathbf{A}_n}(\omega_j,\omega_l)\right\}^{-1}.$$

as estimators of $\mathbf{h}_m(\omega_l)$ and $\mathbf{h}_{m,cap}(\omega_l)$, respectively, where

$$\widehat{\mathbf{A}}_m(\omega_j,\omega_l) = \frac{2}{m}\sum_{t=1}^{m}\sum_{s=1}^{m}\mathbf{W}_m^{t*}(\omega_j)\widehat{\mathbf{\Lambda}}_m(\omega_j)\mathbf{W}_m^{s}(\omega_j)e^{-i(t-s)\omega_l}$$

and

$$\widetilde{\mathbf{A}_m}(\omega_j,\omega_l) = \frac{2}{m}\sum_{t=1}^{m}\sum_{s=1}^{m}\mathbf{W}_m^{t*}(\omega_j)\widehat{\mathbf{\Lambda}}_m^{-1}(\omega_j)\mathbf{W}_m^{s}(\omega_j)e^{-i(t-s)\omega_l}.$$

Using the relation between the spectral density matrix of a periodically correlated time series and a stationary vector time series, we propose the following two estimators, as the estimates for the spectral density matrix of PC time series, $\mathbf{f}(\omega_l)$, namely,

$$\widehat{\mathbf{f}}_m(\omega_l) = \frac{1}{T}\mathbf{U}(\omega_l)\widehat{\mathbf{h}}_m(\omega_l)\mathbf{U}^{-1}(\omega_l),$$

$$\widehat{\mathbf{f}}_{m,cap}(\omega_l) = \frac{1}{T}\mathbf{U}(\omega_l)\widehat{\mathbf{h}}_{m,cap}(\omega_l)\mathbf{U}^{-1}(\omega_l).$$

6 Numerical illustration

As an application of these methods for estimating the spectral density matrix of a real PC time series, we consider time series of mean monthly flows of Fraser River at Hope, BC, from January 1913 to December 1990. This time series has been used by several authors for

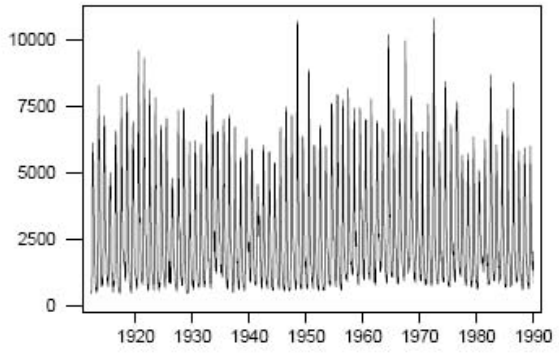


Figure 7: Mean monthly flows of Fraser River at Hope, BC, from January 1913 to December 1990

fitting PC time series models, and is known to have periodicity of 12 months, $T = 12$, (Vecchia and Ballerini (1991), Mcleod, (1994) and see Hipel and Mcleod (1994)). The data is plotted in Figure 7.

There is a twelve-dimensional stationary time series $\mathbf{Y}_t$, associated with this time series. The number of observations are 936 and so we have 78 observations from $\mathbf{Y}_t$, that we consider in $M = 26$ groups, with $m = 3$ elements in each group. The estimates $\widehat{\mathbf{h}}_m(\omega_l)$ and $\widehat{\mathbf{h}}_{m,cap}(\omega_l)$ are calculated with $\omega_j = \frac{2\pi j}{m}$, $j = 0, ..., m-1$. The above estimates are computed at the frequencies $\omega_l = l\pi$, $l = 0(0.1)1$. And then we computed $\widehat{\mathbf{f}}_m(\omega_l)$ and $\widehat{\mathbf{f}}_{m,cap}(\omega_l)$, the truncated estimate and Capon estimate of spectral density matrix of periodically correlated X_t. The estimates of $\widehat{\mathbf{f}}_m(\omega_l)$ and $\widehat{\mathbf{f}}_{m,cap}(\omega_l)$ are both 12×12 matrices and for reasons of space, we don't include these, but we plot the logarithm of $\widehat{f}_0(\omega)$ in Figure 8.

According to a criteria introduced by Nematollahi and Subba Rao (2005), the truncated estimate seems to be preferable to Capon's estimate.

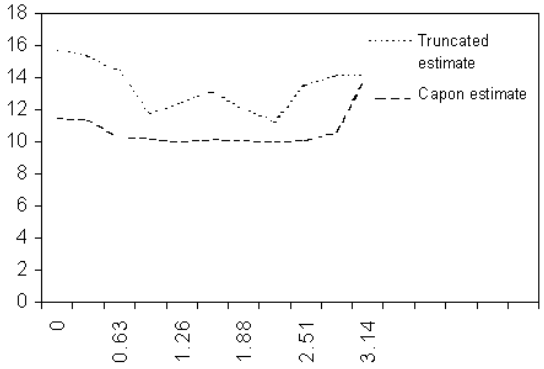


Figure 8: The logarithm of theoretical density, truncated estimate and Capon estimate of $f_0(\omega_l)$, $\omega_l = l\pi$, $l = 0(0.1)1$

References

[1] Adams, G. J. and Goodwin, G. C. (1995) Parameter estimation for periodic ARMA models, Journal of Time Series Analysis 16 (2) 127-145.

[2] Andel, J. (1989) Periodic autoregression with exogenious variables and periodic variance. *Aplicace matematiki* , **34(5)**, 387-395.

[3] Anderson, P., Meerschaert, M. and Vecchia, A., (1999) Innovations algorithm for periodically stationary time series. Stochastic Process. Appl. v83. 149-169.

[4] Bennett, W. R. (1958) Statistics of regenerative digital transmission. *Bell Syst. Tech. J.*, **37**, 1501-42.

[5] Bloomfield, P., Hurd, H. L. and Lund, R. (1994) Periodic correlation in stratospheric ozone time series. *J. Time Series Anal.,* **15(2)**, 127-150.

[6] Boshnakov, G. (1994) Periodically correlated sequences : Some properties and recursions. Research Report 1, division of Quality Technology and Statistics, Luleo University, Sweden.

[7] Boshnakov, G. N. (1996) Recursive computation of the parameters of periodic autoregressive moving-average processes. *J. Time Series Anal.,* **17(4)**, 333-349.

[8] Boshnakov, G. (1997) Periodically correlated solution of a class of stochastic difference equations. *Progress in System and Control Theory*, **23**, 1-9.

[9] Boswijk, H. P. and Franses, P. H. (1996) Unit roots in periodic autoregressions. *J. Time series Anal.*, **17(3)**, 221-245.

[10] Boteva, A. and Boshnakov, G. N. (1992)An algorithm for the computation of the theoretical autocovariances of a periodic autoregression process. Proceedings of the 8-th Seminar on Statistical Data Analysis, Varna, pages 9-15.

[11] Brelsford, W. M. (1967) Probability prediction and time series with periodic structure. Phd Dissertation, Johns Hopkins University, Baltimore maryland.

[12] Brillinger, D. R. (1975) *Time series, Data Analysis and Theory.* Holt, Rinehart and winston, New York.

[13] Broszkiewicz-Suwaj, E. (2003) Methods for determining the presence of periodic correlation based on the bootstrap methodology, Tech. Rep. HSC/03/2, Wroclaw University of Technology, Wroclaw, Poland.

[14] Broszkiewicz-Suwaj, E, Makagon, A., Weron, R. and Wylomanska, A. (2004) On detecting and modeling periodic correlation in financial data, Physica A 336, 196-205.

[15] Capon, J. (1969) High resolution frequency-wave number spectrum analysis. *Proc. IEEE* **57**, 1408-18.

[16] Gladyshev, E. G. (1961) Periodically correlated random sequences. *Soviet Math. Dokl.,* **2**, 385-388.

[17] Gardner, W. A. (1994) *Cyclostationary in Communications and Signal Processing*. IEEE Press, New York.

[18] Gardner, W. A. and Franks, L. E. (1975) Characterization of cyclostationary random signal processes. *IEEE Trans. on Information Theory*, **IT-21(1)**, 4-14.

[19] Gray, R. M. (1972) On the asymptotic eigenvalue distribution of Toeplitz matrices. *IEEE Trans. Inf. Theory,* **IT-18(6)**, 725-30.

[20] Gudzenko, L. I. (1959) On periodically nonstationary processes. *Radiotekhnika i elektronika*, **4(6)**, 1062-1064.

[21] Hannan, E. J. (1970) *Multiple Time Series*. New York, Wiley.

[22] Hipel, K. W. and Mcleod, A. I. (1994) *Developments in water Science, 45, Time Series Modelling of Water Resources and Environmental Systems.* Elsevier Science B.V., Netherlands.

[23] Hurd, H.L. and Gerr, N.L. (1991). Graphical methods for determining the presence of periodic correlation. J. Time Ser. Anal. 12, 337-350.

[24] Hurd, H. L. and Miamee, A. G. (2007) *Periodically correlated random sequences: Spectral theory and practice.* John wiley.

[25] Hrafnkelsson, B and Newton, H. J. (2000) asymptotic simultaneous confidence bands for vector autoregressive spectra *Biometrika*, **87(1)**, 173-182.

[26] Jones, R. H. and Brelsford, W. M. (1967) Time series with periodic structure. *Biometrika*, **54**, 403-408.

[27] Li, W.K. and Hui, Y.V. (1988). An algorithm for the exact likelihood of periodic autoregressive moving average models. Commun. Statist. Simulation Comput. 14, 1483-1494.

[28] Loeve, M. (1963) *Probability Theory.* Van Nostrand, New York.

[29] Lund, R. B. and Basawa, I. V. (1999). Modeling and iference for periodically correlated Time series. *Asymptotic, Nonparametrics, And Time Series,* Ed. S. Ghosh, **158,** STATISTICS: textbook and monsgraphs.

[30] Lund, R. B. and Basawa, I. V. (2000). Recursive prediction and likelihood evaluation for periodic ARMA models. *J. Time Series Anal.*, bf 21(1), 75-93.

[31] Lund, R. B. and Symour, L. (2001) Time series, periodic, Encyclopedia of environmetrics, Vol. 4, edt. El-Shaarawi, A. H. and Piegorsch, W. W., John Wiley.

[32] Markelov, V. A. (1966) Axis crossings and relative time of existence of periodically nonstationary random processes. *Soviet Radiophysics*, **9(4)**, 440-443.

[33] Martin, D. E. K. (1999) Detection of periodic autocorrelation in time series data via zero-crossings, J. of time series analysis, 20(4), 435-452.

[34] Maryan, K. and Lavorsky, I. (2002) The estimate of period of Gaussian periodically correlated stochastic process, in Modern Problems of Radio Engineering, Telecommunications and Computer Science, Proceedings of the International Conference, 101-102.

[35] Mcleod, A. I. (1993) Parsimony, model adequacy and periodic correlation in time series forecasting. *International Statistical Review*, **61(3)**, 387-393.

[36] Mcleod, A. I. (1994) Diagnostic checking of periodic autoregression models with application, *J. Time Series Anal.,* **15(2)**, 221-233.

[37] Nematollahi, A. R. and Shishebor, Z. (2005) A note on the equivalence of covariance matrix of a multivariate stationary time series with the related circular symmetric matrix, Journal of Statistical Theory and Applications, Vol. **4(2)**, 209-217.

[38] Nematollahi, A. R. and Subba Rao, T. (2005) On the spectral estimation of periodically correlated (cyclostationary) time series, *Sankhyā A*. **67(3)**, 568-589.

[39] Ogura, H. (1971). Spectral representation of periodic nonstationary random processes. *IEEE Trans. on Information Theory*, **IT-17**, 143-149.

[40] Pagano, M. (1978) On periodic and multiple autoregressions. *Ann. Statist.*, **6**, 1310-1317.

[41] Papoulis, (1965). *Probability, Random Variables and Stochastic Processes*, New York.

[42] Pisarenko, V. F. (1972) On the estimation of spectra by means of nonlinear functions of the covariance matrix, Geophys., J. R. Astr. Soc., **28**, 511-531.

[43] Priestley, M. B. (1989) *Spectral Analysis and Time Series*. London, Academic Press

[44] Shao, Q. and Lund, R.B., (2004) Computation and characterization of autocorrelations and partial autocorrelations in periodic ARMA models. J. Time Ser. Anal. 25(3), 359-372.

[45] Shishebor, Z., Nematollahi, A.R. and Soltani, A.R. (2006) On the covariance functions of periodically correlated AR processes, Journal of Applied Mathematics and Stochastic Analysis, 1-17

[46] Subba Rao, T. and Gabr, M. M. (1989) The estimation of spectrum, inverse spectrum and inverse autocovariances of a stationary time series. *J. Time Series Anal.,* **10(2)**, 183-202.

[47] Tian, C. J. (1988) A limiting property of sample autocovariances of periodically correlated processes with application to period determination, J. Time Ser. Anal., vol. 9, no. 4, pp. 411417.

[48] Troutman, B. (1979) Some results on periodic autoregression. *Biometrika*, bf 66(2), 219-228.

[49] Vecchia, A.V. (1985). Maximum likelihood estimation for periodic autoregressive-moving average models. Technometrics 27, 375-384.

[50] Vecchia, A. V. and Ballerini, R. (1991) Testing for periodic autocorrelations in seasonal time series data. *Biometrika,* **78,** 53-63.

[51] Wang, J., Chen, T. and Huang, B. (2006) Cyclo-period estimation for discrete-time cyclo-stationary signals, IEEE Transactions on Signal Processing 54(1), 83-94.

Variational approximations for Logistic Mixed Models

J.T. Ormerod, and M.P. Wand

School of Mathematics and Applied Statistics, University of Wollongong,
Northfields Avenue, Wollongong 2522, AUSTRALIA

Abstract: Variational approximations have been used extensively in Statistical Physics and Computer Science as a means of simplifying probability calculus. We investigate the transferral and adaptation of variational approximation methodology for the logistic mixed model setting, which is hindered by intractable multivariate integrals.Comparisons with accurate approximations based on Markov Chain Monte Carlo sampling shows that our methodology is comparable, while being significantly faster.

Keywords: Logistic Mixed Models; Variational Approximations; Penalized Splines.

1 Introduction

Logistic mixed models are an extremely useful class of models for binary data. They are a fundamental in longitudinal data analysis, a common analysis in biomedical statistics, where they can be used to model correlation in grouped data and include simple, hierarchical, crossed and nested random effect models (McCulloch & Searle, 2001; Zhao, Staudenmayer, Coull & Wand, 2006). They also can be used for function estimation including scatterplot smoothing, random coefficient and kriging models (Ruppert, Wand & Carrol, 2003; Zhao, Staudenmayer, Coull & Wand, 2006).

Unfortunately the analysis of logistic mixed models is hindered by the presence of analytically intractable integrals and approximations must be made. Approximations include Laplace-like approximations such as Penalized Quasi-Likelihood (PQL, Breslow & Clayton, 1993), Gauss-Hermite quadrature (Naylor & Smith, 1982) and

Monte Carlo methods (Robert & Casella, 1999; Gilks, Richardson & Spiegelhalter, 1996; McCullagh, 1997). Each of these methods have computational shortcomings associated with them. PQL approximations can be severely biased (Breslow & Lin, 1995; Lin & Breslow, 1996), Gauss-Hermite quadrature does not scale well to high dimensional integrals and Monte Carlo methods suffer from the problems of the slowness and difficulties accessing convergence (although some progress has been made, see Rosenthal, 1995 and Cowles & Rosenthal, 1998 for instance). Excellent summaries of existing approximations for generalized linear mixed models (GLMM), of which logistic mixed models are a special case, may be found in McCulloch & Searle (2001, Chapter 10) and Tuerlinckx, Rijmen, Verbeke & de Boeck (2006). However these overviews do not include variational approximations.

Variational methods are a class of analytic approximations which offer a fresh approach to many statistical problems. These are a class of analytic approximations with origins in physics which have recently been applied by computer scientists to a variety of statistical models (MacKay, 1995; Attias, 2000; Beal & Ghahramani, 2002; Beal, 2003; Bishop & Winn, 2003; Winn & Bishop, 2005; Consonni & Marin, 2007). In this article we describe a variational approach, similar to Rijmen & Vomlel (2007), for approximation of integrals arising in logistic mixed models and Bayesian versions thereof based on the ideas of Jaakkola & Jordan (2000). We also develop *grid-based* variational approximations for calculating marginal poster densities. We illustrate these methods with a case study.

2 Logistic Mixed Models

Consider the logistic mixed model with normally distributed random effects given by

$$\begin{array}{rl} [\mathbf{y}|\boldsymbol{\beta},\mathbf{u}] & = \exp\left\{\mathbf{y}^T(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u}) - \mathbf{1}^T\log(\mathbf{1}+e^{\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u}})\right\} \\ [\mathbf{u}|\boldsymbol{\sigma}^2] & = |2\pi\mathbf{G}_{\boldsymbol{\sigma}^2}|^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{u}^T\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\mathbf{u}\right\} \end{array} \qquad (1)$$

where $\mathbf{y} \equiv (y_1, \ldots, y_n)$ is a response vector, $\mathbf{X}$ and $\mathbf{Z}$ are $n \times p$ and

$n \times q$ matrices, $\boldsymbol{\beta}$ is a vector corresponding to the fixed effects of length p, $\mathbf{u}$ are random effects of length q with covariance matrix $\mathbf{G}_{\boldsymbol{\sigma}^2}$ which is parameterized by variance components $\boldsymbol{\sigma}^2 \equiv (\sigma_1^2, \ldots, \sigma_v^2)$.

This model arises in a number of common applications including scatterplot smoothing, additive models and longitudinal models (Ruppert *et al.*, 2003; Zhao *et al.*, 2006) which specify the particular values for the matrices $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{G}_{\boldsymbol{\sigma}^2}$.

The likelihood for $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}^2$ is obtained by integrating out the random effects and is given by

$$\ell(\boldsymbol{\beta}, \boldsymbol{\sigma}^2) = \int_{\mathbb{R}^q} \exp\left\{\mathbf{y}^T(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) - \mathbf{1}^T \log(\mathbf{1} + e^{\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u}})\right\} \frac{\exp\left(-\frac{1}{2}\mathbf{u}^T\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\mathbf{u}\right)}{|2\pi\mathbf{G}_{\boldsymbol{\sigma}^2}|^{q/2}} d\mathbf{u}. \quad (2)$$

Bayesian logistic mixed models are also of considerable interest. A common Bayesian approach is to place vague priors on the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}^2$. A common choice is

$$\boldsymbol{\beta} \sim N(\mathbf{0}, \sigma_\beta^2\mathbf{I}) \quad \text{and} \quad \sigma_i^2 \sim IG(\alpha_i, \alpha_i) \quad (3)$$

where σ_β^2 is chosen to be a large positive constant, α_i is chosen to be a small positive constant and $IG(\alpha_i, \alpha_i)$ denotes the inverse-gamma distribution parameterized such that $\mathbb{E}(\sigma_i^2) = 1$. Gelman (2006) advocates the use of other priors, particularly for the σ_i^2s, but for simplicity we will focus on these.

Often the marginal posterior distributions for the β_is are of interest. These require calculation of integrals of the form

$$[\beta_i|\mathbf{y}] = \frac{\int [\mathbf{y}|\boldsymbol{\beta}, \mathbf{u}][\boldsymbol{\beta}][\mathbf{u}|\boldsymbol{\sigma}^2][\boldsymbol{\sigma}^2] d\boldsymbol{\beta}_{-i} d\mathbf{u} d\boldsymbol{\sigma}^2}{\int [\mathbf{y}|\boldsymbol{\beta}, \mathbf{u}][\boldsymbol{\beta}][\mathbf{u}|\boldsymbol{\sigma}^2][\boldsymbol{\sigma}^2] d\boldsymbol{\beta} d\mathbf{u} d\boldsymbol{\sigma}^2} \quad (4)$$

where $\boldsymbol{\beta}_{-i}$ is the vector $\boldsymbol{\beta}$ with the ith element removed.

Unfortunately there is no known closed form for (2) or (4) and so we must pursue approximations. We will now pursue variational approximations to these integrals.

3 Variational Approximations

There are two standard ways of finding such parameterized bounds. One method of deriving parameterized lower bounds exploits convexity properties of the integrand, we call *tangent transforms* (see Jaakkola & Jordan, 2000), while the other uses Jensen's inequality and leads to EM like algorithms (Hinton & van Camp 1993; MacKay, 1995; Attias, 2000; Beal & Ghahramani, 2002; Beal, 2003; Bishop & Winn, 2003; Winn & Bishop, 2005). For simplicity we will only focus on the first of these.

Tangent transforms exploit the fact that for any convex function $f : \mathbb{R}^d \to \mathbb{R}$ that

$$f(\mathbf{x}) \geq f(\boldsymbol{\xi}) + \mathsf{D}_{\mathbf{x}} f(\boldsymbol{\xi})(\mathbf{x} - \boldsymbol{\xi}) \quad \text{for all } \mathbf{x}, \boldsymbol{\xi} \in \mathbb{R}^d \tag{5}$$

and also

$$f(\mathbf{x}) = \max_{\boldsymbol{\xi}} f(\boldsymbol{\xi}) + \mathsf{D}_{\mathbf{x}} f(\boldsymbol{\xi})(\mathbf{x} - \boldsymbol{\xi}). \tag{6}$$

where $\mathsf{D}_{\mathbf{x}} f(\boldsymbol{\xi}) = (\partial f(\mathbf{x})/\partial x_i|_{\mathbf{x}=\boldsymbol{\xi}})_{1 \leq i \leq d}$ (e.g. Rockafellar, 1972). Similarly if f is concave we replace $\geq$ with $\leq$ in (5) and "max" with "min" in (6). While the approximation appears to be simplistic, since it only relies on first derivative information, often this type of approximation is adequate.

For logistic mixed models the difficult in calculating the likelihood (2) stems from the non-quadratic term $b(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) = \log(\mathbf{1} + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}))$. Jaakkola & Jordan (1997) noticed that an upper bound for $b(x)$ can be obtained due to the fact that $g(x) = \log(e^{-x/2} + e^{x/2}) = b(x) - x/2$ is a concave function of x^2 and hence

$$b(x) \leq b_U(x, \xi) = \frac{x}{2} + \log(e^{-\xi/2} + e^{\xi/2}) + \frac{\tanh(\xi/2)}{4\xi}(x^2 - \xi^2) \tag{7}$$

which holds for all $x, \xi \in \mathbb{R}$. This bound is illustrated in Figure 1.

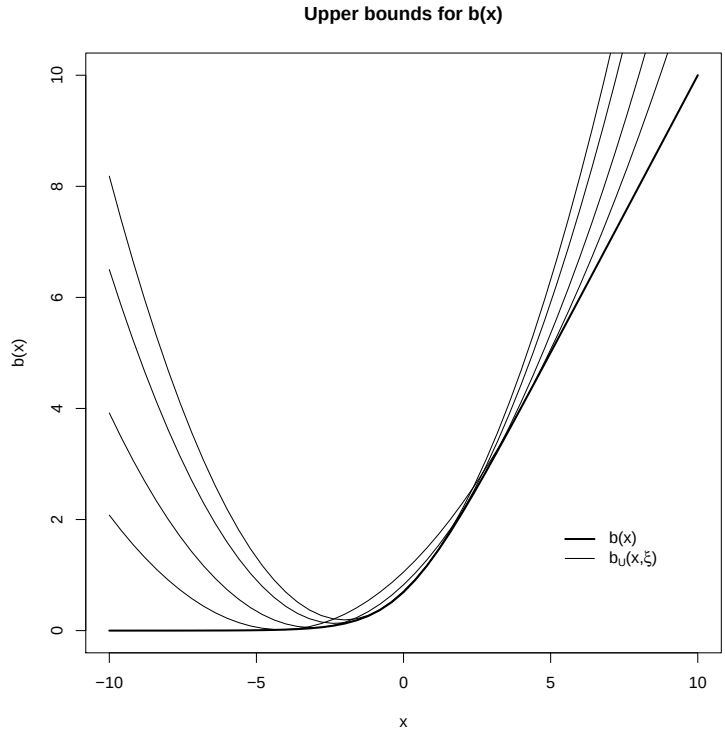


Figure 1: *Upper bounds for* $b(x)$.

Now we can use (7) to find a lower bound for the joint log-likelihood of $\mathbf{y}$ and $\mathbf{u}$, ignoring additive constants,

$$\log[\mathbf{y}, \mathbf{u}] \geq \left(\mathbf{y} - \tfrac{\mathbf{1}}{2}\right)^T (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) - \tfrac{1}{2} (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})^T \boldsymbol{\Lambda} (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) - \tfrac{1}{2}\mathbf{u}^T \mathbf{G}_{\boldsymbol{\sigma}^2}^{-1} \mathbf{u} \quad (8)$$

where $\mathbf{\Lambda} = \text{diag}\left\{\tanh(\xi_i/2)\xi_i/2\right\}_{1\leq i\leq n}$ and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ are additional parameters, called *variational parameters*. Noting that since the right hand side of (8) is, ignoring additive constants, the log of a multivariate Gaussian function of $\mathbf{u}$ we might approximate the posterior distribution $\mathbf{u}|\mathbf{y}$ by $[\mathbf{u}|\mathbf{y}] \approx \delta(\mathbf{u}) = \phi_{\mathbf{\Sigma}}(\mathbf{u}-\boldsymbol{\mu})$ where $\phi_{\mathbf{\Sigma}}(\mathbf{u}-\boldsymbol{\mu}) = |2\pi\mathbf{\Sigma}|^{-1/2}\exp(-(\mathbf{u}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{u}-\boldsymbol{\mu})/2)$,

$$\boldsymbol{\mu} = \mathbf{\Sigma}\mathbf{Z}^T\left(\mathbf{y} - \tfrac{\mathbf{1}}{2}\right) \quad \text{and} \quad \mathbf{\Sigma} = \left(\mathbf{Z}^T\mathbf{\Lambda}\mathbf{Z} + \mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\right)^{-1}. \tag{9}$$

We can now use this as the basis for an approximate EM algorithm.

Let $\boldsymbol{\mu}_{\text{old}}$ and $\mathbf{\Sigma}_{\text{old}}$ be the values for $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ evaluated at the current values for $\boldsymbol{\beta}$, $\boldsymbol{\sigma}^2$ and $\boldsymbol{\xi}$. The expectation step of the EM approach uses, ignoring additive constants,

$$\begin{aligned} Q(\boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\xi}|\boldsymbol{\mu}_{\text{old}}, \mathbf{\Sigma}_{\text{old}}) &= \mathbb{E}_\delta\left\{\log[\mathbf{y}, \mathbf{u}]\right\} \\ &\geq \log(\mathbf{1}^T g(\boldsymbol{\xi})) - \tfrac{1}{2}\log|\mathbf{G}_{\boldsymbol{\sigma}^2}| + \left(\mathbf{y} - \tfrac{\mathbf{1}}{2}\right)^T(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\mu}_{\text{old}}) \\ &\quad -\tfrac{1}{2}\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\mu}_{\text{old}}\right)^T\mathbf{\Lambda}\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\mu}_{\text{old}}\right) - \tfrac{1}{2}\boldsymbol{\mu}_{\text{old}}^T\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\boldsymbol{\mu}_{\text{old}} \\ &\quad -\tfrac{1}{2}\text{tr}\left(\mathbf{\Sigma}_{\text{old}}(\mathbf{Z}^T\mathbf{\Lambda}\mathbf{Z} + \mathbf{G}_{\boldsymbol{\sigma}^2}^{-1})\right). \end{aligned} \tag{10}$$

where $\mathbb{E}_\delta$ denotes expectations with respect to the approximate posterior distribution δ.

The maximisation step then maximizes Q with respect to $\boldsymbol{\beta}$, $\boldsymbol{\sigma}^2$ and $\boldsymbol{\xi}$. In order to optimize Q with respect to $\boldsymbol{\sigma}^2$ we need to specify a particular structure for $\mathbf{G}_{\boldsymbol{\sigma}^2}$. For simplicity we will use $\mathbf{G}_{\boldsymbol{\sigma}^2} = \text{blockdiag}_{1\leq i\leq v}(\sigma_i^2\mathbf{\Omega}_i)$. Other covariance structures for $\mathbf{u}$ are desirable in some contexts (see for example Verbeke & Molenberghs, 2000; McCulloch & Searle, 2001; Zhao *et al.*, 2006), however the covariance structure we have chosen here is quite general.

Differentiating Q with respect to $\boldsymbol{\beta}$, $\boldsymbol{\sigma}^2$ and $\boldsymbol{\xi}$, we find that first order optimality conditions require

$$\begin{aligned} \boldsymbol{\beta} &:= (\mathbf{X}^T\mathbf{\Lambda}\mathbf{X})^{-1}\mathbf{X}\left(\mathbf{y} - \tfrac{\mathbf{1}}{2} - \mathbf{\Lambda}\mathbf{Z}\boldsymbol{\mu}_{\text{old}}\right) \\ \xi_i &:= \sqrt{(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\mu}_{\text{old}})_i^2 + \tfrac{1}{2}(\mathbf{Z}\mathbf{\Sigma}_{\text{old}}\mathbf{Z}^T)_{ii}},\ 1 \leq i \leq n \\ \sigma_i^2 &:= \left(\boldsymbol{\mu}_{\text{old}}^T\mathbf{D}_i\boldsymbol{\mu}_{\text{old}} + \text{tr}\left(\mathbf{\Sigma}_{\text{old}}\mathbf{D}_i\right)\right)/q_i,\ 1 \leq i \leq v \end{aligned} \tag{11}$$

where $\mathbf{D}_i = \text{blockdiag}_{1\le j\le v}(\boldsymbol{\Omega}_j \mathbb{I}_{j=i})$, the q_is are the sizes of the square matrices $\boldsymbol{\Omega}_i$ and $\mathbb{I}_{\{x\}}$ denotes the indicator variable which takes the value 1 if x is true and 0 otherwise. We use (11) as the update equations for the EM algorithm and (9) to update the approximation of the posterior distribution $\mathbf{u}|\mathbf{y}$. Similar updates we first developed by Jaakkola & Jordan (1997) in the context of Bayesian logistic linear models and later by Rijmen & Vomlel (2007) for logistic random effects models. Our approach is an extension of these approaches to a more general model.

A similar approach may be used for Bayesian logistic mixed models. Suppose that we apply the priors (3) to our logistic mixed model. Integrating out the variance components $\boldsymbol{\sigma}^2$ we obtain, ignoring additive constants,

$$\begin{array}{l}\log[\mathbf{y},\boldsymbol{\nu}] = \mathbf{y}^T\mathbf{C}\boldsymbol{\nu} - \mathbf{1}^T b(\mathbf{C}\boldsymbol{\nu}) - \\ \frac{\|\boldsymbol{\beta}\|^2}{2\sigma_{\boldsymbol{\beta}}^2} - \sum_{i=1}^v (\alpha_i + q_i/2)\log\left(\alpha_i + \frac{\mathbf{u}^T\mathbf{D}_i\mathbf{u}}{2}\right)\end{array} \tag{12}$$

where $\mathbf{C} \equiv [\mathbf{X}, \mathbf{Z}]$ and $\boldsymbol{\nu} = (\boldsymbol{\beta}, \mathbf{u})$.

Similar to the logistic mixed model case $[\mathbf{y}, \boldsymbol{\nu}]$ is not multivariate Gaussian in shape. However, noting that $-\log(x) \ge -\xi x + 1 + \log(\xi)$, see for example Jordan, Ghahramani, Jaakkola & Saul (1999), and again using (7) we obtain the following lower bound for (12), again ignoring additive constants,

$$\log[\mathbf{y},\boldsymbol{\nu}] \ge h(\widetilde{\boldsymbol{\xi}}) + \left(\mathbf{y} - \tfrac{1}{2}\right)^T \mathbf{C}\boldsymbol{\nu} - \tfrac{1}{2}\boldsymbol{\nu}^T\left(\mathbf{C}^T\boldsymbol{\Lambda}\mathbf{C} + \mathbf{B}\right)\boldsymbol{\nu} \tag{13}$$

where $h(\widetilde{\boldsymbol{\xi}}) = \mathbf{1}^T g(\boldsymbol{\xi}) + \sum_{i=1}^v (\alpha_i + q_i/2)(\log(\xi_{n+i}) - \alpha_i\xi_{n+i})$,

$$\mathbf{B} = \begin{bmatrix} \sigma_{\boldsymbol{\beta}}^{-2}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sum_{i=1}^v (\alpha_i + q_i/2)\xi_{n+i}\mathbf{D}_i \end{bmatrix} \tag{14}$$

and $\widetilde{\boldsymbol{\xi}} = (\boldsymbol{\xi}, \xi_{n+1}, \ldots, \xi_{n+v})$ are additional variational parameters. Now the right hand side of (13) is, up to additive constants, the log of a

multivariate Gaussian distribution in $\boldsymbol{\nu}$. We use this to approximate the posterior $\boldsymbol{\nu}|\mathbf{y}$ by $\delta(\boldsymbol{\nu}) = \phi_{\boldsymbol{\Sigma}}(\boldsymbol{\nu} - \boldsymbol{\mu})$ where

$$\boldsymbol{\Sigma} = \left(\mathbf{C}^T\boldsymbol{\Lambda}\mathbf{C} + \mathbf{B}\right)^{-1} \quad \text{and} \quad \boldsymbol{\mu} = \boldsymbol{\Sigma}\mathbf{C}^T\left(\mathbf{y} - \tfrac{\mathbf{1}}{2}\right) \tag{15}$$

noting that $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ depend on $\widetilde{\boldsymbol{\xi}}$ through $\boldsymbol{\Lambda}$ and $\mathbf{B}$.

We select $\widetilde{\boldsymbol{\xi}}$ by maximising a lower bound for the marginal likelihood. As before this is most elegantly done via the EM algorithm. Let $\boldsymbol{\mu}_{\text{old}}$ and $\boldsymbol{\Sigma}_{\text{old}}$ be the values for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ evaluated at the current values for $\widetilde{\boldsymbol{\xi}}$. Then the Q function is given by

$$\begin{aligned} Q(\widetilde{\boldsymbol{\xi}}|\boldsymbol{\mu}_{\text{old}}, \boldsymbol{\Sigma}_{\text{old}}) &= \mathbb{E}_\delta\left\{\log\left[\mathbf{y}, \boldsymbol{\nu}\right]\right\} \\ &\leq h(\widetilde{\boldsymbol{\xi}}) + \left(\mathbf{y} - \tfrac{\mathbf{1}}{2}\right)^T \mathbf{C}\boldsymbol{\mu}_{\text{old}} - \tfrac{1}{2}\boldsymbol{\mu}_{\text{old}}^T\left(\mathbf{C}^T\boldsymbol{\Lambda}\mathbf{C} + \mathbf{B}\right)\boldsymbol{\mu}_{\text{old}} \\ &+\text{tr}\left[\boldsymbol{\Sigma}_{\text{old}}\left(\mathbf{C}^T\boldsymbol{\Lambda}\mathbf{C} + \mathbf{B}\right)\right]. \end{aligned} \tag{16}$$

The maximisation step then maximizes Q with respect to $\widetilde{\boldsymbol{\xi}}$. Differentiating Q with respect to $\widetilde{\boldsymbol{\xi}}$ and solving the first order optimality conditions we obtain

$$\begin{aligned} \xi_i &:= \sqrt{(\mathbf{C}\boldsymbol{\mu}_{\text{old}})_i^2 + \tfrac{1}{2}(\mathbf{C}\boldsymbol{\Sigma}_{\text{old}}\mathbf{C}^T)_{ii}},\ 1 \leq i \leq n \\ \xi_{n+i} &:= 2/(2\alpha_i + \left(\boldsymbol{\mu}_{\text{old}}^T\mathbf{B}_i\boldsymbol{\mu}_{\text{old}} + \text{tr}\left(\boldsymbol{\Sigma}_{\text{old}}\mathbf{B}_i\right)\right)),\ 1 \leq i \leq v \end{aligned} \tag{17}$$

where $\mathbf{B}_i = \text{blockdiag}(\mathbf{0}_p, \mathbf{D}_i)$, $1 \leq i \leq v$. We then iterate over (15) and (17) until convergence.

4 Grid-Based Posterior Density Approximations

Approximating marginal posterior densities is a common problem in Bayesian statistics. Unfortunately, as we will see in Section 5, using (15) can underestimate the variances of the marginal posterior densities. This tendency of variational approximations has been shown to arise in various settings (Humphreys & Titterington, 2000; Wang & Titterington, 2005; Consonni & Marin, 2007). Here we consider

approximating the marginal posterior densities by directly approximating (4).

Firstly, the joint distribution of $\mathbf{y}$ and $\boldsymbol{\nu}_{-i}$ is, ignoring additive constants,

$$\begin{aligned}\log[\mathbf{y},\boldsymbol{\nu}_{-i}] \;&\geq (\mathbf{C}_{-i}^T\left(\mathbf{y}-\tfrac{1}{2}\right)-\left[\mathbf{C}^T\boldsymbol{\Lambda}\mathbf{C}+\mathbf{B}\right]_{-i,i}\nu_i)\boldsymbol{\nu}_{-i}\\ &\quad -\tfrac{1}{2}\boldsymbol{\nu}_{-i}^T\left[\mathbf{C}^T\boldsymbol{\Lambda}\mathbf{C}+\mathbf{B}\right]_{-i,-i}\boldsymbol{\nu}_{-i}\end{aligned} \tag{18}$$

where $\mathbf{C}_i$ is the ith column of $\mathbf{C}$, $\mathbf{C}_{-i}$ is the matrix $\mathbf{C}$ with the ith column removed, $\boldsymbol{\nu}_{-i}$ is the vector with the ith element removed, $[\mathbf{C}^T\boldsymbol{\Lambda}\mathbf{C}+\mathbf{B}]_{-i,-i}$ is the matrix $\mathbf{C}^T\boldsymbol{\Lambda}\mathbf{C}+\mathbf{B}$ with the ith row and column removed and $[\mathbf{C}^T\boldsymbol{\Lambda}\mathbf{C}+\mathbf{B}]_{-i,i}$ is the ith column of $\mathbf{C}^T\boldsymbol{\Lambda}\mathbf{C}+\mathbf{B}$ with the ith row removed. As before, the right hand side of (18) is, ignoring additive constants, the log of a multivariate Gaussian we make the approximation $[\boldsymbol{\nu}_{-i}|\mathbf{y}]\approx\delta(\boldsymbol{\nu}_{-i})=\phi_{\boldsymbol{\Sigma}}(\boldsymbol{\nu}_{-i}-\boldsymbol{\mu})$ where

$$\boldsymbol{\Sigma}=\left[\mathbf{C}^T\boldsymbol{\Lambda}\mathbf{C}+\mathbf{B}\right]_{-i,-i}^{-1}\quad\text{and}\quad\boldsymbol{\mu}=\boldsymbol{\Sigma}(\mathbf{C}_{-i}^T\mathbf{y}-\left[\mathbf{C}^T\boldsymbol{\Lambda}\mathbf{C}+\mathbf{B}\right]_{-i,i}\nu_i). \tag{19}$$

Let $\boldsymbol{\mu}_{\text{old}}$ and $\boldsymbol{\Sigma}_{\text{old}}$ be the values for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ evaluated at the current values for $\widetilde{\boldsymbol{\xi}}$. It can be shown (e.g. Hinton & van Camp 1993; MacKay, 1995; Attias, 2000), for any density δ, that

$$\begin{aligned}\log[\mathbf{y},\nu_i] &\geq Q(\widetilde{\boldsymbol{\xi}}|\boldsymbol{\mu}_{\text{old}},\boldsymbol{\Sigma}_{\text{old}})+\mathcal{H}_\delta\\ &\geq h(\widetilde{\boldsymbol{\xi}})+\left(\mathbf{y}-\tfrac{1}{2}\right)^T\mathbf{C}\boldsymbol{\mu}_{\text{old}}\\ &\quad -\tfrac{1}{2}\boldsymbol{\mu}_{\text{old}}^T\left(\mathbf{C}^T\boldsymbol{\Lambda}\mathbf{C}+\mathbf{B}\right)\boldsymbol{\mu}_{\text{old}}+\text{tr}\left[\boldsymbol{\Sigma}_{\text{old}}\left(\mathbf{C}^T\boldsymbol{\Lambda}\mathbf{C}+\mathbf{B}\right)\right]\\ &\quad +\tfrac{1}{2}\log|2e\pi\boldsymbol{\Sigma}_{\text{old}}|=\log[\mathbf{y},\nu_i]_L\end{aligned} \tag{20}$$

where $Q(\widetilde{\boldsymbol{\xi}}|\boldsymbol{\mu}_{\text{old}},\boldsymbol{\Sigma}_{\text{old}})=\mathbb{E}_\delta\{\log[\mathbf{y},\boldsymbol{\nu}]\}$ and $\mathcal{H}_\delta=\frac{1}{2}\log|2e\pi\boldsymbol{\Sigma}_{\text{old}}|$ is the entropy function for δ. Maximising for $\widetilde{\boldsymbol{\xi}}$ we obtain

$$\begin{aligned}\xi_i &:= \sqrt{(\mathbf{C}\boldsymbol{\mu}_{\text{old}})_i^2+\tfrac{1}{2}(\mathbf{C}\boldsymbol{\Sigma}_{\text{old}}\mathbf{C}^T)_{ii}},1\leq i\leq n\\ \xi_{n+i} &:= 2/(2\alpha_i+\left(\boldsymbol{\mu}_{\text{old}}^T\mathbf{B}_i\boldsymbol{\mu}_{\text{old}}+\text{tr}\left(\boldsymbol{\Sigma}_{\text{old}}\mathbf{B}_i\right)\right)),1\leq i\leq v.\end{aligned}$$

In order to tighten the bound (20) we maximise $\log[\mathbf{y}, \nu_i; \boldsymbol{\xi}]_L$ with respect to $\boldsymbol{\xi}$. Let

$$\widetilde{\boldsymbol{\xi}}^*(\nu_i) = \underset{\widetilde{\boldsymbol{\xi}}}{\text{argmax}}\{\log[\mathbf{y}, \nu_i; \widetilde{\boldsymbol{\xi}}]_L\} \tag{21}$$

so that $[\mathbf{y}, \nu_i; \widetilde{\boldsymbol{\xi}}^*]_L$ is also a tight lower bound for $[\mathbf{y}, \nu_i]$ noting we write $\widetilde{\boldsymbol{\xi}}^*(\nu_i)$ because of the fact that $\widetilde{\boldsymbol{\xi}}^*$ depends implicitly on ν_i via the optimisation problem (21).

Given $[\mathbf{y}, \nu_i; \widetilde{\boldsymbol{\xi}}^*(\nu_i)]_L$ we could approximate the marginal likelihood by $[\mathbf{y}]_L \equiv \int [\mathbf{y}, \nu_i; \widetilde{\boldsymbol{\xi}}^*(\nu_i)]_L d\nu_i$ The complicated dependency of $\widetilde{\boldsymbol{\xi}}^*$ on ν_i means that it may be impossible to find a closed form expression for $[\mathbf{y}, \nu_i; \widetilde{\boldsymbol{\xi}}^*(\nu_i)]_L$. Instead we evaluate $\widetilde{\boldsymbol{\xi}}_j^* = \max_{\boldsymbol{\xi}}[\mathbf{y}, \widehat{\nu}_{ij}; \widetilde{\boldsymbol{\xi}}]_L$ for a grid of values $(\widehat{\nu}_{i1}, \ldots, \widehat{\nu}_{iN})$ for some integer N. We then approximate $\log[\mathbf{y}, \nu_i; \widetilde{\boldsymbol{\xi}}^*(\nu_i)]_L$ by some curve $\log[\mathbf{y}, \nu_i]_G$ (where the subscript G denotes a grid based approximation) such that

$$\log[\mathbf{y}, \widehat{\nu}_{ij}]_G = \log[\mathbf{y}, \widehat{\nu}_{ij}; \widetilde{\boldsymbol{\xi}}_j^*]_L \quad \text{for} \quad 1 \leq j \leq N, \tag{22}$$

i.e. $\log[\mathbf{y}, \widehat{\nu}_{ij}]_G$ interpolates the points $(\widehat{\nu}_{ij}, \log[\mathbf{y}, \widehat{\nu}_{ij}; \widehat{\boldsymbol{\xi}}_j]_L)$ for $1 \leq j \leq N$. Finally a grid based variational posterior approximation (GBVPA) for $[\nu_i|\mathbf{y}]$ is given by

$$[\nu_i|\mathbf{y}]_G \equiv [\mathbf{y}, \nu_i]_G / [\mathbf{y}]_G \tag{23}$$

where the one dimensional integral $[\mathbf{y}]_G \equiv \int [\mathbf{y}, \nu_i]_G d\nu_i$ is evaluated numerically.

There are a number of choice to be made. These include the choice and number of grid values, type of interpolation used to approximate $\log[\mathbf{y}, \nu_i; \widetilde{\boldsymbol{\xi}}^*(\nu_i)]_L$ and quadrature method to approximate $[\mathbf{y}, \nu_i]_G$. The choices we have made in the following examples are as follows: 1) Suppose that (ν_{iL}, ν_{iR}) is a 95% highest posterior density credible region for ν_i based on (15). Then we let $(\widehat{\nu}_{i1}, \ldots, \widehat{\nu}_{iN})$

be equally spaced on the interval $\nu_i \in (\nu_{iL} - \triangle/2, \nu_{iR} + \triangle/2)$ where $\triangle = \nu_{iR} - \nu_{iL}$. 2) We used the `R` function `spline()` for the interpolater $[\mathbf{y}, \nu_i]_G$. 3) A 5,000 point trapezoid rule was used to approximate $[\mathbf{y}]_G \equiv \int [\mathbf{y}, \nu_i]_G d\nu_i$ on the interval $\nu_i \in (\nu_{iL} - \triangle/2, \nu_{iR} + \triangle/2)$.

One possible downside of GBVPA is that N optimisation problems of the form (21) need to be solved for each marginal posterior density. Thus, in practice, we seek to choose the grid $(\widehat{\nu}_{i1}, \ldots, \widehat{\nu}_{iN})$ with as few points as possible but enough points to ensure that we have a reasonable approximation for $[\nu_i|\mathbf{y}]_G$. We note that GBVPA could potentially be easily improved however we propose GBVPA as a starting place for such improvements.

5 Numerical Experience

As outlined in the introduction there are many applications to logistic mixed models. For simplicity we will examine the effectiveness of variational approximations for logistic mixed models for the context of additive penalized spline smoothing (e.g. Eilers & Marx, 1996; Ruppert *et al.*, 2003; Wood, 2003; Welham, Cullis, Kenward & Thompson, 2007; Wand & Ormerod, 2008). Wand & Ormerod (2008) Section 5 considered a penalised spline analysis of union membership for a sample of 534 U.S. workers (source: Berndt, 1991) with the subset of covariates

$$\mathbf{x}_i = [\texttt{south}_i,\ \texttt{female}_i,\ \texttt{married}_i,\ \texttt{years.educ}_i,\ \texttt{wage}_i,\ \texttt{age}_i].$$

The variables `years.educ`, `years.experience`, `wage` and `age` are continuous and the variables `south`, `female` and `married` are binary. We consider a model of the form

$$\begin{aligned}\text{logit}\left\{\mathbb{P}\left(\texttt{union.member}_i = 1|\mathbf{x}_i\right)\right\} &= \beta_0 + \beta_1 \texttt{south}_i + \beta_2 \texttt{female}_i \\ &+ \beta_3 \texttt{married}_i + f_{\texttt{years.educ}}(\texttt{years.educ}_i) + f_{\texttt{wage}}(\texttt{wage}_i) \\ &+ f_{\texttt{age}}(\texttt{age}_i) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}\end{aligned}$$

and use the mixed model formulation of cubic O'Sullivan splines (Wand & Ormerod 2008, Section 4) to model $f_{\texttt{years.educ}}$, $f_{\texttt{wage}}$ and $f_{\texttt{age}}$. We

used $K = 25$ quantile spaced inner knots for each of the continuous variables. Let $\mathbf{Z}_{\texttt{years.educ}}$, $\mathbf{Z}_{\texttt{wage}}$ and $\mathbf{Z}_{\texttt{age}}$ be the spline matrices for `years.educ`, `wage` and `age` respectively. Each of these matrices has $q_i = K + 2$ columns and

$$\mathbf{X} = [1, \texttt{years.educ}_i, \texttt{wage}_i, \texttt{age}_i]_{1 \le i \le n}\,, \quad \mathbf{Z} = [\mathbf{Z}_{\texttt{years.educ}}, \mathbf{Z}_{\texttt{wage}}, \mathbf{Z}_{\texttt{age}}]$$

and $\mathbf{D}_{\sigma^2} = \text{blockdiag}\left\{\sigma_i^{-2}\mathbf{I}_{q_i}\right\}_{1 \le i \le 3}$.

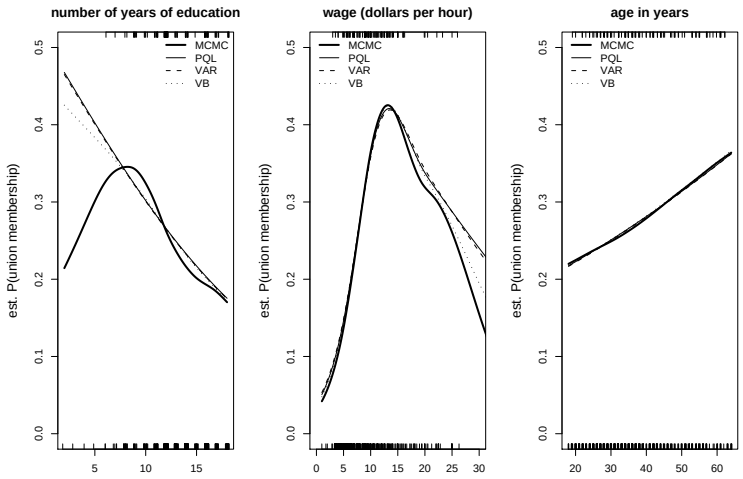


Figure 2: *Smooth function fits for the* `Trade Union` *model using the MCMC, PQL, VAR and VB approximations.*

From Figure 2 we see that the VAR and VB approximations produce fits which are quite similar to PQL. The approximations for the VB were, to the eye, slightly closer to the MCMC fit than the other approximations. While the analytic approximations are less accurate than MCMC methods they are extremely fast; taking about 4 minutes to fit each model. In contrast, for this example, the MCMC

approximation using `WinBUGS` (Spiegelhalter, Thomas & Best, 2000) took a little over an hour.

Finally, for Bayesian logistic mixed we can compare posterior density approximations for `south`, `female` and `married` using (9) and grid-based posterior approximations described in Section 4 with kernel density estimates (Scott, 1992; Wand & Jones, 1995) of posterior samples obtained via MCMC. The kernel density estimates use the Gaussian kernel with the bandwidth chosen via a direct plug-in method (Wand & Jones, 1995, Section 3.6) using the `R` package `KernSmooth`. Alternatively, the Sheather-Jones method (Sheather & Jones, 1991) can deliver excellent results.

It has been well-established in kernel smoothing literature that the choice of kernel has little effect on density estimates (e.g. Marron & Nolan, 1988, Wand & Jones, Chapter 2). However, the how the bandwidth is chosen does matter. Extensive simulation studies (e.g. Park & Turlach, 1992; Cao, Cuevas & Gonzalez-Manteiga, 1994; Jones, Marron & Sheather, 1996) have shown that, for large sample sizes and densities that are Gaussian in shape, automatic bandwidth methods such as the direct plug-in methods and the Sheather-Jones method lead to quite accurate density estimates.

For the MCMC fit we used `WinBUGS` to generate chains of length 50,000 after a burn-in of 5,000 and applied a thinning factor of 5, resulting in posterior samples of size 10,000 and then used the `R` package `KernSmooth` to estimate the densities of these posterior samples. Figure 3 illustrates these estimates. From this figure we notice that densities approximations based on (9) underestimate the amount of variance of the posteriors. The GBVPA approximations, on the other hand, were significantly better. Each GBVPA approximation took roughly 5 minutes to compute while the MCMC approach via `WinBUGS` took a little over 6 hours.

6 Conclusion

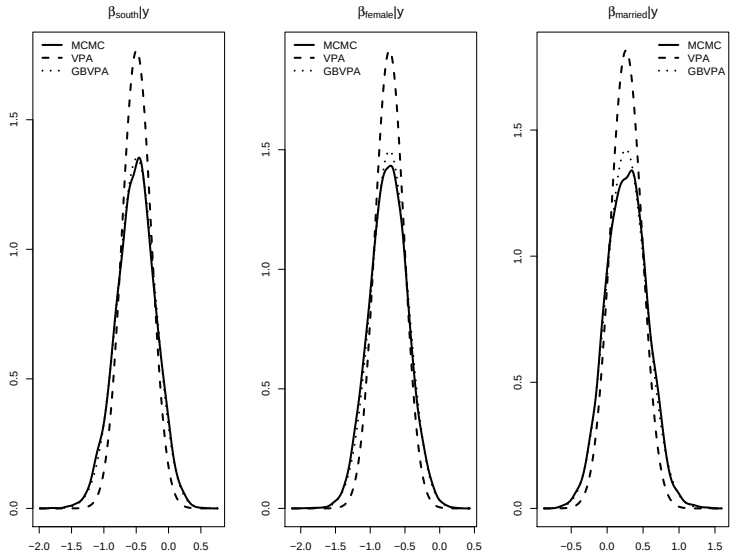


Figure 3: *Illustration of the kernel density estimates of MCMC posterior samples, variational posterior approximations (VPA) and grid-based variational posterior approximations (GBVPA) for* `south`, `female` *and* `married` *coefficients.*

In this article we focused on the logistic mixed model case. The extension to Generalised linear mixed models requires additional ideas which we do not have the space to cover here. In an upcoming paper Ormerod & Wand (2008) cover these cases for general responses including Poisson, gamma and inverse-Gaussian response types.

References

[1] Attias, H (2000). A variational Bayesian framework for graphical models. In Solla, S., Leen, T., and Muller, K.-R. (Eds.), *Advances in Neural Information*

Processing Systems, Volume 12, 209-215. Cambridge, MA: MIT press.

[2] Beal, M. (2003). *Variational algorithms for approximate Bayesian inference*. Ph.D. thesis, Gatsby Computational Neuroscience Unit.

[3] Beal, M. and Ghahramani, Z. (2002). The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures. In *Bayesian Statistics*, Volume 7. Oxford: Oxford University Press.

[4] Berndt, E.R. (1991). *The Practice of Econometrics: Classical and Contemporary*. Reading, Massachusetts: Addison-Wesley.

[5] Bishop, C.M. and Winn, J. (2003). Structural variational distributions in VIBES. In Bishop, C.M. and Frey, B. (Eds.), *Proceedings of Artificial Intelligence*, Florida, USA.

[6] Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal American Statistical Association* 88(1), 9-25.

[7] Breslow, N.E. and Lin, X. (1995). Bias correction in generalized linear mixed models with a single component of dispersion. *Biometrika* 82(1), 81-91.

[8] Cao, R., Cuevas, A., and González Manteiga, W. (1994). A comparative study of several smoothing methods in density estimation. *Computational Statistics & Data Analysis* 17(2), 153-176.

[9] Consonni, G. and Marin, J.-M. (2007). Mean-field variational approximate Bayesian inference for latent variable models. *Computational Statistics Data Analysis* 52(2), 790-798.

[10] Cowles, M.K. and Carlin, B.P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of American Statistics Association* 91, 883-904.

[11] Cowles, M.K. and Rosenthal, J.S. (1998). A simulation approach to convergence rates for markov chain Monte Carlo. *Statistics and Computing* 8, 115-124.

[12] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, Series B 339, 1-38.

[13] Eilers, P. and Marx, B.D. (1996). Flexible smoothing with B-splines and penalties (with comments and rejoinder). *Statistical Science* 11(2), 89-121.

[14] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. (Comment on an article by Browne & Draper, 2006). *Bayesian Analysis* 1(3), 515-534.

[15] Gilks,W. R., Richardson, S., and Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

[16] Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized Additive Models*. London: Chapman & Hall.

[17] Hastings,W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1), 97-109.

[18] Hinton, G.E. and van Camp, D. (1993). Keeping the neural networks simple by minimizing the description length of the weights. In *COLT 93: Proceedings of the sixth annual conference on Computational learning theory*, New York, USA, 5-13. Association for Computing Machinery.

[19] Humphreys, K. and Titterington, D. (2001). Some examples of recursive variational approximations for Bayesian inference. In Opper, M. and Saad, D. (Eds.), *Advances Mean Field Methods: Theory and Practice*. Cambridge, MA: MIT Press.

[20] Jaakkola, T.S. (2001). Tutorial on variational approximation methods. In Opper, M. and Saad, D. (Eds.), *Advanced Mean Field Methods: Theory and Practice*. Cambridge, MA: MIT Press.

[21] Jaakkola, T.S. and Jordan, M.I. (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing* 10, 25-37.

[22] Jones, M. C., Marron, J.S., and Sheather, S.J. (1996). Progress in data-based bandwidth selection for kernel density estimation. *Computational Statistics* 11(3), 337-381.

[23] Jordan, M.I., Ghahramani, Z., Jaakkola T.S., and Saul, L.K. (1999). An introduction to variational methods for graphical models. *Machine Learning* 37(2), 183-233.

[24] Lin, X. and Breslow, N.E. (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association* 91(435), 1007-1016.

[25] MacKay, D.J.C. (1995). Developments in probabilistic modelling with neural networks - ensemble learning. In Kappen, B. and Gielen, S. (Eds.), *Neural Networks: Artificial Intelligence and Industrial Applications. Proceedings of the 3rd Annual Symposium on Neural Networks*, Netherlands. Nijmegen.

[26] Marron, J.S. and Nolan, D. (1988). Canonical kernels for density estimation. *Statistics & Probability Letters* 7(3), 195-199.

[27] McCullagh, P. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* 92, 162-170.

[28] McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models* (2nd ed.). London: Chapman & Hall.

[29] McCulloch, C.E. and Searle, S.R. (2001). *Generalized, Linear, and Mixed Models*. New York: John, Wiley & Sons.

[30] Neal, R.M. and Hinton, G.E. (1998). A new view of the EM algorithm that justifies incremental, sparse and other variants. In Jordan, M.I. (Ed.), *Learning in Graphical Models*, 355-368. Kluwer Academic Publishers.

[31] Park, B.U. and Turlach, B.A. (1992). Practical performance of several data driven bandwidth selectors (with discussion). *Computational Statistics* 7(3), 251-270. Correction in Vol. 9, p. 79.

[32] Rijmen, F. and Vomlel, J. (2007). Assessing the performance of variational methods for mixed logistic regression models. *Journal of Statistical Computation and Simulation*. (accepted).

[33] Robert, C. and Casella, G. (1999). *Monte Carlo Statistical Methods* (2nd ed.). New York: Springer-Verlag.

[34] Rockafellar, R. (1972). *Convex Analysis*. Princeton, New Jersey: Princeton University Press.

[35] Rosenthal, J.S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of American Statistics Association* 90, 558-566.

[36] Ruppert, D., Wand, M.P., and Carroll, R.J. (2003). *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics. New York: Cambridge University Press.

[37] Scott, D.W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York, Chichester: John Wiley & Sons.

[38] Sheather, S.J. and Jones, M.C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Series B* 53, 683-690.

[39] Spiegelhalter, D., Thomas, A., and Best, N. (2000). WinBUGS Version 1.3 User Manual. www.hrc-bsu.cam.ac.uk/bugs.

[40] Tuerlinckx, F., Rijmen, F., Verbeke, G., and de Boeck, P. (2006). Statistical inference in generalized linear mixed models: A review. *British Journal of Mathematical and Statistical Psychology* 59, 225-255.

[41] Verbeke, G. and Molenberghs,G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.

[42] Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. London: Chapman & Hall.

[43] Wand, M.P. and Ormerod, J.T. (2008). On semiparametric regression with O'Sullivan penalised splines. *Australian & New Zealand Journal of Statistics*. (to appear).

[44] Wang, B. and Titterington, D.M. (2005). Inadequacy of interval estimates corresponding to variational Bayesian approximations. In Cowell, R.G. and Ghahramani, Z. (Eds.), *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, 373-380. Society for AISTATS.

[45] Welham, S.J., Cullis, B.R., Kenward, M.G., and Thompson, R. (2007). A comparison of mixed model splines for curve fitting. *Australian & New Zealand Journal of Statistics* 49, 1-23.

[46] Winn, J. and Bishop, C. M. (2005). Variational message passing. *Journal of Machine Learning Research* 6, 661-694.

[47] Wood, S.N. (2003). Thin-plate regression splines. *Journal of the Royal Statistical Society, Series B* 65, 95-114.

[48] Zhao, Y., Staudenmayer, J., Coull, B.A., and Wand, M.P. (2006). General Design Bayesian Generalized Linear Mixed Models. *Statistical Science* 21, 35-51.

On Nonparametric Density Estimation in Mixture Models

Reza Pakyari

Department of Mathematics, Arak University

Abstract: A new nonparametric estimation method for the component density functions of a multivariate mixture model is proposed. The estimation procedure is based on inverting the mixture model that is expressed in terms of densities. A monte Carlo simulation study shows that the nonparametric estimator performs relatively better than its parametric competitor when estimating the most probable component and when the value of the mixing proportion is close to one.

Keywords: Nonparametric density estimation, Independent marginals, Kernel methods, Bandwidth.

1 Introduction

In many statistical applications, it is known or suspected that observations arise from two or more populations with different distributions mixed in varying proportions. For example, the distribution of height in a population of adults might be decomposed into male and female populations, or similarly in fisheries research, where usually fish lengths are available, but not their sexual identities.

Hettmansperger and Thomas (2000) proposed an almost nonparametric approach for estimation of the mixing proportions, and also for estimation of the number of components in the mixture.

Hall et al. (2005) presented a nonparametric approach based on inversion the mixture model in terms of distribution functions for estimation of the mixing proportions and component distributions in multivariate mixtures. They also suggest the density estimation by differentiating the distribution function estimators. In this talk, we suggest another density estimators based on inverting the mixture model that is expressed in terms of densities rather than distribution functions.

2 Density Estimation

Consider a k-variate two-term mixture model in terms of densities,

$$\phi = \pi \prod_{i=1}^{k} f_{1i} + (1-\pi) \prod_{i=1}^{k} f_{2i}\,. \tag{1}$$

This model implies a set of lower-dimensional submodels

$$\phi_{i_1 \ldots i_\ell} = \pi \prod_{m=1}^{\ell} f_{1i_m} + (1-\pi) \prod_{m=1}^{\ell} f_{2i_m}\,, \tag{2}$$

where $1 \le \ell \le k$, $1 \le i_1 < \ldots < i_\ell \le k$, and $\phi_{i_1 \ldots i_l}$ denotes the ℓ-variate "marginal" density of ϕ corresponding to vector components with indices $i_1, \ldots, i_\ell$. Following the steps in Hall et al. (2005) by inversion of the mixture model (1) we get

$$f_{1i} = \pm\left(\frac{(1-\pi)\,\psi_{ii_1}\,\psi_{ii_2}}{\pi\,\psi_{i_1 i_2}}\right)^{1/2} + \phi_i\,, \quad f_{2i} = \mp\left(\frac{\pi\,\psi_{ii_1}\,\psi_{ii_2}}{(1-\pi)\,\psi_{i_1 i_2}}\right)^{1/2} + \phi_i\,, \tag{3}$$

where $\psi_{i_1 i_2} = \phi_{i_1 i_2} - \phi_{i_1}\,\phi_{i_2}$, and the $+$ and $-$ signs have the same interpretation they do in estimators of Hall et al. (2005). Formulae (3) could be used to estimate f_{ji}. In particular, we could construct kernel estimators of ϕ_i and $\phi_{i_1 i_2}$, leading to the obvious estimators of $\psi_{i_1 i_2}$. We could substitute these, and the estimator of π into (3), to obtain estimators of f_{1i} and f_{2i}. We would average over suitable pairs (x_{i_1}, x_{i_2}) in the sample space.

3 Simulation Studies

Our estimator of the component density function depend on the threshold ϵ_3. We used an empirical method to find the optimal values for the threshold parameter. We generated 300 datasets each of size $n = 500$ from the trivariate mixture model

$$\phi(\boldsymbol{x}) = \pi\, n_3(\boldsymbol{x}, \boldsymbol{\mu_1}, \boldsymbol{I_3}) + (1-\pi)\, n_3(\boldsymbol{x}, \boldsymbol{\mu_2}, \boldsymbol{I_3})\,, \tag{4}$$

where $n_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the trivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. In particular, we chose the 3×3 identity matrix $\boldsymbol{I_3}$ as the variance-covariance matrix, and $\boldsymbol{\mu_1} = (0,0,0)$ and $\boldsymbol{\mu_2} = (4,4,4)$ as the mean vectors of the first and second components, respectively.

We also fitted a Gaussian mixture model and estimated the parameters by the maximum likelihood estimation method via the EM algorithm.

Figure 1 shows the results of the simulation study for both parametric and nonparametric methods. In each figure, the three panels in the first column graph the root MISE for the respective marginal density functions of the first component, and the three panels in the second column correspond to the three marginal density functions of the second component.

It can be seen from this figure that the parametric estimator performs better than the nonparametric estimator when estimating the first component. However, when used to estimate the most probable component, the nonparametric estimator tends to performs relatively close to its parametric competitor and even better when the mixing proportion is greater than 0.30.

References

[1] Everitt, B.S. & Hand, D.J. (1981). *Finite Mixture Distributions.* Chapman & Hall, London.

[2] Hall, P. and Neeman, A. and Pakyari, R. and Elmore, R. (2005). Nonparametric Inference in Multivariate Mixtures. *Biometrika* **92**, 667-678.

[3] Hettmansperger, T. P. and Thomas, H. (2000). Almost Nonparametric inference for repeated measures in mixture models. *Journal of the Royal Statistical Society, Series B* **62**, 811-825.

[4] Lindsay, B.G. (1995). *Mixture Models: Theory, Geometry, and Applications.* Institute of Mathematical Statistics, Hayward, CA.

[5] McLachlan, G.J. & Basford, K.E. (1988). *Mixture Models: Inference and Applications to Clustering.* Wiley, New York.

[6] McLachlan, G.J. & Peel, D. (2000). *Finite Mixture Models.* Wiley, New York.

[7] Titterington, D.M., Smith, A.F.M. & Makov, U. (1985). *Statistical Analysis of Finite Mixture Distributions.* Wiley, New York.

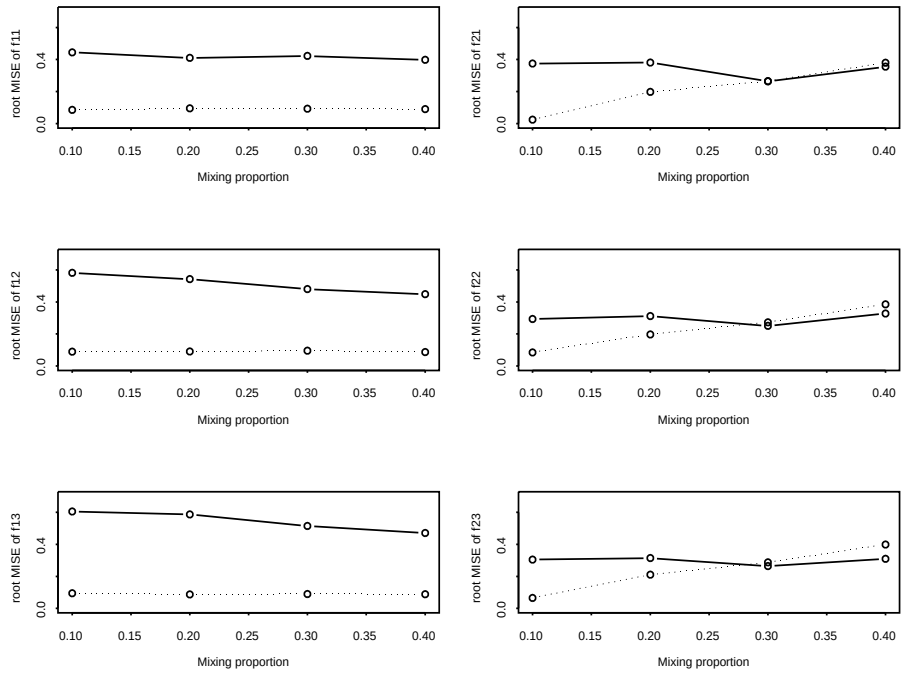


Figure 1: Root mean integrated squared errors of parametric, dotted lines, and nonparametric, solid lines, estimators of marginal densities. The three plots in the first [respectively, second] column depict root MISE against the mixing proportion π, for estimates of the three marginal densities of the first [second] component in the normal model at equation (4). Sample size is 500.

Licensing A New Treatment; A Frequentist Condition on a Bayesian Approach

Hamid Pezeshk

Center of Excellence in Biomathematics, School of Mathematics, Statistics and Computer Science, College of Science, University of Tehran

Abstract: The determination of sample size is an important issue in designing any trial. Particularly it is of key importance in medical studies. Sample size computations are largely based on frequentist or classical methods. Many authors have looked at the problem and many papers have been written to discuss the problem form both the frequentist and Bayesian standpoints. In the Bayesian approach the prior information on the unknown parameters is taken into account.

In this work we consider a fully Bayesian approach to the sample size determination problem which treats the problem as a decision problem and employs a utility function to find the optimal sample size of the trial. Furthermore, we assume that a regulatory authority, which is deciding on whether or not to grant a license to a new treatment, uses a frequentist approach. We then find the optimal sample size for the trial by maximizing the expected net benefit, which is the expected benefit of subsequent use of the new treatment minus the cost of the trial. The aim is to apply the methodology on an actual trial to show how this might work.

Keywords: Sample Size Determination, Fully Bayesian Approach, Regulatory Authority, Expected Net Benefit, Cost, Normal Distribution, Number of Subsequent Users.

1 Introduction

One of the most important questions in planning any experiment is how big to make the trial. Particularly it is of key importance in medical studies. Several researchers have recognized the value of using prior distributions rather than point estimates in sample size calculation. Bayesian publications on the sample size question, which uses a prior distribution for the unknown parameters, may be divided into two groups: inferential (see, for example, Adcock (1988), Spiegelhalter

et al (1994)) and fully Bayesian or decision theoretic methods, which treat the problem as a decision problem and employ a loss or utility function (see, for example, Grundy *et al* (1956), Lindley (1997), Gittins and Pezeshk (2000), Pezeshk and Gittins (2006a), (2006b), Willan (2007) and Kikuchi *et al* (2008)).

Bayesian sample size determination for estimating the success probability in binomial sampling has received considerable attention (see, for example, Adcock (1995), Joseph *et al* (1995), Pham-Gia and Turkkan (1992), Pezeshk and Gittins (2002)).The case of the difference between two binomial parameters has been considered by Joseph *et al* (1997) and Pham-Gia and Turkkan (2003).

Here we present the results of applying a fully Bayesian analysis to an actual trial for which we assume that the regulator takes a frequentist's approach to grant a license to the new service or the new treatment.

Pezeshk and Gittins (2006a) introduce the function shown in figure 1 to represent the number of subsequent users of the new treatment.

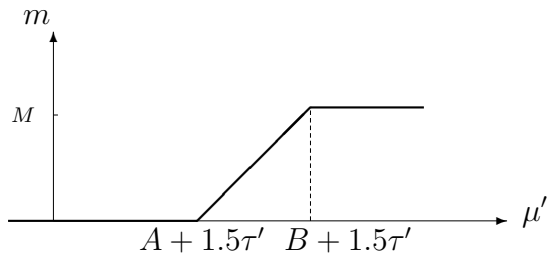


Figure 1: Number of subsequent users,

Here μ' is the posterior mean and τ' is the posterior standard deviation of the parameter of interest, δ, the difference between the

performance, on some appropriate scale, of the new treatment and of the existing one. M is the expected total number of users if the new treatment turns out to be very good in every respect, and A and B are constants which need to be estimated in advance.

$$m = \begin{cases} 0, & \mu' < A + 1.5\tau' \\ \frac{M}{B-A}[\mu' - A - 1.5\tau'], & A + 1.5\tau' < \mu' < B + 1.5\tau' \\ M, & B + 1.5\tau' < \mu'. \end{cases} \quad (1)$$

This implies that if the amount of improvement is less than $A + 1.5\tau'$, then nobody switches to the new treatment; if it is between $A + 1.5\tau'$ and the higher value $B + 1.5\tau'$, then some patients switch to the new treatment; and if it is higher than $B + 1.5\tau'$ then all patients switch to the new treatment.

To simplify the presentation we assume throughout that the variance of the observations σ^2 is known.

2 Formal Statement of the Problem

Assume that Z_1, Z_2, ... are *i.i.d.* random variables with a normal density, $N(\delta, \sigma^2)$, and let us also assume that $\delta \sim N(\mu, \tau^2)$. Z_i's might be considered as the difference between the performance of the new service or treatment and that of an existing one. As introduced above, let m denote the number of subsequent users of the new treatment and let the cost of carrying out a trial with n users or patients be cn.

The benefit per user or patient of using the new treatment is assumed to be b. The objective (i.e. expected net benefit) function which is proposed, consisting of the total benefit from the resulting change in the number of patients using the new treatment minus the cost of the trial, can be written as

$$r(n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} mbd\Pi^n(\delta|\overline{z}_n) f(\overline{z}_n) d\delta d\overline{z}_n - cn. \quad (2)$$

Here $f(\overline{z}_n)$ is the predictive density function of $\overline{Z_n}$ and $\Pi^n(\delta|\overline{z}_n)$ is the posterior distribution of the unknown parameter δ given $\overline{z}_n$. This is $N(\mu'(\overline{z}_n), \tau'^2)$, where $\mu'(\overline{z}_n) = \frac{\sigma^2\mu + n\tau^2\overline{z}_n}{\sigma^2+n\tau^2}$ and $\tau'^2 = \frac{\sigma^2\tau^2}{\sigma^2+n\tau^2}$. Thus we may carry out the inner integration with respect to δ in (2) to give

$$r(n) = bE[m\mu'(\overline{Z}_n)] - cn. \tag{3}$$

Since $\delta \sim N(\mu, \tau^2)$, $\overline{Z}_n - \delta \sim N(0, \sigma^2/n)$, and δ and $\overline{Z}_n - \delta$ are independent, it follows that $f(\overline{z}_n)$ is $N(\mu, \tau^2 + \sigma^2/n)$.
It is convenient to express the problem in terms of the scaled variables $R(n) = \frac{r(n)}{Mb}$ and $C = \frac{c}{Mb}$. Numerical calculations show that in all cases $R(n)$ has a unique maximum as a function of n (for more details see Kikuchi, Pezeshk and Gittins (2008)).

3 Licensing the New Service or the New Treatment

Following Pezeshk (2006) let us assume that the regulator intends to test the hypothesis H_0 that $\delta = 0$ against the alternative H_1 that $\delta > 0$. If the data are normally distributed and if the *size* of the test is α then the regulator's criterion for rejecting H_0 in favor of H_1 is

$$\text{reject } H_0 \text{ if and only if } \overline{z}_n > z_\alpha\sqrt{\frac{\sigma^2}{n}}, \tag{4}$$

where z_α is such that $P(Z \geq z_\alpha) = \alpha$ for a standard normal variate Z. The regulator grants a license to the new treatment, provided that (4) holds.

Pezeshk (2006) show that the sample size determination with the above setting for a trial could be dealt with using the following theorem.

Theorem If the regulator's requirement is as in (4) and the benefit of using the new treatment per patient is b, then the expected net

benefit for carrying out a trial of size n is

$$r(n) = bE[m\mu'(\overline{Z}_n)] - cn.$$

The scaled expected net benefit can be written as

$$\begin{aligned} R(n) = \frac{1}{B-A}\int_{max(h_1,f_r)}^{max(h_2,f_r)} [D_c + \sqrt{n}T_c(T_c^{-2}+n)^{-1/2}u]\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}du \\ - \frac{A}{B-A}\int_{max(h_1,f_r)}^{max(h_2,f_r)} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}du \\ - \frac{1}{B-A}1.5(T_c^{-2}+n)^{-1/2}\int_{max(h_1,h_r)}^{max(h_2,f_r)} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}du \\ + \int_{max(h_2,f_r)}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}du - Cn. \end{aligned} \tag{5}$$

Here $T_c = \frac{\tau_c}{\sigma}$, $D_c = \frac{\mu_c}{\sigma}$, $C = \frac{c}{Mb}$, $R(n) = \frac{r(n)}{Mb}$, $f_r(\alpha, n) = \frac{z_\alpha - \sqrt{n}T_c}{\sqrt{1+nT_c^2}}$,

$$h_1(A,n) = \frac{[A + 1.5(T_c^{-2}+n)^{-1/2} - D_c](T_c^{-2}+n)^{1/2}}{T_c\sqrt{n}},$$
$$h_2(B,n) = \frac{[B + 1.5(T_c^{-2}+n)^{-1/2} - D_c](T_c^{-2}+n)^{1/2}}{T_c\sqrt{n}}.$$

Note that we have used the above change of variable to convert

the inequalities $z_\alpha\sigma/\sqrt{n} < \overline{z}_n < \frac{\sigma(\sigma^2+n\tau_c^2)(B+1.5\tau_c'/\sigma)-\sigma^2\mu_c}{n\tau_c^2}$ to $f_r(\alpha,n) < u < h_2(B,n)$.

Proof Evaluating the function m in equation (3) and taking the

condition expressed in (4) into account after some calculations we will arrive at (5).

Maximizing (5), one can find the optimal sample size of the trial.

A computer program has been written to calculate the expected net benefit function $r(n)$ and the maximizing value n^* of n. The program uses the numerical integration and optimization routines written in the *MatLab* system. Extensive numerical calculation shows that, for all values of the various parameters, the expected net benefit $r(n)$ has a unique local maximum that is also the overall maximum. In what follows we present the results of applying our fully Bayesian methodology to an actual trial and show how it may deal with a frequentist regulatory authority.

4 A Trial for the Suitability of a Treatment

A trial for the suitability of a certain treatment for juveniles was carried out by an international drug company. A 0-1 quantity (some hair loss=0, no hair loss=1) for severely ill hospitalized patients (ie, patients suffering from the effects of chemotherapy) was observed. We worked in term of the log odds-ratio δ and assumed central limit theorem normality to calculate the optimal sample size which maximized the expected net benefit resulting from the trial. The value $\sigma = 2$ was based on the company's previous experience with similar trials. For the binomial data $\overline{Z}_n$ was redefined as $log[\frac{1}{n}\Sigma X_i/(1-\frac{1}{n}\Sigma X_i)] - log[\frac{1}{n}\Sigma Y_i/(1-\frac{1}{n}\Sigma Y_i)]$ and δ as $log[p_1/(1-p_1)] - log[p_2/(1-p_2)]$, where $p_1 = P_r(\text{cure} \mid \text{new treatment})$ and $p_2 = P_r(\text{cure} \mid \text{alternative treatment})$. Now $X_i = 1$ if the i^{th} patient responds favourably to the new treatment, and otherwise $X_i = 0$, and Y_i is defined similarly for the i^{th} patient receiving the alternative treatment. Inference was based on the fact that $\overline{Z}_n$ has a distribution which is approximately normal with the mean δ and the variance $\frac{2(p_1(1-p_1)+p_2(1-p_2))}{np_1(1-p_1)p_2(1-p_2)}$ see Coad and Rosenberger (1999). For rare events, Spiegelhalter *et al* (1994) suggest

an approximation of $4/n$ for the variance of $\overline{Z}_n$.

$\sigma = 2$. This was based on the company's previous experience with similar trial.

$\mu = 2.09$. This is the log odds-ratio with success probabilities of 0.2 with placebo and 0.67 with the active cream, which we assume to be the expected value.

$\tau = \mu/2 = 1.045$. A convenient representation of prior ignorance. It means that

$$P_r(\delta < 0) = P_r(\delta > 2\mu) \simeq 0.025.$$

$A = 0.8\mu = 1.672$. Maximum critical difference for no sale.

$B = 1.2\mu = 2.058$. Minimum critical difference for maximum sale.

$c = \$3800$. The total cost per patient, to include payment to the doctor and hospital costs.

$Mb = \$9.5M$. An estimate for the sale, arrive at along the following lines, for the total additional return if the trial gives a clear positive result.

Let

$I =$ Current annual income from the treatment,

$I(1+x) =$ Future annual income if this trial gives a clear positive result,

$(1 + r) =$ (value of being able to show now that the trial result is clearly positive)/(value of being able to show in one year from now that the trial result is clearly positive)

Thus r is a discount rate. It is convenient to express it in 'real' terms, with the effect if inflation removed. Its main components are attributable to obsolescence because of advances in alternative therapies and to reduction of the remaining period of patent protection.

Thus the expected net benefit if the trial is clearly successful is

$$Mb = Ix(1 + \frac{1}{1+r} + \frac{1}{(1+r)^2} + \frac{1}{(1+r)^3} + ...) = Ix(1 + \frac{1}{r}).$$

For example if $I = \$9.5M$, $x = 0.1$ and $r = 0.1$, then $Mb = \$10.45M$.

Taking $\alpha = 0.05$, the function $r(n)$ in equation (5) was maximized and the optimal sample size n^* with an intermediate regulator requirement was found to be 215, resulting in an expected net benefit $r(n^*)$ of \$4.18 million. Figure 2 shows the variation of $r(n)$ as a function of n.

It should be noted that, based on the typical requirements for calculating the sample size in a classical or frequentist framework with a size of 5% and a power of 80% the company concerned recruited 40 patients in the above trial.

So, the optimal sample size based on our analysis is larger than the one used in the practice. Since the cost of carrying out the trial against the possible benefits of the new therapy, the classical procedure gives, in particular, no indication of those cases for which the likely benefits do not justify carrying out a trial at all, or of those cases for which the possible benefits justify a larger trial to increase the confidence of both the potential users and the authorities in new therapy and thereby persuade more of them to use it.

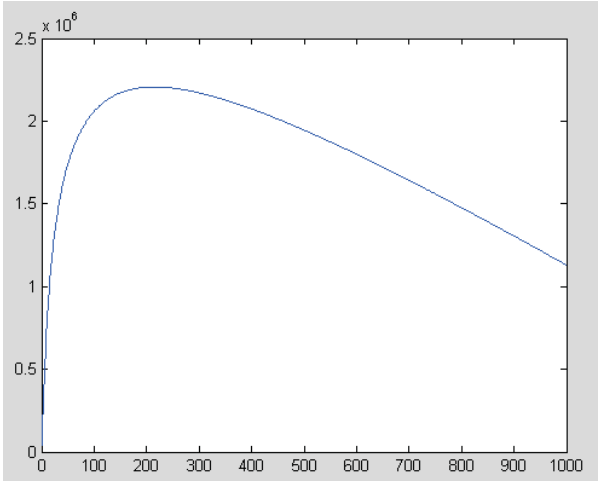


Figure 2: Expected net benefit fuction as a function of n, when $D = 1.045$, $T = 0.5225$, $C = 0.0004$, and $(A, B) = (1.672, 2.058)$, as discussed in the case study.

References

1. Adcock, C. J. (1988) A Bayesian Approach to Calculating Sample Sizes. *The Statistician* **37,** 433 – 439.

2. Adcock, C. J. (1995) The Bayesian Approach to the Determination of Sample Sizes - some comments on the paper by Joseph, Wolfson, and Berger. *The Statistician* **44,** 155 – 161.

3. Coad, D. S., Rosenberger, W. F. (1999) A comparison of the randomized play-the-winner rule and triangular test for clinical trials with binary responses. *Statistics in Medicine* **18,** 761–769.

4. Gittins, J. C. and Pezeshk H. (2000) How Large Should a Clinical Trial Be? *The Statistician* **49,** 177 – 187.

5. Grundy, P. M., Healy, M. J. R. and Rees D. H. (1956) Economic Choice of the Amount of Experimentation. *Journal of the Royal Statistical Society A* **18,** 32 – 38.

6. Joseph, L., Wolfson, D. B., and du Berger, R. (1995) Sample Size Calculations for Binomial Proportions via a Highest Posterior Density Intervals. *Statistics in Medicine* **44,** 143 – 154.

7. Joseph, L., du Berger, R., and Bélisle, P. (1997) Bayesian and Mixed Bayesian/Likelihood Criteria for Sample Size Determination. *Statistics in Medicine* **16,** 769–781.

8. Kikuchi, T. Pezeshk, H. and Gittins, J. C. (2008) A Bayesian Cost-Benefit Approach to the Determination of Sample Size in Clinical Trials *Statistics in Medicine* **27,** 68–82.

9. Lindley, D. V. (1997) The Choice of Sample Size. *The Statistician* **46,** 129–138.

10. Pezeshk, H. (2006) The Choice of Sample Size; A Fully Bayesian Approach. *The Proceeding of the* 8^{th} *Iranian Statistical Conference.* Shiraz University, August 22 - 24, pp 216–225.

11. Pezeshk, H. and Gittins, J. C. (2002) A Fully Bayesian Approach to Calculating Sample Sizes for Clinical Trials with Binary Responses. *Drug Information Journal* **36,** 143–150.

12. Pezeshk, H. and Gittins, J. C. (2006a) A Bayesian Approach to the Determination of the Number of Subsequent Users of a New Treatment. *Statistical Method in Medical Research* **15,** 585–592.

13. Pezeshk, H. and Gittins, J. C. (2006b) A Cost-Benefit Approach to the Amount of Experimentation of Clinical Trials *Drug Information Journal* **40,** 407–411.

14. Pham–Gia, T. and Turkkan, N. (1992) Sample Size Determination in Bayesian Analysis. *The Statistician* **41,** 389 – 392.

15. Pham–Gia, T. and Turkkan, N. (2003) Determination of Exact Sample Sizes in the Bayesian Estimation of the Difference of two Proportions. *The Statistician* **52,** 131 – 150.

16. Spiegelhalter, D. J., Freeman, L. S. and Parmar, M. K. B. (1994) Bayesian Approaches to Randomized Trials. *J. R. Statist. Soc. A* **157,** 357 – 416.

17. Willan, A. R. (2007) Clinical Decision Making and Expected Value of Information. *Clinical Trials* **4,** 279 – 285.

Comparing Quantile Regression and Linear Regression to Analysis the Risk Factors of Reflux Duration

Mohamad Amin Pourhoseingholi, Soghrat Faghihzadeh, Afsaneh Zarghi, Manijeh Habibi, Azadeh Safaee, Fatemeh Qafarnejad , Mohammad Reza Zali

Research Center for Gastroenterology and Liver Diseases, Shaheed Beheshti Medical University

Abstract: Multiple linear regression allows researchers to assess how the mean of a conditional distribution changes with respect to patient characteristics But in most cases of interest, the biometrical measurements like as duration data are not normally distributed and this leads to unreliable results. Quantile regression is an econometric regression model in which a specified conditional quantile of the outcome variable is expressed as a linear function of subject characteristics and allows researcher to assess how any quantile of a conditional distribution changes with patient characteristics.

Our goal of this study was to determine the associations between reflux duration and demographic characteristics by quantile regression and compare this model to linear regression and Cox semi parametric model.

The results have demonstrated that using quantile regression leads to better interpretation and richer inference about association between reflux duration and patient characteristics.

Keywords: quantile regression, regression models, reflux duration.

1 Introduction

Gastroesophageal reflux disease (GERD) is currently one of the most prevalent gastrointestinal disorders (Liang and Chen (2005)). Symptoms associated with gastroesophageal reflux disease (GERD) are common, with an incidence of approximately 20% in the general population (Rosaida and Goh (2004) and El-Serag and Petersen (2004)). It is often detrimental for patient quality of life and can develop into

oesophagitis or cause other complications (Dimenas (1993) and Lagergren et al.(1999)) Statistical tools such as linear regression are widely used to assess the associated factors on reflux severity and frequency. Multiple linear regression allows researchers to assess how the mean of a conditional distribution changes with respect to patient characteristics and Cox regression allows one to assess the impact of patient characteristics on the relative instantaneous hazard of an event. But in most cases of interest, the biometrical measurements like as duration data are not normally distributed, this data depends on time so parametric model such as linear regression that needs normality assumption can not informative enough (Green and Silverman (1994)). In this situation quantiles are more favorable to picture the distribution of response variable. Quantile regression (Koenker and Bassett (1987) and Koenker and Hallock(2001)) is an econometric regression model in which a specified conditional quantile (or percentile) of the outcome variable is expressed as a linear function of subject characteristics. This is in contrast to linear regression, that the mean of a continuous response is expressed as a linear function of a set of independent or predictor variables. This model is an important tool for estimating conditional quantile models that has been used in many empirical studies and has been studied extensively in theoretical statistics; One of quantile regression's most appealing features is its ability to estimate quantile-specific effects that describe the impact of covariates not only on the center but also on the tails of the outcome distribution. Therefore this technique is capable of providing a more complete statistical analysis of the stochastic relationships among random variables. Not only this model frequently used in the econometrics literature (Levin (2001) and Buchinsky (2004)) but also recently in the field of biostatistics, it is used to find conditional quantile functions to provide a more complete view (Marcotte and Wilcox-Gok (2003), Austin and Schull (2003) and Austin et al.(2005)).
Our goal of this study was to determine the associations between reflux duration and demographic characteristics by quantile regression and compare this model to linear regression and Cox semi parametric

model.

2 Methods

2.1 Quantile regressiom analysis

Duration data are commonly used in applied econometrics and biometrics. There is a variety of readily available estimators for popular models such as the accelerated failure time model and the proportional hazard model (Cox (1973) and Koenker and Bassett (1978)) introduced quantile regressions as a regression based method to model the quantiles of the response variable conditional on the covariates (Koenker and Bassett (1987)). Quantile regression is recently emerging as an attractive alternative to these popular models (Koenker and Bilias (2001) and Koenker and Geling (2001)). By modeling the distribution of the duration in a flexible semiparametric way, quantile regression does not impose modeling assumptions that may not be hold.

This technique is intended to estimate, and conduct inference about conditional quantile functions. Let a regression model be denoted as follows:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \varepsilon_i. \tag{1}$$

In least square technique, the coefficients are estimated so as to minimize the sum of squared deviations from the regression lines.

The pth quantile of the distribution of a random variable y is a number q such $P(y < q) = p$.

In quantile regression, the pth quantile of the conditional distribution can be modeled as a linear function of subject characteristics, One defines the following function:

$$\rho_i = \begin{cases} p & if \quad e_i > 0 \\ (1-p) & otherwise \end{cases} \tag{2}$$

The regression coefficient are estimated with respect to minimize the following function

$$\sum_{i=1}^{n} |e_i| o_i \tag{3}$$

Where

$$e_i = y_i - (b_0 + b_1 x_{1i} + b_2 x_{2i} + \ldots + b_k x_{ki}) \quad (4)$$

And $(b_0, b_1, \ldots, b_k)$ are the regression parameters.
If someone wants to estimate median regression (Birkes and Dodge (1993)), then ρ_i is equal to 0.5 for all observations.
But there is no closed-form expression to estimation and an iterative algorithm must be used to obtain the coefficient estimates (Koenker and Bassett (1987) and Gilchrist (2003)).
Similarly, the standard errors of the regression coefficients can be estimated using methods described by Koenker and Bassett (Koenker and Bassett (1982)) and elaborated by Rogers (Rogers (1992)) and another technique that recommended estimating the standard errors of regression coefficients for quantile regression models is bootstrap (Gould (1992)) In practical works, quantile regression is similar to linear regression. Some macros have been developed in software such as SAS (PROC QUANTREG) and R (package quantreg) to fit the models of interest and researcher also can use either forward or backward model selection techniques similar to those available for multiple linear regression. Quantile regression models can include interaction terms like linear regression and the coefficients are interpreted in the same way.

2.2 Cox Proportional Hazards Model

Cox's model (Cox (1972)) has become the most used procedure for modeling the relationship of covariates to a survival or other time-to-event data (Therneau and Grambsch (2000)).
Cox proposed a semi-parametric model for the hazard function that allows the addition of explanatory variables, or covariates, but keeps the baseline hazard as an arbitrary, unspecified, nonnegative functional of time. The Cox hazard function for fixed-time covariates, x, is

Due to the construction of equation above, the baseline hazard $\lambda_0(t)$is defined as the hazard function for that individual with zero on all covariates. Because the baseline hazard is not assumed to be of a parametric form, Cox's model is referred to as a semi-parametric

model for the hazard function.
However, it has some restrictions. One of the restrictions to using the Cox model with time-fixed covariates is its proportional hazards assumption; it means the hazard ratio between two sets of covariates is constant over time. This is due to the common baseline hazard function canceling out in the ratio of the two hazards.

2.3 Reflux duration risk factors

A door to door questionnaire have been used to evaluate gastrointestinal symptoms contain reflux and its history (Zarghi et al. (2007)). The questionnaire has been completed through the cross sectional study that conducted during May and December 2006 in Tehran. Approximately 6000 families with nearly 24000 members reside. Study subjects were interviewed by using questionnaire. The questionnaire included socio demographic, health relevant life style and clinical factors and GI symptoms. Those who had any GI symptom or sign were referred to a general practitioner for further evaluation and Functional Gastrointestinal questionnaire were completed. We selected 790 adults' gastrointestinal patients using random sampling method that 208 numbers were found with reflux symptoms based on room III. Demographic characteristics contain age, sex, body mass index (BMI), feeling of depression and stress, smoking and the duration of reflux during the previous six months (in week) were reported by means of validated questionnaires and a quantile regression was performed to find the results. Linear regression and Cox model were used to compare the efficacy of models with quantile regression. We conducted all analyses with the use of SAS software (version 9.1).

3 Results

3.1 Baseline data

A total of 208 patients with reflux symptom entered to this study. The mean of age is 52.5 ± 17.8 and the median is 51; 72.2 per cent of these patients are female, 46.5 per cent have pain in sever stage,

35.6 per cent reported pain after eating, 53.8 per cent viewed a kind of pain after having some specific food, 38.9 per cent have feeling of depression, 65.9 per cent have some kind of stress and 7.7 per cent are smokers.
The mean of reflux duration based on number of weeks reported by patients is 16.4 ± 9.3 and the median is 24. Figure 1 shows the distribution of reflux duration and it clearly indicates a dramatic high negative skewness equals -0.58 with a large group of patients who tend to experience symptom more than 15 weeks. This negative skewness is also high for female (-0.65), pain in sever stage (-0.58), pain after eating (-1.03), pain after having some specific food (-0.59), feeling of depression (-1.4), stress (-0.69) and smoking (-1.03).

3.2 Linear regression

An ordinary linear regression has been conducted with both untransformed and log-transformed reflux duration. Final model based on backward selection in regression with untransformed response produced significant coefficients for pain after eating, feeling of depression, smoking and age and in regression with log-transformed response produced these significant coefficients for this variables again except smoking (Table 1). These results indicate that the mean of reflux duration for patients who have a kind of pain after eating is 2.81 week longer than others in linear regression with untransformed and 28.4 pre cent longer in regression with log-transformed ($\exp(0.25)=1.28$). Feeling of depression is other risk factors based on results in both regression models. The mean of reflux duration in regression with untransformed response is 4.79 week longer for someone who feels depression and it is 60 per cent longer in regression with log-transformed response.
Age has a little effect on the duration of reflux. From the linear model with untransformed response it seems that the mean of reflux duration increases 0.08 week when the age increases one year. In log-transformed regression the increasing is just 0.7 per cent. Smoking is only significance in regression with untransformed response. The

mean of reflux duration is 4.4 weeks more than the patients who did not smoke. Although the results clinically favorable, the goodness of fit are low for both transformed and untransformed ones. The R square for both models is lower than 0.13 and this case obviously is the result of sever skewness in reflux duration, which may lead to misleading inference about the parameters of interest.

3.3 Cox Proportional Hazards Regression

The proportional hazard regression model produced no significant coefficient neither in full model nor in final model based on backward selection. In final model only pain after eating remains with insignificant P-value equals 0.2.

3.4 Quantile Regression

The results from quantile regression analysis are contained in table 2. estimated coefficients for sex, pain severity and pain after having some specific food describe the increase in the given decile of reflux duration associated with a unit increase in given patient characteristic. The length of reflux duration for each decile (except sex where the first and second given decile are not significance) in male are longer than that similar proportion of female patients and it increases significantly in expected decile for patients who experienced sever pain and pain after eating some specific food in compare to the others considerably ($P<0.001$ for all deciles). The results for feeling of stress and BMI seem skeptical. Although their estimated coefficients are negative, the amount values are not clinically important. Therefore we can ignore these doubtful results. In addition all significant variables in linear and log-linear regression have no significant effect on any decile in quantile regression.

4 Discussion

Researchers have traditionally used either OLS regression or proportional hazard model to evaluate the impact of patient characteristics

on data duration like as waiting times for receiving medical treatment (GUSTO (1993)) or length of stay in hospital (Yoon et al. (2003)). OLS seeks to model the mean of a random response as a function of observed explanatory factors; quantile regression seeks to model this function for each given quantile separately. By modeling the distribution of the duration in a flexible semiparametric way, quantile regression does not impose modeling assumptions that may not be hold.

Although quantile regression is frequently used in econometric (Levin (2001) and Buchinsky (1994)) , in the field of medicine where some duration data tends to concentrate on first or last quantiles researchers recently are interested in using this model. Our study has demonstrated that the use of quantile regression allows for richer inferences to be drawn a picture of impact of patient characteristics on reflux duration. In linear regression model with transformed and untransformed response, study fund pain after eating, depression, smoking and age significant. However some psychosomatic symptom such as depression (Siupsinskiene et al. (2007)) and lifestyle like smoking (Lu et al. (2005)) are reported as the risk factors for reflux, the diagnostics of regression models indicated low reliability in modeling because of sever skewness in response. Therefore this result seems unreliable. In contrast to linear regression models, quantile regression fund a significant relation between sex, pain severity and having pain after some specific foods with duration of reflux. Age and BMI are fund insignificant. However the relationship between GERD and age is controversial, some studies have showed a direct relationship (Isolauri and Laippala (1995)), others an inverse relationship (Drossman et al. (1993)) and others still no association whatsoever (Ho et al. (1998)). Like age, in the some study results for BMI address the association between BMI and gastroesophageal reflux disease (Nilsson et al. (2003) and Kulig et al. (2004)) and in others have found no significant relationship (Andersen and Jensen (1991)).

Cox proportional hazard showed no significant association for any factors entered to the model. This misleading is maybe due to the fact

that the Cox model imposes a fully parametric relationship between the relative hazard of duration time and the event, while quantile regression dose not impose any such relationship. Censored quantile regression discusses this relationship (Portnoy (2003)) but in our study we have no assumption about censoring. In conclusion, we have

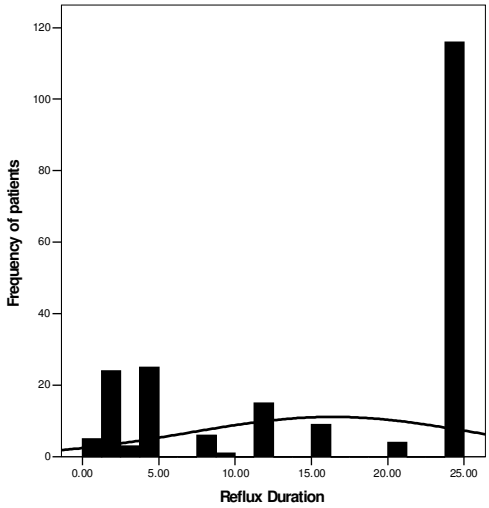


Figure 1. Distribution of reflux duration in patient with GERD symptom

demonstrated that using quantile regression allows researchers to find the impact of patient characteristic on the time duration response. The linear regression provides at best case an incomplete picture of the association between time response and characteristic factors and leads to unreliable results in the case of sever skewness. Cox proportional hazard dose not allow one to characterize the distribution of time to factor of interest and can be modeled in unexpected results due to its assumption. As a result researchers should be interested in quantile regression's advantages to evaluate and provide a complete view of changing in time response distribution with patient factors and characteristics.

Acknowledgments We would like to thank Health Center of Firoozkouh city and Damavand city for data gathering and Research Center of Gastroenterology and Liver Diseases for their valuable collaboration in this study.

References

[1] Liang H., Chen J.D. Assessment of the esophageal pressure in gastroesophageal reflux disease by the local regression Ann Biomed Eng. 2005 Jun; 33(6):847-53.

[2] Rosaida MS, Goh KL. Gastroesophageal reflux disease, reflux oesophagitis and non-erosive reflux disease in a multiracial Asian population: a prospective, endoscopy based study. Eur. J. Gastroenterol. Hepatol. 2004; 16: 495-501.

[3] El-Serag H.B., Petersen N.J., Carter J. et al. Gastroesophageal reflux among different racial groups in the United States. Gastroenterology 2004; 126: 1692-9.

[4] Dimenas E. Methodological aspects of evaluation of quality of life in upper gastrointestinal disease. Scand J Gastroenterol 1993; 28(Suppl. 199): 18-21.

[5] Lagergren J., Bergstrom R., Lindgren A., Nyren O. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. N Engl J Med 1999; 340: 825-31.

[6] Green P.J., Silverman B.W. nonparametric regression and generalized linear models Chapman & Hall London 1994.

[7] Koenker, R. and Bassett G. Regression quantiles. Econometrica, 1987; 46, 33-50.

[8] Koenker R., Hallock, K. (2001), Quantile Regression. Journal of Economic Perspectives, 2001; Vol. 51, No. 4, 143-156.

[9] Levin J. For whom the reductions count: a quantile regression analysis of class size and peer effects on scholastic achievement. Empirical Economics 2001; 26:221-246.

[10] Buchinsky M. Changes in the U.S. wage structure 1963-1987: application of quantile regression. Econometrica 1994; 62:405-458.

[11] Marcotte D.E., Wilcox-Gok V. J Estimating earnings losses due to mental illness: a quantile regression approach. Ment Health Policy Econ 2003; 6(3):123-34.

[12] Austin P.C., Schull M.J. Quantile regression: a statistical tool for out-of-hospital research. Acad Emerg Med. 2003; 10(7):789-97.

[13] Austin P.C., Tu J.V., Daly P.A. and Alter D.A. The use of quantile regression in health care research: a case study examining gender differences in the timeliness of thrombolytic therapy. Statist. Med. 2005; 24: 791-816.

[14] Cox DR. Regression models and life-table. (With discussion). Journal of Royal Statistical Society B 1972; 34: 187-220.

[15] Koenker, R. Bilias, Y. Quantile Regression for Duration Data: A Reappraisal of the Pennsylvania Reemployment Bonus Experiments. Empirical Economics, 2001; 26, 199-220.

[16] Koenker, R., Geling O. Reappraising Medfly Longevity: A Quantile Regression Survival Analysis. Journal of the American Statistical Association, 2001; 96(454), 458-468.

[17] Birkes D., Dodge Y. Alternative Methods of Regression. Wiley, Inc.: New York, NY, 1993.

[18] Gilchrist WG. Statistical Modelling with Quantile Functions. Chapman & Hall/CRC: Boca Raton, FL, 2000.

[19] Koenker R., Bassett Jr. G. Robust tests for heteroscedasticity based on regression quantiles. Econometrica 1982; 50:43-61.

[20] Rogers W.H. Sg11: quantile regression standard errors. Stata Technical Bulletin 1992; 9:16-19 (Reprinted in Stata Technical Bulletin Reprints 1992; 2:133-137).

[21] Gould W.W. Sg11.1: quantile regression with bootstrapped standard errors. Stata Technical Bulletin 1992; 9: 19-21 (Reprinted in Stata Technical Bulletin Reprints 1992; 2:137-139).

[22] Therneau T., Grambsch, P. Modeling Survival Data: Extending the Cox Model, New York: Springer-Verlag 2000.

[23] Zarghi A., Pourhoseingholi M.A., Habibi M., Rostami Nejad M., Ramezankhani A. and Zali M.R. Prevalence of Gastrointestinal Symptoms in Population of Tehran, Centre of Iran. 5th European Conference on Tropical Medicine and International Health, 2007.

[24] The GUSTO Investigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. New England Journal of Medicine 1993; 329:673-682.

[25] Yoon P., Steiner I., Reinhardt G. Analysis of factors influencing length of stay in the emergency department. CJEM. 2003; 5(3):155-61.

[26] Siupsinskiene N., Adamonis K., Toohill R.J. Quality of life in laryngopharyngeal reflux patients. Laryngoscope 2007; 117(3):480-4.

[27] Lu C.L., Lang H.C., Chang F.Y., Chen T.J., Chen C.Y., Luo J.C., Lee S.D. Social and medical impact, sleep quality and the pharmaceutical costs of heartburn in Taiwan. Aliment Pharmacol Ther 2005; 22(8):739-47.

[28] Isolauri J., Laippala P. Prevalence of symptoms suggestive of gastro-oesophageal reflux disease in an adult population. Ann Med 1995; 27: 67- 70.

[29] Drossman D.A., Li Z., Andruzzi E., et al. U.S. householder survey of functional gastrointestinal disorders. Prevalence, sociodemography, and health impact. Dig Dis Sci 1993; 38: 1569- 80.

[30] Ho K.Y., Kang J.Y., Seow A. Prevalence of gastrointestinal symptoms in a multiracial Asian population, with particular reference to reflux-type symptoms. Am J Gastroenterol 1998; 93: 1816-22.

[31] Nilsson M., Johnsen R., Ye W., Hveem K., Lagergren J. Obesity and estrogen as risk factors for gastroesophageal reflux symptoms. JAMA 2003; 290:66-72.

[32] Kulig M., Nocon M., Vieth M., et al. Risk factors of gastroesophageal reflux disease: methodology and first epidemiological results of the ProGERD study. J Clin Epidemiol 2004; 57:580-589.

[33] Andersen L.I., Jensen G. Risk factors for benign oesophageal disease in a random population sample. J Intern Med 1991; 230:5-10.

[34] Portnoy S. Censored regression quantiles. Journal of the American Statistical Association 2003; 98:1001-1012.

Variable	Linear Regression		Linear Regression on log transformed		Proportional Hazards Regression	
	Full Model	Final Model	Full Model	Final Model	Full Model	Final Model
Sex	-0.45(0.819)		-0.07(0.713)		0.06(0.805)	
Pain Severity	-0.45(0.762)		-0.05(0.722)		0.009(0.906)	
Pain after eating	2.74(0.084)	2.81(0.066)	0.25(0.116)	0.25(0.099)	0.19(0.299)	0.22(0.200)
Pain after some food	-0.28(0.851)		-0.03(0.839)		-0.01(0.932)	
Feeling of Depression	4.98(0.006)	4.79(0.001)	0.54(0.003)	0.47(0.001)	0.21(0.306)	
Feeling of Stress	-0.32(0.867)		-0.05(0.774)		0.8(0.743)	
Smoking	4.63(0.119)	4.40(0.094)	0.39(0.193)		0.28(0.429)	
BMI	-0.07(0.587)		-0.003(0.835)		0.006(0.669)	
Age	0.08(0.038)	0.08(0.031)	0.008(0.043)	0.007(0.058)	-0.005(0.344)	

Estimated coefficients and p-value in linear regression and proportional hazard regression, each cell contains the estimated coefficient and (P-value) associated with the given variable and given mean response

Variable	Decile								
	1	2	3	4	5	6	7	8	9
Sex	0.01 (0.382)	0.02 (0.499)	0.094 (< 0.001)	0.13 (< 0.001)	0.14 (< 0.001)	0.14 (< 0.001)	0.15 (< 0.001)	0.14 (< 0.001)	0.15 (< 0.001)
Pain Severity	0.32 (< 0.001)	0.34 (< 0.001)	0.395 (< 0.001)	0.40 (< 0.001)	0.41 (< 0.001)	0.42 (< 0.001)	0.45 (< 0.001)	0.46 (< 0.001)	0.51 (< 0.001)
Pain after eating	0.07 (0.204)	0.07 (0.200)	0.071 (0.115)	0.03 (0.339)	0.02 (0.471)	0.03 (0.414)	0.02 (0.637)	0.02 (0.644)	0.01 (0.799)
Pain after some food	0.16 (< 0.001)	0.22 (< 0.001)	0.259 (< 0.001)	0.27 (< 0.001)	0.29 (< 0.001)	0.28 (< 0.001)	0.34 (< 0.001)	0.40 (< 0.001)	0.43 (< 0.001)
Feeling of Depression	-0.01 (0.715)	-0.01 (0.570)	0.004 (0.819)	0.004 (0.715)	0.001 (0.933)	-0.0007 (0.945)	-0.003 (0.862)	-0.01 (0.382)	-0.01 (0.462)
Feeling of Stress	-0.09 (0.004)	-0.09 (0.008)	-0.088 (0.002)	-0.08 (< 0.001)	-0.08 (< 0.001)	-0.09 (< 0.001)	-0.09 (< 0.001)	-0.09 (< 0.001)	-0.09 (< 0.001)
Smoking	0.003 (0.923)	0.009 (0.743)	0.022 (0.474)	0.02 (0.341)	0.02 (0.207)	0.02 (0.203)	0.02 (0.318)	0.003 (0.851)	0.000 (0.999)
BMI	-0.04 (0.056)	-0.05 (0.007)	-0.047 (0.001)	-0.05 (< 0.001)	-0.06 (< 0.001)	-0.06 (< 0.001)	-0.07 (< 0.001)	-0.07 (< 0.001)	-0.07 (< 0.001)
Age	0.005 (0.775)	0.005 (0.910)	0.045 (0.318)	0.02 (0.464)	0.01 (0.609)	0.005 (0.809)	-0.001 (0.945)	0.003 (0.897)	-0.005 (0.886)

Quantile regression coefficients and p-value for the nine Decile of reflux duration, each cell contains the estimated coefficient (P-value) associated with the given variable and given quantile regression model.

A note on the mean past and the mean residual life of a $(n-k+1)-out-of-n$ system under multi monitoring

M. H. Poursaeed

Department of Mathematics, University of Lorestan

Abstract: In the study of the reliability of technical systems, $k-out-of-n$ systems play an important role. In the present paper, we consider a $(n-k+1)-out-of-n$ system consisting of n identical components such that the lifetimes of components are independent and have a common distribution function F. It is assumed that the number of monitoring is l and the total number of failures of the components at time t_i is m_i, $i=1,...,l-1$. Also at time t_l $(t_1<...<t_l)$ the system have failed or the system is still working. Under these conditions, the mean past lifetime (MPL), the mean residual lifetime (MRL) of system and their properties are investigated.

Keywords: Markovian's property, Order statistics, Reliability function, Reversed hazard rate, Truncated expectation.

1 Introduction

The $(n-k+1)-out-of-n$ system structure is a very popular type of redundancy in technical systems. A $(n-k+1)-out-of-n$ system consisting of n components functions if and only if at least $n-k+1$ out of n components are operating. Two important particular cases of $(n-k+1)-out-of-n$ systems are parallel systems and series systems corresponding to $k=n$ and $k=1$, respectively. Let $T_1,\cdots,T_n$ denote the lifetimes of the components of the system and assume that the $T_i's, i=1,\cdots,n$ are independent and have a common distribution F. Also let $T_{1:n},\cdots,T_{n:n}$ be the order statistics corresponding to $T_i's$.

The MPL and the MRL of T at time t, which we denote by $M(t)$ and $m(t)$, respectively, are defined as: $M(t)=E(t-T|T<t)=\frac{\int_0^t F(x)dx}{F(t)}$ and $m(t)=E(T-t|T>t)=\frac{\int_t^\infty \overline{F}(x)dx}{\overline{F}(t)}$, provided that $0<F(t)<1$, where f and $\overline{F}=1-F$ denote the density function and the

reliability function, respectively. Many properties and applications of these measures are obtained in the literature (we refer to Kotz and Shanbhag (1980) and Guess and Proschan (1988) and references there in).

The MPL and the MRL for the $(n-k+1)-out-of-n$ system described above, are defined to be $E(t-T_{k:n}|T_{k:n}\leq t)$ and $E(T_{k:n}-t|T_{k:n}\geq t)$, respectively. Asadi (2006) has considered the structure of parallel systems and has introduced the measure

$$E(t-T_{r:n}|T_{n:n}\leq t),\quad r=1,2,\cdots,n, \tag{1}$$

as the mean past lifetime of components of the system. Tavangar and Asadi (2007) studied a $(n-k+1)$-out-of-n system with identical components where the independence of component's lifetime and having a common distribution function is assumed. They considered

$$E(t-T_{r:n}|T_{k:n}\leq t),\quad r=1,2,\cdots,k, \tag{2}$$

and obtained some of its properties.

A new definition for the mean residual life function of a parallel system is proposed by Asadi and Bairamov (2005), such that $T_{k:n}>t$, i.e., at least $(n-k+1)$, $k=1,2,\cdots,n$ components of the system are still working. Indeed, they have proposed the following measure

$$E(T_{n:n}-t|T_{k:n}\geq t),\quad k=1,2,\cdots,n, \tag{3}$$

and obtained several properties of it (see also Bairamov *et al.*, 2002).

Recently, Poursaeed and Nematollahi (2008), considered a parallel system such that the total number of failures at time t_1 is m and at time t_2 $(t_1<t_2)$ all components of the system have failed or the system is still working. Under these conditions, they studied the MPL and the MRL of system.

In this paper, under the similar condition, we studied the MPL and the MRL of a $(n-k+1)-out-of-n$ system, as the number of monitoring is l, i.e.,

$$M_{n,k}^{l,m_{l-1}}(t_{l-1},t_l)=E(t_l-T_{k:n}|T_{m_1:n}<t_1<T_{m_1+1:n},...,T_{m_{l-1}:n}<t_{l-1}<T_{m_{l-1}+1:n},T_{k:n}<t_l) \tag{4}$$

and

$$W_{n,k}^{l,m_{l-1}}(t_{l-1},t_l) = E(T_{k:n} - t_l | T_{m_1:n} < t_1 < T_{m_1+1:n}, ..., T_{m_{l-1}:n} < t_{l-1} < T_{m_{l-1}+1:n}, T_{k:n} > t_l). \quad (5)$$

Note that in special case, when $k = n$ and $l = 2$, the definition of $M_{n,k}^{l,m_{l-1}}(t_{l-1},t_l)$ and $W_{n,k}^{l,m_{l-1}}(t_{l-1},t_l)$ reduces to what was studied by Poursaeed and Nematollahi (2008). Indeed, the present paper has generalized some of the existing results in the literature.

This paper was organized as follows: In section 2, the MPL of the $(n-k+1)-out-of-n$ systems under multi monitoring,$M_{n,k}^{l,m_{l-1}}(t_{l-1},t_l)$, was obtained. In section 3, it was shown that MPL is a decreasing function of k, also is an increasing function of m_{l-1} and n. In addition, the behavior of it with respect to t_{l-1} and t_l was studied. In sections 4 and 5, the MRL function, i.e., $W_{n,k}^{l,m_{l-1}}(t_{l-1},t_l)$ and some of its properties were also investigated.
Since the functional form of $M_{n,k}^{l,m_{l-1}}(t_{l-1},t_l)$ compared to $M_{n,k}^{2,m_1}(t_1,t_2)$ and $W_{n,k}^{l,m_{l-1}}(t_{l-1},t_l)$ compared to $W_{n,k}^{2,m_1}(t_1,t_2)$ are similar, therefore in the next chapters, at first the case of double monitoring ($l = 2$) was considered. This can be easily generalized to the case that the number of monitoring is l times.

2 The MPL of $(n-k+1)-out-of-n$ systems under multi monitoring

Let $M_{n,k}^{2,m}(t_1,t_2)$ be the MPL function of a $(n-k+1)-out-of-n$ system under double monitoring. It is assumed that the lifetimes of the components are independent and identically distributed with common distribution F. In the following we obtain the functional form of MPL.

$$
\begin{aligned}
& P(t_2 - T_{k:n} > s, T_{m:n} < t_1 < T_{m+1:n}, T_{k:n} < t_2) \\
= & \int_{t_1}^{t_2-s} \int_{t_1}^{z} \int_{0}^{t_1} \frac{n!}{(m-1)!(k-m-2)!(n-k)!} \\
& \times (F(x))^{m-1} (F(z) - F(y))^{k-m-2} (1 - F(z))^{n-k} \\
& \times dF(x) dF(y) dF(z) \\
= & \ ... = \binom{n}{m} (F(t_1))^m \sum_{i=k-m}^{n-m} \binom{n-m}{i} (F(t_2 - s) - F(t_1))^i \\
& \times (1 - F(t_2 - s))^{n-m-i},
\end{aligned}
$$

and so,

$$
\begin{aligned}
& P(t_2 - T_{k:n} > s | T_{m:n} < t_1 < T_{m+1:n}, T_{k:n} < t_2) \\
= & \frac{\sum_{i=k-m}^{n-m} \binom{n-m}{i} (F(t_2 - s) - F(t_1))^i (1 - F(t_2 - s))^{n-m-i}}{\sum_{i=k-m}^{n-m} \binom{n-m}{i} (F(t_2) - F(t_1))^i (1 - F(t_2))^{n-m-i}}, \quad (6)
\end{aligned}
$$

when $0 < s < t_2 - t_1$.

Note that to obtain (6), it is not actually needed to restrict the support of F on $(0, \infty)$. In general, $M_{n,k}^{2,m}(t_1, t_2)$ can be defined for the distribution functions with left extremity $a > 0$ and right extremity $b < \infty$, respectively.

Remark 2.1. If

$$
\begin{aligned}
& P(t_2 - T_{k:n} > s | T_{0:n} < t_1 < T_{1:n}, T_{k:n} < t_2) \\
& := P(t_2 - T_{k:n} > s | t_1 < T_{1:n} < T_{k:n} < t_2)
\end{aligned}
$$

the expression of (6) will be valid for $m = 0$.

Remark 2.2. Interestingly, note that the conditional probability given by (6) can also be represented as follows:

$$
P(t_2 - T_{k:n} > s | T_{m:n} < t_1 < T_{m+1:n}, T_{k:n} < t_2) = \frac{P(Y_{t_1, t_2 - s}^{n-m} \geq k - m)}{P(Z_{t_1, t_2}^{n-m} \geq k - m)}, \quad (7)
$$

where Y_{t_1,t_2-s}^{n-m} and Z_{t_1,t_2}^{n-m} are distributed as binomial with parameters $(n-m, \Theta_{t_1}(t_2-s))$ and $(n-m, \Theta_{t_1}(t_2))$, respectively such that $\Theta_{t_1}(.) = 1-\frac{\bar{F}(.)}{\bar{F}(t_1)}$, which is denoted by $Y_{t_1,t_2-s}^{n-m} \sim Bin(n-m, \Theta_{t_1}(t_2-s))$ and $Z_{t_1,t_2}^{n-m} \sim Bin(n-m, \Theta_{t_1}(t_2))$.

According to remark 2.2, the MPL of system for $k = m+1, \cdots, n$ and $0 \leq t_1 < t_2 < \infty$ can be stated as:

$$M_{n,k}^{2,m}(t_1,t_2) = \int_0^{t_2-t_1} \frac{P(Y_{t_1,t_2-s}^{n-m} \geq k-m)}{P(Z_{t_1,t_2}^{n-m} \geq k-m)} ds = \int_{t_1}^{t_2} \frac{P(Y_{t_1,s}^{n-m} \geq k-m)}{P(Z_{t_1,t_2}^{n-m} \geq k-m)} ds \tag{8}$$

Note that the special case of (8), i.e., when $m = 0$ and $t_1 = 0$, is studied by Tavangar and Asadi (2007).

3 Some properties of $M_{n,k}^{2,m}(t_1,t_2)$

The following theorem shows the behavior of $M_{n,k}^{2,m}(t_1,t_2)$ in terms of n, m and k.

Theorem 3.1.
a) For fixed k and m, $M_{n,k}^{2,m}(t_1,t_2)$ is an increasing function of n.
b) For fixed m and n, $M_{n,k}^{2,m}(t_1,t_2)$ is a decreasing function of k.
c) For fixed n and k, $M_{n,k}^{2,m}(t_1,t_2)$ is an increasing function of m.
Proof:

a) This follows from noting that

$$\begin{aligned}
&\frac{\sum_{i=k-m}^{(n+1)-m}\binom{(n+1)-m}{i}(\theta_{t_1}(u))^i(1-\theta_{t_1}(u))^{(n+1)-m-i}}{\sum_{i=k-m}^{(n+1)-m}\binom{(n+1)-m}{i}(\theta_{t_1}(t_2))^i(1-\theta_{t_1}(t_2))^{(n+1)-m-i}}\\
&-\frac{\sum_{i=k-m}^{n-m}\binom{n-m}{i}(\theta_{t_1}(u))^i(1-\theta_{t_1}(u))^{n-m-i}}{\sum_{i=k-m}^{n-m}\binom{n-m}{i}(\theta_{t_1}(t_2))^i(1-\theta_{t_1}(t_2))^{n-m-i}}\\
=\;& [(\theta_{t_1}(u))^{k-m}(1-\theta_{t_1}(u))^{n-k+1}\\
&\sum_{i=k-m}^{n-m}\binom{n-m}{i}(\theta_{t_1}(t_2))^i(1-\theta_{t_1}(t_2))^{n-m-i}-(\theta_{t_1}(t_2))^{k-m}\\
&(1-\theta_{t_1}(t_2))^{n-k+1}\sum_{i=k-m}^{n-m}\binom{n-m}{i}(\theta_{t_1}(u))^i(1-\theta_{t_1}(u))^{n-m-i}]\\
&/[(\sum_{i=1}^{(n+1)-m}\binom{(n+1)-m}{i}(\theta_{t_1}(t_2))^i(1-\theta_{t_1}(t_2))^{(n+1)-m-i})\\
&\times(\sum_{i=1}^{n-m}\binom{n-m}{i}(\theta_{t_1}(t_2))^i(1-\theta_{t_1}(t_2))^{n-m-i})],
\end{aligned}$$

which the numerator of the above expression equals to:

$$\begin{aligned}
&\{(\theta_{t_1}(u))^{k-m}(1-\theta_{t_1}(u))^{n-k+1}(\theta_{t_1}(t_2))^{k-m}(1-\theta_{t_1}(t_2))^{n-k+1}\\
&\times[\frac{1}{1-\theta_{t_1}(t_2)}\sum_{i=k-m}^{n-m}\binom{n-m}{i}(\frac{\theta_{t_1}(t_2)}{1-\theta_{t_1}(t_2)})^{i-(k-m)}\\
&-\frac{1}{1-\theta_{t_1}(u)}\sum_{i=k-m}^{n-m}\binom{n-m}{i}(\frac{\theta_{t_1}(u)}{1-\theta_{t_1}(u)})^{i-(k-m)}]\}. \qquad (9)
\end{aligned}$$

Since $\frac{1}{1-\theta_{t_1}(t_2)} \geq \frac{1}{1-\theta_{t_1}(u)}$ and $\frac{\theta_{t_1}(t_2)}{1-\theta_{t_1}(t_2)} \geq \frac{\theta_{t_1}(u)}{1-\theta_{t_1}(u)}$ for all $u \in (t_1, t_2)$, It is deduced that (9) is positive and consequently, $M_{n,k}^{2,m}(t_1, t_2)$ is an increasing function of n.

b) This follows easily on noting that

$$\begin{aligned}
&\frac{\sum_{i=(k+1)-m}^{n-m}\binom{n-m}{i}(\theta_{t_1}(u))^i(1-\theta_{t_1}(u))^{n-m-i}}{\sum_{i=(k+1)-m}^{n-m}\binom{n-m}{i}(\theta_{t_1}(t_2))^i(1-\theta_{t_1}(t_2))^{n-m-i}}\\
&-\frac{\sum_{i=k-m}^{n-m}\binom{n-m}{i}(\theta_{t_1}(u))^i(1-\theta_{t_1}(u))^{n-m-i}}{\sum_{i=k-m}^{n-m}\binom{n-m}{i}(\theta_{t_1}(t_2))^i(1-\theta_{t_1}(t_2))^{n-m-i}}\\
= \;& \binom{n-m}{k-m}(\theta_{t_1}(t_2))^{k-m}(1-\theta_{t_1}(t_2))^{n-k}(\theta_{t_1}(u))^{k-m}(1-\theta_{t_1}(u))^{n-k}\\
&\times\sum_{i=k-m+1}^{n-m}\binom{n-m}{i}[(\frac{\theta_{t_1}(u)}{1-\theta_{t_1}(u)})^{i-(k-m)}-(\frac{\theta_{t_1}(t_2)}{1-\theta_{t_1}(t_2)})^{i-(k-m)}]/\\
&\{\sum_{i=(k+1)-m}^{n-m}\binom{n-m}{i}(\theta_{t_1}(t_2))^i(1-\theta_{t_1}(t_2))^{n-m-i}\\
&\times\sum_{i=k-m}^{n-m}\binom{n-m}{i}(\theta_{t_1}(t_2))^i(1-\theta_{t_1}(t_2))^{n-m-i}\}
\end{aligned}$$

which can be shown that the numerator of the above expression is negative.

c) The result follows from the fact that

$$P(X\geq k-m)=P(Y\geq k-m)+\binom{n-m-1}{k-m-1}p^{k-m}(1-p)^{n-k},$$

where $0<p<1$, $X\sim Bin(n-m,p)$ and $Y\sim Bin(n-m-1,p)$.

Remark 3.2.

$$0\leq M_{n,k}^{2,m}(t_1,t_2)\leq t_2-t_1$$

can be shown by using (7) and (8).

Remark 3.3. We can show that

$$M^{2,m}_{n+1,k+1}(t_1,t_2) \le M^{2,m+1}_{n+1,k+1}(t_1,t_2) = M^{2,m}_{n,k}(t_1,t_2) \le M^{2,m}_{n+1,k-1}(t_1,t_2).$$

Theorem 3.4. $M^{2,m}_{n,k}(t_1,t_2)$ is a decreasing function of t_1.

Proof: It can be shown that the partial derivation of (7) with respect to t_1 is negative. Thus for $t_1^{'} < t_1^{''} < t_2$, we have:

$$\frac{P(Y^{n-m}_{t_1'',t_2-s} \ge k-m)}{P(Z^{n-m}_{t_1'',t_2} \ge k-m)} \le \frac{P(Y^{n-m}_{t_1',t_2-s} \ge k-m)}{P(Z^{n-m}_{t_1',t_2} \ge k-m)}$$

and

$$M^{2,m}_{n,k}(t_1'',t_2) = \int_{t_1''}^{t_2} \frac{P(Y^{n-m}_{t_1'',s} \ge k-m)}{P(Z^{n-m}_{t_1'',t_2} \ge k-m)} ds \le \int_{t_1''}^{t_2} \frac{P(Y^{n-m}_{t_1',s} \ge k-m)}{P(Z^{n-m}_{t_1',t_2} \ge k-m)} ds$$

$$\le \int_{t_1'}^{t_2} \frac{P(Y^{n-m}_{t_1',s} \ge k-m)}{P(Z^{n-m}_{t_1',t_2} \ge k-m)} ds = M^{2,m}_{n,k}(t_1',t_2)$$

where $Y^{n-m}_{t,s} \sim Bin(n-m,\Theta_t(s))$ and $Z^{n-m}_{t,t_2} \sim Bin(n-m,\Theta_t(t_2))$.

Theorem 3.5. When $\frac{f(t)}{F(t)-F(t_1)}$ is a decreasing function of t for $t > t_1$, then $M^{2,m}_{n,k}(t_1,t_2)$ is an increasing function of t_2.

Proof: The numerator of partial derivation of (7) with respect to t_2 is equal to

$$\begin{aligned}
&(n-m)\bar{F}(t_1)(\bar{F}(t_2-s))^{n-k}(\bar{F}(t_2))^{n-k} \\
&\times (F(t_2-s)-F(t_1))^{k-m}(F(t_2)-F(t_1))^{k-m} \\
&\times \{\frac{f(t_2-s)}{F(t_2-s)-F(t_1)} \frac{1}{\bar{F}(t_2)} \sum_{i=k-m}^{n-m} \binom{n-m}{i} (\frac{F(t_2)-F(t_1)}{\bar{F}(t_2)})^{i-(k-m)} \\
&- \frac{f(t_2)}{F(t_2)-F(t_1)} \frac{1}{\bar{F}(t_2-s)} \sum_{i=k-m}^{n-m} \binom{n-m}{i} (\frac{F(t_2-s)-F(t_1)}{\bar{F}(t_2-s)})^{i-(k-m)}\}.
\end{aligned}$$

which can be shown that is positive.
Note, $\frac{f(t)}{F(t)-F(t_1)}$ is the reversed hazard rate of the truncated random variable $T|T \geq t_1$.

Remark 3.6. When $f(t)$ is a decreasing function of t for $t > t_1$, then $\frac{f(t)}{F(t)-F(t_1)}$ will be a decreasing function of t and consequently, $M_{n,k}^{2,m}(t_1, t_2)$ will be an increasing function of t_2. This can be used for the cases $F(t)$ does not have a closed form.

Remark 3.7. Suppose that the number of monitoring is three times (i.e., $l = 3$), the total number of failures at time t_1 and t_2 are m and r. Also at time t_3 ($t_1 < t_2 < t_3$) the system have failed. In this case, for $1 \leq m < m+1 < r < r+1 \leq k \leq n$, $t_1 < t_2 < t_3$ and $0 < s < t_3 - t_2$, can be shown:

$$\begin{aligned} P(t_3 \; - \; & T_{k:n} > s | T_{m:n} < t_1 < T_{m+1:n}, T_{r:n} < t_2 < T_{r+1:n}, T_{k:n} < t_3) \\ = \; & \frac{P(Y_{t_2,t_3-s}^{n-r} \geq k-r)}{P(Z_{t_2,t_3}^{n-r} \geq k-r)}, \end{aligned} \tag{10}$$

where $Y_{t_2,t_3-s}^{n-r} \sim Bin(n-r, \Theta_{t_2}(t_3-s))$, $Z_{t_2,t_3}^{n-r} \sim Bin(n-r, \Theta_{t_2}(t_3))$ and $\Theta_{t_2}(.) = 1 - \frac{\bar{F}(.)}{\bar{F}(t_2)}$.

It can be noted that, (7) and (10) have similar structures. Therefore, it can be generalized to the case that the number of monitoring is l times. Thus, for $1 \leq m_1 \leq m_2 \leq ... \leq m_{l-1} < m_{l-1} + 1 \leq k \leq n$ and $t_1 < ... < t_l$, we have

$$M_{n,k}^{l,m_{l-1}}(t_{l-1}, t_l) = \int_{t_{l-1}}^{t_l} \frac{P(Y_{t_{l-1},s}^{n-m_{l-1}} \geq k - m_{l-1})}{P(Z_{t_{l-1},t_l}^{n-m_{l-1}} \geq k - m_{l-1})} ds \tag{11}$$

The properties of the above function revealed a similar pattern to double monitoring case as well.

4 The MRL of $(n-k+1)-out-of-n$ systems

This section is included at first, studying of the MRL of a $(n-k+1)-out-of-n$ system under double monitoring and then generalizing to the multi monitoring. Therefore, it is assumed that the total number of failures at time t_1 is m and at time t_2 $(t_1 < t_2)$ the system is still working. Under these conditions, the MRL i.e.,

$$W_{n,k}^{2,m}(t_1,t_2) = E(T_{k:n} - t_2 | T_{m:n} < t_1 < T_{m+1:n}, T_{k:n} > t_2),\ 1 \le m < k,$$

is obtained. For $2 \le m+1 < k$ and $0 < s$,

$$\begin{aligned} & P(T_{k:n} - t_2 > s, T_{m:n} < t_1 < T_{m+1:n}, T_{k:n} > t_2) \\ = & \int_{t_2+s}^{\infty}\int_{t_1}^{z}\int_{0}^{t_1} \frac{n!}{(m-1)!(k-m-2)!(n-k)!} \\ & \times (F(x))^{m-1}(F(z)-F(y))^{k-m-2}(1-F(z))^{n-k}dF(x)dF(y)dF(z) \\ = & \binom{n}{m}(F(t_1))^m \sum_{i=0}^{k-m-1}\binom{n-m}{i}(F(t_2+s)-F(t_1))^i \\ & (1-F(t_2+s))^{n-m-i}, \end{aligned}$$

and

$$\begin{aligned} & P(T_{k:n} - t_2 > s | T_{m:n} < t_1 < T_{m+1:n}, T_{k:n} > t_2) \\ = & \frac{\sum_{i=0}^{k-m-1}\binom{n-m}{i}(F(t_2+s)-F(t_1))^i(1-F(t_2+s))^{n-m-i}}{\sum_{i=0}^{k-m-1}\binom{n-m}{i}(F(t_2)-F(t_1))^i(1-F(t_2))^{n-m-i}} \end{aligned} \quad (12)$$

By separate calculation, it can be found that (12) is valid for $k = m+1$.

Remark 4.1. (12) can also be represented as follows:

$$\begin{aligned} P(T_{k:n} &- t_2 > s | T_{m:n} < t_1 < T_{m+1:n}, T_{k:n} > t_2) \\ &= \frac{\sum_{i=0}^{k-m-1} \binom{n-m}{i} (1 - \frac{\bar{F}(t_2+s)}{\bar{F}(t_1)})^i (\frac{\bar{F}(t_2+s)}{\bar{F}(t_1)})^{n-m-i}}{\sum_{i=0}^{k-m-1} \binom{n-m}{i} (1 - \frac{\bar{F}(t_2)}{\bar{F}(t_1)})^i (\frac{\bar{F}(t_2)}{\bar{F}(t_1)})^{n-m-i}} \\ &= \frac{P(Y_{t_1}^{t_2+s} \geq n-k+1)}{P(Z_{t_1}^{t_2} \geq n-k+1)} \end{aligned} \tag{13}$$

where $Y_{t_1}^{t_2+s} \sim Bin\,(n-m, \vartheta_{t_1}(t_2+s))$ and $Z_{t_1}^{t_2} \sim Bin\,(n-m, \vartheta_{t_1}(t_2))$, s.t.

$$\vartheta_{t_1}(.) = \frac{\bar{F}(.)}{\bar{F}(t_1)}.$$

Therefore, for $1 \leq m < k$, we have

$$\begin{aligned} W_{n,k}^{2,m}(t_1, t_2) &= E(T_{k:n} - t_2 | T_{m:n} < t_1 < T_{m+1:n}, T_{k:n} > t_2) \\ &= \int_0^\infty \frac{P(Y_{t_1}^{t_2+s} \geq n-k+1)}{P(Z_{t_1}^{t_2} \geq n-k+1)} ds \\ &= \int_{t_2}^\infty \frac{P(Y_{t_1}^{u} \geq n-k+1)}{P(Z_{t_1}^{t_2} \geq n-k+1)} du \end{aligned} \tag{14}$$

Remark 4.2. Similar to remark 2.1, if

$$E(T_{k:n} - t_2 | T_{0:n} < t_1 < T_{1:n}, T_{k:n} > t_2) := E(T_{k:n} - t_2 | t_1 < T_{1:n}, T_{k:n} > t_2)$$

(i.e., all the failure times are greater than t_1), it is found that (14) is valid for m=0.

Remark 4.3. For $r = 1, ..., m$ the Markovian's property of ordered statistics implies

$$\begin{aligned} &E(t_1 - T_{r:n} | T_{m:n} < t_1 < T_{m+1:n}, T_{k:n} < t_2) \\ &= E(t_1 - T_{r:n} | T_{m:n} < t_1 < T_{m+1:n}, T_{k:n} > t_2) \\ &= \int_0^{t_1} P(Y_{t_1}^u \geq r) du, \end{aligned} \tag{15}$$

where $Y_{t_1}^u \sim Bin(m, \frac{F(u)}{F(t_1)})$.

Note, (15) is similar to the results given by Asadi (2006).

5 Some properties of $W_{n,k}^{2,m}(t_1, t_2)$

In this section, some of important properties of $W_{n,k}^{2,m}(t_1, t_2)$ are studied.

Theorem 5.1.

i) Similar to the theorem 3.1, it can be shown that

a) For fixed k and m, $W_{n,k}^{2,m}(t_1, t_2)$ is an increasing function of n.

b) For fixed n and k, $W_{n,k}^{2,m}(t_1, t_2)$ is a decreasing function of m.

c) For fixed n and m, $W_{n,k}^{2,m}(t_1, t_2)$ is an increasing function of k.

ii) Similar to the theorem 3.4, can be shown that $W_n^{2,m}(t_1, t_2)$ is an increasing function of t_1.

iii) When $f(t)$ is an increasing function of t and $\bar{F}(t) > \frac{n-k}{n-m-1}\bar{F}(t_1)$ for $t > t_1$, then $W_{n,k}^{2,m}(t_1, t_2)$ can be shown as a decreasing function of t_2. Obviously just the first condition (being $f(t)$ as an increasing function of t for $t > t_1$) would be needed for the parallel systems.

iv) It is easy to show that

$$W_{n,k}^{2,m+1}(t_1, t_2) \leq W_{n+1,k}^{2,m}(t_1, t_2) \;=\; W_{n,k-1}^{2,m-1}(t_1, t_2) \leq W_{n+1,k}^{2,m-1}(t_1, t_2).$$

Remark 5.2. If the number of monitoring is three times (i.e., $l = 3$), $1 \leq m < m+1 \leq r < r+1 \leq k$, $t_1 < t_2 < t_3$ and $0 < s$, we will have

$$\begin{aligned} E(T_{k:n} - t_3 | T_{m:n} &< t_1 < T_{m+1:n}, T_{r:n} < t_2 < T_{r+1:n}, T_{k:n} > t_3) \\ &= \int_{t_3}^{\infty} \frac{P(Y_{t_2}^u \geq n-k+1)}{P(Z_{t_2}^{t_3} \geq n-k+1)} du \end{aligned} \tag{16}$$

where $Y_{t_2}^u \sim Bin\ (n-r, \vartheta_{t_2}(u))$ and $Z_{t_2}^{t_3} \sim Bin\ (n-r, \vartheta_{t_2}(t_3))$, s.t.

$$\vartheta_{t_2}(.) = \frac{\overline{F}(.)}{\overline{F}(t_2)}.$$

Therefore, the remark 5.2. can be generalized to the case that the number of monitoring is l times. In this case, for $1 \leq m_1 \leq m_2 \leq ... \leq m_{l-1} < m_{l-1}+1 \leq k \leq n$, $t_1 < ... < t_l$ and $0 < s$, we have:

$$\begin{aligned} & W_{n,k}^{l,m_{l-1}}(t_{l-1}, t_l) \\ & = E(T_{k:n} - t_l | T_{m_1:n} < t_1 < T_{m_1+1:n}, ..., T_{m_{l-1}:n} < t_{l-1} \\ & < T_{m_{l-1}+1:n}, T_{k:n} > t_l) = \int_{t_l}^{\infty} \frac{P(Y_{t_{l-1}}^u \geq n-k+1)}{P(Z_{t_{l-1}}^{t_l} \geq n-k+1)} du \quad (17) \end{aligned}$$

where $Y_{t_{l-1}}^u \sim Bin\ (n-m_{l-1}, \vartheta_{t_{l-1}}(u))$ and $Z_{t_{l-1}}^{t_l} \sim Bin\ (n-m_{l-1}, \vartheta_{t_{l-1}}(t_l))$, s.t.

$$\vartheta_{t_{l-1}}(.) = \frac{\overline{F}(.)}{\overline{F}(t_{l-1})}.$$

It can be displayed that all properties of (17) are according to double monitoring case.

References

[1] Asadi, M. (2006). On the mean past lifetime of the components of a parallel system. *Journal of Statistical planning and Inference*, **136**, 1197-1206.

[2] Asadi, M., Bairamov, I., (2005). A note on the mean residual life function of a parallel system. *Communications in Statistics- Theory and Methods*, **34**, 475-484.

[3] Bairamov, I., Ahsanullah, M., Akhundov, I.,(2002). A Residual Life Function of a System having Parallel or Series Structures. *J. Statist. Theory Appl.*, **1**, 119-132.

[4] Guess, F. and Proschan, F. (1988). Mean residual life: Theory and applications. Krishnaiha, P.R. et al., eds., *Handbook of Statistics*, **7**, 215-224.

[5] Kotz, S., Shanbhag, D. N. (1980). Some new approaches to probability distributions. *Adv. Appl. Probab.*, **12**, 903-921.

[6] Poursaeed, M. H., Nematollahi, A. R., (2008). On the Mean Past and the Mean Residual Life Under Double Monitoring. Communications in Statistics- Theory and Methods, **37**, 1119-1133.

[7] Tavangar, M. and Asadi, M., (2007). A study on the mean past lifetime of the components of $n-k+1-out-of-n$ system at the system level. *Submitted for publication.*

A Nonparametric Method for Assessment of Interactions in a Multi-factor Design with Censored Data

M. H. Rahbar[1,*], H. Talebi[2], J. Sangchoon[3]

[1,3] *Michigan State University*
[2] *University of Isfahan*

Abstract: Statistical procedures and methodology for assessment of interventions or treatments based on medical data often involves complexities due to incompleteness of the available data as a result of drop out or the inability of complete follow up until the endpoint of interest. In this article we propose a nonparametric method for assessment of interactions when we are concerned with investigation of the simultaneous effects of two or more factors in a regression model based on censored data. Specifically, we will assess the interaction between a treatment (dose) and a covariate (e.g., age categories) on the mean survival time of subjects assigned to combinations of the levels of these factors when the effect of treatment (dose) is dependent on the covariate (age). The proposed method allows for varying levels of censorship in the outcome among different groups of subjects at different levels of the independent variables (factors). We present the asymptotic distribution of the estimators of the parameters in our model, which then allows for statistical inference. Through a simulation study we assess the effect of the censoring rate on the standard error of these types of estimators and power of our new proposed method with that of using the Cox Proportional Hazard Model. Finally, we will demonstrate real life application of our proposed method, by using data from a nursing intervention trial of cancer patients undergoing chemotherapy.

Keywords: Kaplan-Meier ; Survival Time; Censoring; Regression; Multi-factor design; Interactions; Simulations; Cancer.

1 Introduction

Statistical procedures and methodology for assessment of interventions or treatments based on medical data often involves complexities due to incompleteness of the available data as a result of drop out or the

Invited speaker

inability of complete follow up until the endpoint of interest which results in censored data. A nonparametric estimator of the survival curve based on randomly right censored data was first introduced by Kaplan and Meier (1958). Later, Cox(1972) introduced the Proportional Hazard (PH) Model as a semi-parametric method for assessing covariate effects including interactions from randomly right censored data. Nonparametric estimation of regression coefficients based on right censored data was first addressed by Miller (1976) who obtained estimators by minimizing a weighted sum of squares of the residuals, with weights determined by the Kaplan-Meier estimator of the error distribution. Buckley and James (1979) suggested an alternative estimator of the regression coefficients based on an expectation identity. Both these procedures require iterative methods for computing the estimators. Koul, Susarla, and van Ryzin (1981) introduced a non-iterative estimator of these parameters and derived their asymptotic distribution. More recent work in this area include James and Smith (1984), Leurgens (1987), Rahbar (1990), and Rahbar and Gardiner (1995, 2004).

Rahbar and Gardiner (2004) provided methodology for investigating the simultaneous effects of two or more factors on the mean survival time of subjects in an additive regression model when the covariates have finitely many levels. This model does not consider interactions. For example, in laboratory animal studies designed to assess the effects of varying doses of a carcinogen, the time to some adverse event at dose $d_j$of the carcinogen might be dependent on age categoryc_k. In this situation a non-additive model, which allow assessment of potential interaction between does and age may be more appropriate. In this article we use a regression approach to model the mean survival time of subjects in a treatment when its effect is dependent on the level of covariates. For example, different treatments could correspond to increasing levels of a drug in a non-additive manner. We assume that patient-specific characteristics associated with these treatments can be categorized into a finite number of levels. An additional feature of our model is that the outcome might not be observable in all

subjects. With survival time, this allows for right censoring. We believe development of these nonparametric techniques for assessment of treatments effects while controlling for important covariates and taking into account potential interactions will complement the existing literature for the analysis of censored data. Our approach differs from the other methods referenced here in that it allows varying levels of variance and censorship at different covariate levels. This new method is mainly based on results documented by Rahbar (1990) and Rahbar and Gardiner (1995, 2004).

The remainder of the article is organized as follows. In section 2 we present our regression model, assumptions and estimation scheme for the regression parameters. In section 3 we will present the properties of our estimators, including their asymptotic distribution. In section 4 we report the results of our simulation study. In section 5 we demonstrate application of these techniques using real life data. The last section is devoted to concluding remarks with some discussion.

2 Regression model

Suppose the treatment regimen has $m_1$levels with values $d_1, \ldots, d_{m_1}$and there are finite levels of a covariate, categorized into $m_2$levels with values $c_1, \ldots, c_{m_2}$. Thus there are $m_1 \times m_2$combinations of treatment by the covariate categories represented in our data. Let T_{ijk} denote the response of the i-th subject in the j-th treatment regimen and k-th covariate category. If $T_{ijk}$is a survival time or some time-to-event, it may not be observable in some subjects due to censoring at time U_{ijk}. This restricts observation toX_{ijk}= min(T_{ijk},U_{ijk}) and the indicator δ_{ijk} of the event $[T_{ijk} \leq U_{ijk}]$. In laboratory animal studies designed to assess the effects of varying doses of a carcinogen, T_{ijk}may denote the survival time or time to some adverse event at dose $d_j$of the carcinogen which may be dependent on age category c_k. For assessment of interaction between does and age, we assume an interactive regression model for T_{ijk} given by

$$T_{ijk} = \alpha + \beta_1 d_j + \beta_2 c_k + \gamma d_j c_k + \varepsilon_{ijk}$$

where $\alpha, \beta_1, \beta_2, \gamma$ are unknown parameters and ε_{ijk}an unobservable error. The interaction term in the above model is similar to one degree of freedom interaction effect considered by Tukey (1949). The objective is estimation of these parameters based on n_{jk}observations on the censored sample (X_{ijk},δ_{ijk}), i=1,. . . ,n_{jk}. Throughout we will assume that $(\varepsilon_{ijk}:1 \leq i \leq n_{jk}, 1 \leq j \leq m_1, 1 \leq k \leq m_2)$ are independent identically distributed with zero mean, and continuous distribution $1-F$. The survival time $T_{ijk}$is assumed to have finite support on the interval $[0, \tau]$. Also$\varepsilon_{ijk}$and $U_{ijk}$are assumed to be independent withU_{ijk}having a continuous distribution $1-G_{jk}$.

Let H_{jk}denote the survival distribution of X_{ijk}, that is, $H_{jk}(t)$= $P[X_{ijk} > t]$, and $\tau_{H_{jk}} = \inf\{t : H_{jk}(t) = 0\}$the upper limit of its support. If $\theta_{jk} = \alpha+\beta_1 d_j+\beta_2 c_k+\gamma d_j c_k$, then$H_{jk} = F_{jk}G_{jk}$where$F_{jk}(\cdot) = F(\cdot - \theta_{jk})$is the survival distribution of T_{ijk}.

Estimation procedure

For each $(j,\ k)$, we use the observations $\{X_{ijk},\delta_{ijk},d_j,c_k:1 \leq i \leq n_{jk}\}$ to estimate $\theta_{jk}$by

$$\hat{\theta}_{jk} = \int_0^\infty \hat{F}_{jk}(t)dt \tag{1}$$

where $\hat{F}_{jk}(t)$ is the product-limit estimator of $F_{jk}(t)$ with the value of $\hat{F}_{jk}(t)$ set to zero for $t > X_{jk}^{(n)} = \max\{X_{ijk} : 1 \leq i \leq n_{jk}\}$, (Kaplan and Meier, 1958). Let $\theta_{jk} = \alpha + \beta_1 d_j + \beta_2 c_k + \gamma d_j c_k$be the mean survival time of subjects in thej^{th} treatment and k^{th} covariate level, $1 \leq j \leq m_1$ and $1 \leq k \leq m_2$. For each k,$1 \leq k \leq m_2$ and any pair(j, j'),$1 \leq j, j' \leq m_1$ we have $\theta_{jk} - \theta_{j'k} = \beta_1(d_j - d_{j'}) + \gamma(d_j - d_{j'})c_k$ which leads to $\beta_1(k, j, j') = \frac{\theta_{jk}-\theta_{j'k}}{d_j-d_{j'}} - \gamma c_k$. Similarly, one can calculate an expression for $\beta_2(j, k, k')$ at each $j, 1 \leq j \leq m_1$ and any pair(k, k'),$1 \leq k, k' \leq m_2$. Finally, for any pairs (j, j') and (k, k') of the levels of covariates we have $(\theta_{jk} - \theta_{j'k}) - (\theta_{jk'} - \theta_{j'k'}) = \gamma(d_j - d_{j'})(c_k - c_{k'})$ which suggest an estimator $\hat{\gamma}(j, j', k, k')$ of γgiven by

$$\hat{\gamma}(j, j', k, k') = \left[(\hat{\theta}_{jk} - \hat{\theta}_{j'k}) - (\hat{\theta}_{jk'} - \hat{\theta}_{j'k'})\right] /(d_j - d_{j'})(c_k - c_{k'}), \tag{2}$$

where $\hat{\theta}_{jk} = \int_0^\infty \hat{F}_{jk}(t)dt$. The numerator in $\hat{\gamma}(j, j', k, k')$ represents the difference between an estimate of the expected difference in survival time for treatment levels j and j' at the covariate levels k and k'. Algebraically, this can be rearranged to represent the sum of diagonal elements minus the sum of off-diagonal elements of a matrix with rows j and j', and columns k and k'. Using all $M_1 = \frac{1}{2}m_1(m_1 - 1)$and $M_2 = \frac{1}{2}m_2(m_2 - 1)$ pairs of (j, j') and (k, k'), we define the estimator $\hat{\gamma}$ of γ given by

$$\hat{\gamma} = (M_1 M_2)^{-1} \sum_{k \neq k'}^{m_2} \sum_{j \neq j'}^{m_1} \hat{\gamma}(j, j', k, k') \tag{3}$$

For each level k, $1 \leq k \leq m_2$, and any pair (j, j'), $1 \leq j, j' \leq m_1$ we can estimate $\beta_1(k, j, j') = \frac{\theta_{jk} - \theta_{j'k}}{d_j - d_{j'}} - \gamma c_k$ by its natural estimator $\hat{\beta}_1(k, j, j') = \frac{\hat{\theta}_{jk} - \hat{\theta}_{j'k}}{d_j - d_{j'}} - \hat{\gamma} c_k$. Using all M_1 pairs of (j, j') we define the estimator $\hat{\beta}_1$ of β_1 by,

$$\hat{\beta}_1 = (m_2 M_1)^{-1} \sum_{k=1}^{m_2} \sum_{j \neq j'}^{m_1} \hat{\beta}_1(k, j, j') = (m_2 M_1)^{-1} \sum_{k=1}^{m_2} \sum_{j \neq j'}^{m_1} \left[\frac{\hat{\theta}_{jk} - \hat{\theta}_{j'k}}{d_j - d_{j'}} \right] - \hat{\gamma}\bar{c}, \tag{4}$$

where $\bar{c} = (m_2)^{-1} \sum_{k=1}^{m_2} c_k$, and $\hat{\gamma}$ is given in (4). Similarly we propose estimator $\hat{\beta}_2$ of β_2 by,

$$\hat{\beta}_2 = (m_1 M_2)^{-1} \sum_{k \neq k'}^{m_2} \sum_{j=1}^{m_1} \left[\frac{\hat{\theta}_{jk} - \hat{\theta}_{jk'}}{c_k - c_{k'}} \right] - \hat{\gamma}\bar{d}, \tag{5}$$

where $\bar{d} = (m_1)^{-1} \sum_{j=1}^{m_1} d_j$. Finally, we obtain the estimator $\hat{\alpha}$ of α

by,

$$\hat{\alpha} =$$
$$n^{-1}\left[\sum_{j=1}^{m_1}\sum_{k=1}^{m_2} n_{jk}\hat{\theta}_{jk} - \hat{\beta}_1\sum_{j=1}^{m_1} n_j d_j - \hat{\beta}_2\sum_{k=1}^{m_2} n_k c_k - \hat{\gamma}\sum_{j=1}^{m_1}\sum_{k=1}^{m_2} n_{jk} d_j c_k\right]$$
$$= n^{-1}(\sum_{j=1}^{m_1}\sum_{k=1}^{m_2} n_{jk}\hat{\theta}_{jk} - \hat{\beta}_1^*\sum_{j=1}^{m_1} n_j d_j - \hat{\beta}_2^*\sum_{k=1}^{m_2} n_k c_k$$
$$-\hat{\gamma}\sum_{j=1}^{m_1}\sum_{k=1}^{m_2}\left[\frac{n_j}{m_2} + \frac{n_k}{m_1} + n_{jk}\right] d_j c_k) = n^{-1}(\hat{\alpha}^* - s\hat{\gamma}) \qquad (6)$$

where $\hat{\alpha}^*$, $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ are, respectively, the same as $\hat{\alpha}$, $\hat{\beta}_1$ and $\hat{\beta}_2$ of Rahbar and Gardiner (2004), $s = \sum_{j=1}^{m_1}\sum_{k=1}^{m_2}\left(\frac{n_j}{m_2} + \frac{n_k}{m_1} + n_{jk}\right)$, $n_j = \sum_{k=1}^{m_2} n_{jk}$, $n_k = \sum_{j=1}^{m_1} m_{jk}$ and n is the total number of observation.

Each of the estimators $\hat{\alpha}$,$\hat{\beta}_1$,$\hat{\beta}_2$,$\hat{\gamma}$ has an asymptotic normal distribution. Since all these estimators can be expressed as a linear combination of $\hat{\theta}_{jk}$'s, say $\sum_{j=1}^{m_1}\sum_{k=1}^{m_2} L_{jk}\hat{\theta}_{jk}$ for constants $L_{jk}, 1 \le j \le m_1, 1 \le k \le m_2$, all asymptotic normality properties of the estimators which were provided by Rahbar and Gardiner (2004) are still valid and applicable for non-additive model with two factors. Therefore, we only show the asymptotic normality for the estimator$\hat{\gamma}$ and construct a consistent estimator of its asymptotic variance.

3 Asymptotic properties of $\hat{\gamma}$

First we express $\hat{\gamma}$ in the form

$$\hat{\gamma} = \sum_{j=1}^{m_1}\sum_{k=1}^{m_2} L_{jk}\hat{\theta}_{jk}, \qquad (7)$$

where the L_{jk} are fixed coefficients but functions of d_j and c_k. We, introduce the following notation. For $t < \tau_{H_{jk}}$ let $A_{jk}(t) = \int_{t-}^{\infty} F(u-\theta_{jk})du$, $\tilde{H}_{jk}(t) = \int_0^t G_{jk}(u)dF(u-\theta_{ij})$, $C_{jk}(t) = \int_0^t H_{jk}^{-2}(u)$

$\tilde{d}H(u)$ and

$$\sigma_{jk}^2 = \int_0^\infty A_{jk}^2(u)\, dC_{jk}(u). \tag{8}$$

***Theorem* 3.1** Suppose $\lambda_{jk} = \lim_{n\to\infty}(n_{jk}/n) > 0$, then under model (1),

$$\sqrt{n}\,(\hat{\gamma} - \gamma) \to N\left(0, \sigma_{\hat{\gamma}}^2\right),$$

where $\sigma_{\hat{\gamma}}^2 = \sum_{j=1}^{m_1}\sum_{k=1}^{m_2}\lambda_{jk}^{-1}\sigma_{jk}^2 L_{jk}^2$.

Proof: The proof of this theorem is identical to that of Theorem 3.1 of Rahbar and Gardiner (2004). Therefore we will not provide the details here.

Estimation of $\sigma_{\hat{\gamma}}^2$

Motivated by (9) we introduce a consistent estimator of the asymptotic variance $\sigma_{\hat{\gamma}}^2$, by

$$\hat{\sigma}_n^2 = \sum\nolimits_{j=1}^{m_1}\sum\nolimits_{k=1}^{m_2} nD_{jk}L_{jk}^2 \tag{9}$$

where

$$D_{jk} = \sum\nolimits_{i=1}^{n_{jk}} \delta_{ijk}(K_{jk}^{-1}(X_{ijk})\int_{X_{ijk}-}^{X_{jk}^{(n)}} \hat{F}_{jk}(u)\, du)^2 \tag{10}$$

and $K_{jk}(t) = 1 + \sum_{i=1}^{n_k}[X_{ijk} > t]$.

Using these results we can establish the strong consistency and asymptotic normality of $\hat{\sigma}_n^2$. The next theorem describes the asymptotic distribution of $\overset{2}{}_n$.

***Theorem* 3.2** Under the assumptions of Theorem 3.1 and the condition

$$\max_{1\leq j\leq m_l}\max_{1\leq j\leq m_2}\left\{\sqrt{n}\left(nn_{jk}^{-1}-\lambda_{jk}^{-1}\right)\right\}\rightarrow 0$$

$$\sqrt{n}\left(\sigma_n^2-\sigma_{\hat{\gamma}}^2\right)\rightarrow N\left(0,\eta\right)$$

where $\eta=\sum_{j=1}^{m_1}\sum_{k=1}^{m_2}\lambda_{jk}^{-3}\gamma_{jk}L_{jk}^4$.

Proof: Using (10) easily we obtain an almost sure representation for . Since the proof of this theorem is also identical to that of Theorem 3.2 of Rahbar and Gardiner (2004), we will not provide the details here.

Remarks

[1] The model in (1) can be modified to include more than two factors. Since the estimators of the regression parameters will be linear combinations of the form in (8), we will not repeat the estimation procedures in the presence of more than two factors. It can be also shown that the estimator of the higher order interactions all can be expressed as linear combinations of cell mean estimates, yet accounting for interactions.[2] The non-negativity assumption on the response variable T can be relaxed. The assumption of finite support for the survival time distribution in (1) could be also relaxed if we assume that its tail decays exponentially to zero. A natural estimator of θ_{jk} is then $\hat{\theta}_{jk}=\int_0^{X_{jk}^{(n)}}\hat{F}_{jk}(t)dt$ with$X_{jk}^{(n)}=\max\{X_{ijk}:1\leq i\leq n_{jk}\}$, where we have defined the product limit estimator to be zero beyond$X_{jk}^{(n)}$. The definition adopted by Gill leaves $\hat{F}_{jk}(t)$constant for $t>X_{jk}^{(n)}$. With this definition we will need some conditions on the tail behavior of the underlying distribution of the survival time, for example $\sqrt{n}\int_{X_{jk}^{(n)}}^{\infty}F\left(t\right)dt\rightarrow 0$ in probability. [3] To ensure the asymptotic normality of the survival mean estimate in each stratum, one should ensure availability of sufficient number of observations in each stratum. Even though the number of required observations in each stratum is dependent on several factors including the shape of the distribution of survival time and censoring rate, we recommend using this proposed

methodology when at least 30 observations are available in each stratum and the censoring rate is less than 50% in each stratum. [4] In presence of a significant interaction the post hoc paired comparisons of various treatment combinations is possible for the proposed model.

Simulations

In a simulation study, we assessed the impact of censoring rate on the standard error of the estimators of the parameters in (1). In addition, we compared the power of detecting an interaction using our proposed procedure with that of the Cox PH Model. Suppose there are two treatment groups labeled as$d_1 = 0$ and d_2=1 and five age levels having values c_1=3, c_2=9, c_3=15, c_4=21 and c_5=27. We first fixed the parameters,$\alpha = 2.5$, $\beta_1 = 1.25$, and $\beta_2 = 0.25$ in the regression model (1). Different scenarios for γ were considered as indicated in Table 1. Then using routines in SAS, we generated 2000 repeated samples of 300 independent observations for each of the two treatment groups ensuring approximately 20% of the data in each of the five levels of age covariate, based on the random error term having a uniform distribution with mean zero.

Table 1. Estimated power for testing the interaction effect when the error term in (1) has a uniform distribution and parameters are α=2.5, β_1=1.25,β_2=0.25, and γ, based on 2000 replications of 300 observations in each of the two treatment groups(%C ≡%Censored).$The censoring samples$,$Y_{1jk}, Y_{2jk}, \ldots, Y_{n_{jk}jk}$, were independent of the survival times generated from a uniform distribution on $(\theta_{jk} - a/b$,$\theta_{jk} + a/b + \psi)$, where ψ was assigned selected fixed values to achieve a desired level of censoring in order to assess the effect of censoring on the standard error of our estimators. We have also considered a situation where there was no censoring. Table 1 provides the mean and the standard deviation of all 2000 estimates for α,β_1,β_2, and γ in the regression model (1) as well as simulated power for detecting whether the interaction is statistically significant. In addition, in Figures 1, 2, and 3, we provide power functions for detecting the interaction using our propose procedure and the Cox PH model by increasing the parameter $\gamma$from zero to 0.05 by increments of 0.01 and for 0, 15, and 30% censoring rate respectively.

γ	%C	$\hat{\alpha}$		$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\gamma}$		Power	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE	New test	Cox PH
0	0	2.4968	0.1781	1.2467	0.2385	0.2503	0.0107	0.0000	0.0138	0.0520	0.0485
	15	2.5019	0.1965	1.2528	0.2556	0.2499	0.0118	-0.0002	0.0148	0.0590	0.0520
	30	2.5050	0.1984	1.2398	0.2602	0.2497	0.0118	0.0005	0.0153	0.0565	0.0630
0.01	0	2.5058	0.1872	1.2381	0.2422	0.2497	0.0110	0.0106	0.0140	0.1200	0.0700
	15	2.4929	0.1946	1.2475	0.2490	0.2503	0.0114	0.0104	0.0146	0.1155	0.0585
	30	2.5133	0.2022	1.2419	0.2635	0.2493	0.0118	0.0103	0.0153	0.1115	0.0520
0.02	0	2.5021	0.1837	1.2443	0.2425	0.2500	0.0110	0.0204	0.0142	0.3310	0.2155
	15	2.5009	0.1864	1.2508	0.2495	0.2499	0.0113	0.0200	0.0148	0.2875	0.1500
	30	2.5009	0.2055	1.2535	0.2663	0.2499	0.0122	0.0198	0.0156	0.2795	0.1125
0.03	0	2.5039	0.1796	1.2465	0.2429	0.2498	0.0108	0.0303	0.0143	0.5910	0.4380
	15	2.5018	0.1902	1.2435	0.2491	0.2501	0.0113	0.0302	0.0143	0.5555	0.3430
	30	2.5018	0.2036	1.2501	0.2686	0.2501	0.0118	0.0300	0.0154	0.5075	0.2450
0.04	0	2.5029	0.1875	1.2449	0.2439	0.2498	0.0111	0.0402	0.0142	0.8245	0.6910
	15	2.4964	0.1865	1.2537	0.2522	0.2503	0.0110	0.0396	0.0144	0.7890	0.5840
	30	2.5037	0.1953	1.2447	0.2602	0.2499	0.0116	0.0403	0.0154	0.7640	0.4470
0.05	0	2.5001	0.1891	1.2429	0.2485	0.2499	0.0111	0.0503	0.0144	0.9505	0.8730
	15	2.4984	0.1879	1.2479	0.2478	0.2502	0.0111	0.0502	0.0145	0.9345	0.7985
	30	2.5025	0.2018	1.2456	0.2668	0.2501	0.0120	0.0500	0.0154	0.9060	0.6575

Concluding Remarks/ Discussion In this paper we have presented a nonparametric regression model for the mean survival time which allows assessment of an interaction between the two factors in a multi-factor design. Our approach differs from the other methods referenced earlier in that it allows varying levels of variance and censorship at different levels of the covariates. The asymptotic normality of the estimators and their estimators of the asymptotic variances provide an opportunity to conduct appropriate test of hypotheses as well as constructing confidence intervals for the regression parameters in model Our simulation results indicate a tendency in higher standard errors as the rate of censoring increases. Since our estimation procedures are highly influenced by the Kaplan-Meier estimator, we recommend using these estimators when the sample size in various combinations of the levels is larger then 30 and each level of censor-

ing rate is less than 50%. The results in Table 1 indicate that our proposed procedure provides reasonable estimates of $\gamma$with consistent standard errors across different levels of censoring. In the absence of an interaction, (when γ is zero), the simulated power for both Cox PH model and our proposed method provide reasonably close to its nominal value of 0.05, type I errors, for all three different censoring levels. However, our proposed method has higher power for detecting the interaction as compared with the Cox PH Model. It is important to note that the simulated powers of our proposed method at 30% censoring rate are even better than that of the Cox PH model without any censoring. The estimated power functions shown in Figure 1, 2, and 3 are plotted based on the simulated powers presented in Table 1. The difference in the power functions between our proposed method and the Cox PH model is amplified by increasing the rate of censoring.

The almost sure representation and the asymptotic results provided for the estimators of the asymptotic variance of the regression parameters in model (1) are also useful when sequential estimation of these parameters is of particular interest. As shown in Rahbar and Gardiner (1995) when censoring is present the estimation procedures proposed in this paper are expected to produce a smaller standard error than the corresponding nonparametric estimators proposed by Koul *et al.* The results presented in this manuscript should be viewed as an extension of the results in Rahbar and Gardiner (2004) to the case of multiple covariates with interactions.

Acknowledgement We acknowledge and appreciate the contributions of Drs. Barbara and Bill Given for providing us with data from their clinical intervention trials of Family Home Care for cancer (CA 79280) and Automated Telephone Monitoring for Symptom Management (CA 30724), funded by the National Cancer Institute (NCI).

REFERENCES

1. Buckley, J. and James, I., 1979. Linear regression with censored data. *Biometrika* **66**, 429-436.

2. Cox, D. R., 1972. Regression Models and Life-Tables With Discussion. *J. Royal Statist Society, Ser* B, **34**, 187-220.

3. Gill, R., 1983. Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.* **11**, 49-58.

4. Given, B., Given, C.W., Sikorskii, A., Jeon, S., Sherwood, P., and Rahbar, M., 2006. The impact of providing symptom management assistance on caregiver reaction: results of a randomized trial. *J. Pain Symptom Manage* **32**, 433-443.

5. James, I. R. and Smith, P. J., 1984. Consistency results for linear regression with censored data. *Ann. Statist.* **12**, 590-600.

6. Kaplan, E.L. and Meier, P., 1958. Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

7. Koul, H., Susarla, V., and van Ryzin, J., 1981. Regression analysis with randomly right-censored data. *Ann. Statist.* **9**, 1276-1288.

8. Leurgans, S., 1987. Linear models, random censoring and synthetic data. *Biometrika* **74**, 301-309.

9. Miller, R.G., 1976. Least squares regression with censored data. *Biometrika* **63**, 449-464.

10. Rahbar, M. H., 1990. Sequential estimation of functionals of the survival curve under random censorship. *Sequential Analysis*, **9 (2)** , 137-150.

11. Rahbar, M. H. and Gardiner J. C., 1995. Nonparametric estimation of regression parameters using censored data with a discrete covariate. *Statist & Probab Letters*, Vol. **24,** 13-20.

12. Rahbar, M. H. and Gardiner J. C., 2004. Nonparametric modeling of the mean survival time in a multi-factor design based on randomly right-censored data. *Biometrical Journal*, Vol. **46**, 497-502.

13. Tukey, J. W., 1949. One degree of freedom for non-additivity. *Biometrics*, Vol. **5**, 232-242.

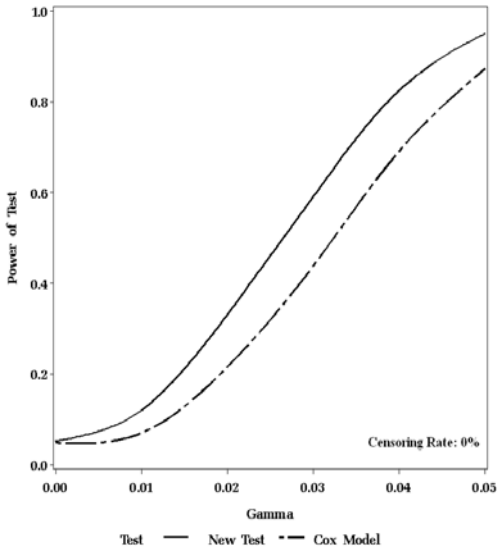


Figure 1: Estimated power for detecting interaction effects using our new test and the Cox proportional model. (No Censoring)

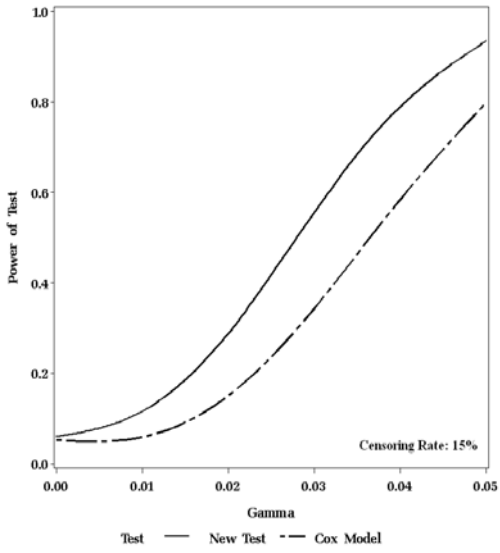


Figure 2: Estimated power for detecting interaction effects using our new test and the Cox Proportional Hazard model. (censoring rate= 15%)

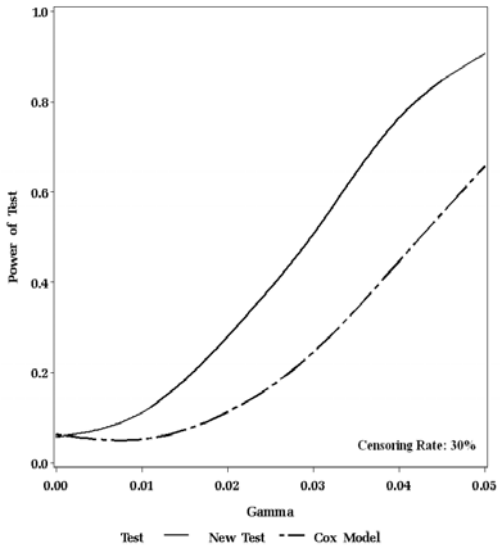


Figure 3: Estimated power for detecting interaction effects using our new test and the Cox proportional model. (Censoring rate =30%)

Asymptotic Behavior of Weighted Sums of Dependent Random Variables

V. Ranjbar. Y, M. Amini, and A. Bozorgnia

Department of Statistics, School of Mathematical Sciences, Ferdowsi University of Mashhad

Abstract: Let $\{X_n; n \geq 1\}$ be a sequence of random variables with common distribution function F. In this paper we study the asymptotic behaviors of tail probability of weighted and randomly weighted sums of heavy-tailed dependent random variables, say, $T_n = \sum_{i=1}^{n} c_i X_i$,$W_n = \sum_{i=1}^{n} \theta_i X_i$ and $W_N = \sum_{i=1}^{N} \theta_i X_i$, where $\{c_i\}$ is a sequence of bounded positive real numbers and $\{\theta_i\}$ is a sequence of positive random variables which are independent of sequence $\{X_n; n \geq 1\}$, and N is a nonnegative integer-valued random variables, independent of θ_i and X_i for all $i \geq 1$.

Keywords: Weakly Negative Dependent, Heavy-tailed, Asymptotic behavior, Weighted Sums.

1 Introduction and preliminaries

The asymptotic behaviors of the tail probability of partial sums and weighted sums of independent heavy-tailed random variables have been extensively investigated by many authors. Kiaw and Tang (2004), Chen et. all (2005) and Gluk et. all (2006), have been provided asymptotic behavior of weighted sums of subexponential random variables. Large deviations of weighted sums and partial sums for ND random variables and NA ones have been extended by Chen et. all (2007), Tang (2006) and Lio (2007), respectively. In this paper we study the asymptotic behaviors of the tail probabilities, $P(\Sigma_{i=1}^{n} c_i X_i > x)$, $P(\max_{1\leq i\leq n} \theta_i X_i > x)$ and $P(\Sigma_{i=1}^{N} \theta_i X_i > x)$ as x tends to infinity; where, $\{c_i\}$ is a sequence of bounded positive real numbers and $\{\theta_i\}$ is a sequence of positive random variables which are independent of the sequence $\{X_n; n \geq 1\}$, for one class of heavy-tailed distributions. Throughout this paper all distribution functions will be defined on

E-mail: vahidranjbar_y@yahoo.com.

$[0,\infty)$ and $f(x) \sim g(x)$ means that $lim_{x\to\infty} f(x)/g(x) = 1$ .We denote the tail of distribution of F by $\bar{F}(x) = 1 - F(x)$, convolution of distributions F and G by $F * G$ and denote n^{th} convolution of F by $F^{(n)}$ and $\overline{F^{(n)}} = 1 - F^{(n)}$. In the following, some well known classes of heavy-tailed distribution function are listed.

1)Dominated variation (D): The distribution function F is belongs to D if for any $1 < t < 1$,

$$\limsup_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} < \infty$$

2)Consistency varying tailed (C): The distribution function F is belong to C if,

$$\lim_{t\uparrow 1} \limsup_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = 1 \quad or \quad \lim_{t\downarrow 1} \limsup_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = 1$$

3)Long-tailed (L): The distribution function F is belongs to L if for any $t > 0$,

$$\lim_{x\to\infty} \frac{\overline{F}(x-t)}{\overline{F}(x)} = 1$$

For more details about the heavy-tiled distribution functions, see Cline and Samorodnitsky (1994), Bingham et. all. (1987) and Jelenkovic and Lazar (1999).

Moreover, as other classes of heavy-tailed distributions, we consider the following relation holds for the distribution functions,

$$0 < lim_{x\to\infty} \int_0^x F(u)du < \infty \qquad (1-1)$$

It is easy to find some evidence for the relation (1-1) like the distribution functions that belong to the class of subexponential distribution functions which was defined by Chistyakov (1946).

Definition 1 *The random variables $X_1, ..., X_n$ are said to be Weekly Negatively Dependent (WND) if for each n and all $x_1, ..., x_n$ there exist some $C \geq 1$ such that*

$$f_{X_1,...,X_n}(x_1, ..., x_n) \leq C.f_{X_1}(x_1) \times ... \times f_{X_n}(x_n) \qquad (1-2)$$

Lemma 1: Let $X_1, ..., X_n$ be WND random variables with distribution function $F(x_1, ..., x_n)$ and marginal distribution $F(x_1), ..., F(x_n)$, respectively, then it is easy to see that $F(x_1, ..., x_n)$ $\leq C.F(x_1), ..., F(x_n)$ and $\overline{F}(x_1, ..., x_n) \leq C.\overline{F}(x_1), ..., \overline{F}(x_n)$.

Lemma 2: Let $X_1, ..., X_n$ be WND random variables and $h_1(.), ..., h_n(.)$ be some monotone measurable functions, then $h_1(X_1), ..., h_n(X_n)$ are WND with the same constant C.

Corollary 1: Lemma 2 is valid if $h_i(X_i) = C_i X_i$, where $c_i \in R$.

Lemma 3: Let $\{a_i\}$ and $\{b_i\}$ be two sequences of positive real numbers, then

$$\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \leq max\left[\frac{a_i}{b_i}; i = 1, ..., n\right]$$

Lemma 4: Let X is a random variable with distribution function F and cX with distribution function of F' , where c is some constant, then

i. $F \in L$ if and only if $F' \in L$.
ii. $F \in D$ if and only if $F' \in D$.

2 Main results

In this section, first we prove a Lemma and then a Theorem, which is based on our work, and then we study the asymptotic behaviors of the tail probabilities of weighted sums of WND random variables. In fact, we prove some equivalence statements for tail probabilities.

Lemma 5: Let X_1 and X_2 be two WND random variables with common distribution function $F \in L$, if F satisfies in condition (1-1), then

for all $x > 2\upsilon > 0$,

$$\lim_{\upsilon\to\infty}\lim_{x\longrightarrow\infty}\int_{\upsilon}^{x}\frac{\overline{F}(x-u)}{\overline{F}(x)}dF(u)=0.$$

***Proof*:** we can write

$$\begin{aligned}\int_{\upsilon}^{x}\frac{\overline{F}(x-u)}{\overline{F}(x)}dF(u) &= \left[\int_{\upsilon}^{x-\upsilon}+\int_{x-\upsilon}^{x}\right]\frac{\overline{F}(x-u)}{\overline{F}(x)}dF(u)\\ &\leq \int_{\upsilon}^{x-\upsilon}\frac{\overline{F}(x-u)}{\overline{F}(x)}dF(u)+\overline{F}(x-\upsilon). \qquad (2-1)\end{aligned}$$

Now, for any $x > 2\upsilon$, we have

$$\int_{\upsilon}^{x-\upsilon}\bar{F}(x-u)dF(u)=\int_{\upsilon}^{x-\upsilon}\bar{F}(u)d\bar{F}(x-u)\leq\int_{\upsilon}^{x+\upsilon}\bar{F}(u)d\bar{F}(x-u)=I_1.$$

Now let

$$I=\int_{0}^{\infty}\bar{F}(u)du=\sum_{n=0}^{\infty}\int_{nh}^{(n+1)h}\overline{F}(u)du.$$

Since $\bar{F}(nh+h+\upsilon)\leq\bar{F}(u)\leq\bar{F}(nh+\upsilon)$ for $u\in[nh+\upsilon,nh+h+\upsilon]$, we have

$$h\sum_{n=1}^{\infty}\bar{F}(nh)=h\sum_{n=0}^{\infty}\bar{F}(nh+h)\leq I\leq h\sum_{n=0}^{\infty}\bar{F}(nh)<\infty, \qquad (2-2)$$

So we can write

$$\begin{aligned}
&\sum_{n=0}^{N_0-1} \bar{F}\left[(n+1)h+\upsilon\right]\left\{\bar{F}(x-(n+1)h-\upsilon)-\bar{F}(x-nh-\upsilon)\right\} \\
= &\sum_{n=0}^{N_0-1} \int_{nh+\upsilon}^{(n+1)h+\upsilon} \bar{F}\left[(n+1)h+\upsilon\right] d\bar{F}(x-u) \\
&\leq \sum_{n=0}^{N_0-1} \int_{nh+\upsilon}^{(n+1)h+\upsilon} \bar{F}(u) d\bar{F}(x-u) = I_1 \\
\leq &\sum_{n=0}^{N_0-1} \int_{nh+\upsilon}^{(n+1)h+\upsilon} \bar{F}\left[nh+\upsilon\right] d\bar{F}(x-u) \\
= &\sum_{n=0}^{N_0-1} \bar{F}\left[nh+\upsilon\right]\left\{\bar{F}(x-(n+1)h-\upsilon)-\bar{F}(x-nh-\upsilon)\right\},
\end{aligned}$$

where $N_0 = [x/h]$. Then we get

$$\frac{I_1}{\bar{F}(x)} \leq \sum_{n=0}^{N_0-1} \bar{F}(nh+\upsilon)\left\{\frac{\bar{F}(x-(n+1)h-\upsilon)-\bar{F}(x-nh-\upsilon)}{\bar{F}(x)}\right\}.$$

Since $F \in L$, then when x tend to infinity, for all value of n and υ,

$$\left\{\frac{\bar{F}(x-(n+1)h-\upsilon)-\bar{F}(x-nh-\upsilon)}{\bar{F}(x)}\right\}$$

tends to zero.Therefore for sufficient large x and for any $\varepsilon > 0$, we have,

$$K_2 \leq \frac{I_1}{\bar{F}(x)} \leq \varepsilon \sum_{n=0}^{N_0-1} \bar{F}(nh+\upsilon) \leq \sum_{n=0}^{\infty} \bar{F}(nh) < \varepsilon.M.$$

The final inequality valid by (2-2). This and (2-1) complete the proof.

Theorem 1: Suppose that X_1 and X_2 are two WND random variables

with the distribution functions F_1 and F_2 , respectively, where $F_i \in L; i = 1, 2$. If F satisfies in the condition (1-1) and $sup_x \overline{G}(x)/\overline{F}(x) < \infty$, then

$$P(X_1 + X_2 > x) \sim P(X_1 > x) + P(X_2 > x) \quad as \quad x \to \infty$$

***Proof*:** By assumption of $sup_x \overline{G}(x)/\overline{F}(x) = K < \infty$ and Lemma 1 , for every $\nu > 0$,we get

$$\begin{aligned}
&\overline{H}(x) = P(X + Y > x; X < \upsilon) + P(X + Y > x; \upsilon < X < x) \\
&+P(X > x) \leq C \int_0^{\upsilon} \overline{G}(x - u) dF(u) + C \int_{\upsilon}^{x} \overline{G}(x - u) dF(u) + \overline{F}(x) \\
&\leq C\overline{G}(x - u) + \overline{F}(x) \times CK \int_{\upsilon}^{x} \frac{\overline{F}(x - u)}{\overline{F}(x)} dF(u) + \overline{F}(x) \\
&\leq \left[\overline{F}(x) + \overline{G}(x)\right] \left\{ \left[1 + CK \int_{\upsilon}^{x} \frac{\overline{F}(x - u)}{\overline{F}(x)} dF(u)\right] \vee \left[\frac{C\overline{G}(x - \upsilon)}{\overline{G}(x)}\right] \right\},
\end{aligned}$$

where $x \vee y = \max\{x, y\}$. So

$$\lim_{\nu \to \infty} \limsup_{x \longrightarrow \infty} \frac{\overline{H}(x)}{\overline{F}(x) + \overline{G}(x)} \leq 1.$$

Since $G \in L$ and F satisfies in the condition (1-1) and by Lemma 5, the last inequality is valid. On the other hand, for any positive x we have,

$$\begin{aligned}
P(X + Y > x) &= P(X > x) + P(Y > x) \\
&+ P(X + Y > x; X < x; Y < x) - P(X > x; Y > x) \\
&\geq P(X > x) + P(Y > x) - P(X > x; Y > x) \\
&\geq P(X > x) + P(Y > x) - C.P(X > x)P(Y > x).
\end{aligned}$$

The second inequality is valid by Lemma 1. Then

$$\frac{P(X + Y > x)}{P(X > x) + P(Y > x)} \geq 1 + o(1).$$

This completes the proof.

Theorem 2: Let X_1 and X_2 be WND random variables with common distribution function $F \in L$. If F satisfies in the condition (1-1), then

$$P(c_1X_1 + c_2X_2 > x) \sim P(c_1X_1 > x) + P(c_2X_2 > x) \quad as \quad x \to \infty \tag{2-3}$$

where $0 < a < c_1, c_2 < b < \infty$.

***Proof*:** Let $c_1 = 1$, then Corollary 1 implies that X_1 and c_2X_2 are WND. Suppose that F' be distribution function of c_2X_2 and $0 < c_2 \leq 1$, we get

$$sup_x \frac{\overline{F}'(x)}{\overline{F}(x)} = sup_x \frac{P(X_1 + c_2X_2 > x)}{P(X_1 > x)} \leq 1,$$

therefore, Theorem 1 implies that

$$P(X_1 + c_2X_2 > x) \sim P(X_1 > x) + P(c_2X_2 > x) \quad as \quad x \to \infty \tag{2-4}$$

Now, suppose that, $c_2 \geq 1$, by (2-4) we obtain

$$\begin{aligned} &P(X_1 + c_2X_2 > x) \\ &= P(c_3X_1 + X_2 > x) \sim P(c_3X_1 > x) + P(X_2 > x) \quad as \quad x \to \infty \\ &= P(X_1 > x) + P(c_2X_2 > x) \end{aligned}$$

where $c_3 = 1/c_2$. Therefore, for any $c \in [a, b]$ we have

$$P(X_1 + cX_2 > x) \sim P(X_1 > x) + P(cX_2 > x) \quad as \quad x \to \infty \tag{2-5}$$

Moreover, for every $c_1, c_2 \in [a, b]$, by (2-3) we get

$$\begin{aligned} &P(c_1X_1 + c_2X_2 > x) \\ &= P(X_1 + c'X_2 > x) \sim P(X_1 > x) + P(c'X_2 > x) \quad as \quad x \to \infty \\ &= P(c_1X_1 > x) + P(c_2X_2 > x) \end{aligned}$$

Where $c' = c_1/c_2$. This completes the proof.

Theorem 3: Let $X_1, ..., X_n$ be WND random variables with common distribution function $F \in C$. If F satisfies in the condition (1-1), then for any positive real numbers a and b such that $0 < a \leq b < \infty$,

$$P\left(\sum_{i=1}^{n} c_i X_i > x\right) \sim \sum_{i=1}^{n} P\left(c_i X_i > x\right) \qquad (2-6)$$

where $a < c_i < b; i = 1, ..., n$.

***Proof*:** We proof (2-6) by induction approach. For $n = 2$, Theorem 2 implies (2-6). Now suppose that for any $m \geq 2$

$$P\left(\sum_{i=1}^{m} c_i X_i > x\right) \sim \sum_{i=1}^{m} P\left(c_i X_i > x\right) \qquad (2-7)$$

Geluk et. all (2004), show that, when (2-7) is true, then for every $x > 0$ and n,

$$P\left(\sum_{i=1}^{n} c_i X_i > x\right) \leq \sum_{i=1}^{n} P\left(c_i X_i > x\right) \qquad (2-8)$$

Since the argument is similar, we omitted it. Moreover, applying Lemma 2, we get

$$\begin{aligned} P\left(\sum_{i=1}^{m+1} c_i X_i > x\right) &\geq \sum_{i=1}^{m+1} P\left(c_i X_i > x\right) - \sum \sum_{i \neq j} P\left(c_i X_i > x; c_j X_j > x\right) \\ &\geq \sum_{i=1}^{m+1} P\left(c_i X_i > x\right) - \sum \sum_{i \neq j} C.P\left(c_i X_i > x\right) P\left(c_j X_j > x\right) \\ &\sim \sum_{i=1}^{m+1} P\left(c_i X_i > x\right) \quad as \quad x \to \infty. \qquad (2-9) \end{aligned}$$

Now (2-8) and (2-9) complete the proof.

Theorem 4: Let$X_1, ..., X_n$ be WND random variables, then

$$P\left(\max_{1\leq i\leq n} X_i > x\right) \sim \sum_{i=1}^{n} P(X_i > x) \quad as \quad x \to \infty$$

***Proof*:** Using Lemma 2, for every $x > 0$, we have

$$\begin{aligned} P\left(\max_{1\leq i\leq n} X_i > x\right) &= P\left(\bigcup_{1\leq i\leq n} (X_i > x)\right) \\ &\geq \sum_{i=1}^{n} P\left(X_i > x\right) - \sum\sum_{i\neq j} P\left(X_i > x; X_j > x\right) \\ &\geq \sum_{i=1}^{n} P\left(X_i > x\right) - \sum\sum_{i\neq j} C.P\left(X_i > x\right) P\left(X_j > x\right) \end{aligned}$$

Applying Lemma 3 we get

$$\sum\sum_{i\neq j} C.P\left(X_i > x\right) P\left(X_j > x\right) = o\left(\sum_{i=1}^{n} P\left(X_i > x\right)\right) \quad as \quad x \to \infty$$

Therefore

$$P\left(\max_{1\leq i\leq n} X_i > x\right) \geq \sum_{i=1}^{n} P\left(X_i > x\right) \quad as \quad x \to \infty$$

Moreover, it is easy to see that for all $x > 0$,

$$P\left(\max_{1\leq i\leq n} X_i > x\right) \leq \sum_{i=1}^{n} P\left(X_i > x\right) \quad as \quad x \to \infty$$

This completes the proof.

Theorem 5: Let $X_1, ..., X_n$ be WND random variables with common distribution function $F \in C$, which satisfies in the condition (1-1).

Moreover, if $\theta_1, ..., \theta_n$ are random variables independent of $X_1, ..., X_n$ and $P(a \leq \theta_i \leq b) = 1; i = 1, ..., n$, for some a and b such that $0 < a \leq b < \infty$, then

$$P\left(\max_{1\leq i\leq n} \theta_i X_i > x\right) \sim \sum_{i=1}^{n} P(\theta_i X_i > x) \sim P\left(\sum_{i=1}^{n} \theta_i X_i > x\right) \quad as \;\; x \to \infty \tag{2-10}$$

***Proof*:** Applying Theorem 4 we get

$$\begin{aligned} P\left(\sum_{i=1}^{n} \theta_i X_i > x\right) &= E_{\underline{\theta}}\left[P\left(\sum_{i=1}^{n} \theta_i X_i > x \mid \underline{\theta}\right)\right] \\ &\sim E_{\underline{\theta}}\left[\sum_{i=1}^{n} P\left(\theta_i X_i > x \mid \underline{\theta}\right)\right] \quad as \quad x \to \infty \\ &= \sum_{i=1}^{n} E_{\underline{\theta}}\left[P\left(\theta_i X_i > x \mid \underline{\theta}\right)\right] = \sum_{i=1}^{n} P\left(\theta_i X_i > x \mid \underline{\theta}\right). \end{aligned}$$

This proofs the second relation in (2-10). For the first one, applying conditional expectation role and Lemma 2, for every $x > 0$ we have

$$P\left(\max_{1\leq i\leq n} \theta_i X_i > x\right) \geq \sum_{i=1}^{n} P\left(\theta_i X_i > x\right) - \sum\sum_{i\neq j} P\left(\theta_i X_i > x; \theta_j X_j > x\right) \tag{2-11}$$

$$= \sum_{i=1}^{n} P\left(\theta_i X_i > x\right) - \sum\sum_{i\neq j} E_{\theta_i,\theta_j}\left[P\left(\theta_i X_i > x; \theta_j X_j > x \mid \theta_i, \theta_j\right)\right]$$

$$\geq \sum_{i=1}^{n} P\left(\theta_i X_i > x\right) - \sum\sum_{i\neq j} C.E_{\theta_i}\left[P\left(\theta_i X_i > x \mid \theta_i\right)\right] E_{\theta_j}\left[P\left(\theta_j X_j > x \mid \theta_j\right)\right]$$

$$= \sum_{i=1}^{n} P\left(\theta_i X_i > x\right) - \sum\sum_{i\neq j} CP\left(\theta_i X_i > x\right) P\left(\theta_j X_j > x\right)$$

The inequality is valid by Corollary 1 and independence of X and θ . Now applying Lemma 3, we get

$$\sum\sum_{i\neq j} CP\left(\theta_i X_i > x\right) P\left(\theta_j X_j > x\right) = o\left(\sum_{i=1}^{n} P\left(\theta_i X_i > x\right)\right) \quad as \;\; x \to \infty$$

Substituting this in (2-11), we obtain

$$P\left(\max_{1\leq i\leq n} \theta_i X_i > x\right) \geq \sum_{i=1}^{n} P\left(\theta_i X_i > x\right) \quad as \quad x \to \infty$$

This completes the proof.

Theorem 6: Let $\{X_n; n \geq 1\}$ be a sequence of WND random variables with common distribution function F, which satisfies in condition (1-1) and let$\{\theta_i; i \geq 1\}$ be a sequence of identical positive random variables which are independent of $\{X_n\}$. If N be a nonnegative integer-valued random variable independent of sequence $\{X_n\}$ and $\{\theta_i\}$ with $E(N) < \infty$, then,

$$P\left(\sum_{i=1}^{n} \theta_i X_i > x\right) \sim E(N).P\left(\theta_1 X_1 > x\right) \quad as \quad x \to \infty$$

***Proof*:** Using conditional expectation and Theorem 5, for every $x > 0$ we have

$$\begin{aligned} P\left(\sum_{i=1}^{n} \theta_i X_i > x\right) &= E_N\left[P\left(\sum_{i=1}^{n} \theta_i X_i > x \mid N\right)\right] \\ &\sim E_N\left[\sum_{i=1}^{n} P\left(\theta_i X_i > x \mid N\right)\right] \quad as \quad x \to \infty \\ &= E_N\left[\sum_{i=1}^{n} P\left(\theta_i X_i > x\right)\right] = \sum_{k=0}^{\infty}\sum_{i=1}^{k} P(\theta_i X_i > x) \\ &\times .P(N \geq i-1) = P(\theta_1 X_1 > x).E(N) \end{aligned}$$

This completes the proof.

Corollary 2: Under the assumption of Theorem 6, if $P(\theta_i = 1) = 1$ for all$i \geq 1$, then

$$P\left(\sum_{i=1}^{n} X_i > x\right) \sim E(N).P\left(X_1 > x\right) \quad as \quad x \to \infty$$

Remark : Let $\{X_n\}$ be a sequence of independent random variables with common distribution function F, then condition $F(x_1, ..., x_n) \leq C.\prod_{i=1}^{n} F(x_i)$ can be replaced by $F(x_1, ..., x_n) = \prod_{i=1}^{n} F(x_i)$. Thus, all above Theorems, Lemmas and Corollaries are true in this case, particularly; Theorem 1 in [4] and Theorem 3.1 in [9] are special case of Theorem 1 and Theorem 3, 4, 5 in this paper, respectively.

References

[1] Cal, J. and Tang, Q. (2004). On max-sum equivalent and convolution closure of heavy-tailed distributions and their applications. J.Appl.Prob.41, 117-130.

[2] Chen, Y. and Kiaw.NG. Tang, Q. (2005). Weighted sums of Subexponential random variables and their maxima. J.Appl.Prob.37, 510-522.

[3] Chen, Y. and Zhang, W. (2007). Large deviations for random sums of negatively dependent random variables with consistence varying tails. Statistics and probability letters.77, 530-538.

[4] Embrechts, P. and M.Goldie, C. (1979). On closure and factorization properties of Subexponential and related distributions. J. Austral. Math. Soc. 29, 243-256.

[5] Geluk, J.L. and DeVries, C.G. (2006). Weighted sums of Subexponential random variables and asymptotic dependence between returns on reinsurance equities. Mathematics and Economics. 38, 39-56.

[6] Kiaw.NG. and Tang, Q. (2004). Asymptotic behavior of tail and local probabilities for sums of Subexponential random variables. J. Appl. Prob. 41, 108-116.

[7] Lio, Y. (2007). Precise large deviations for negatively associated random variables with consistency varying tails. Statistics and probability letters. 77, 181-189.

[8] Tang, Q. (2006). Insensitivity to negative dependence of the asymptotic behavior of precise large deviations.

[9] Tang, Q. and Tisitsiashvili, G. (2003). Randomly weighted sums of subexponential random variables with application to ruin theory. Extremes 6, 171-188.

[10] Wang, D. and Tang, Q. (2004). Maxima of sums and random sums for negatively associated random variables with heavy tails. Statistics and probability letters.68, 287-295.

Influence diagnostic in partial least squares in measurement error models

A. Rasekh, and F. Mazarei

Department of Statistics, Shahid Chamran University

Abstract: Partial least squares (*PLS*) regression has received increasing attention in recent years. However, like other regression methods, *PLS* fitting could be substantially altered by one or a few influential points. This paper assesses the influence of observations on estimation of parameters in *PLS* measurement error models based on SIMPLS algorithm. We derive Cook's Distance (*CD*) and the sum of squared influence functions for sample i on the predicted concentrations of all other samples (*SID*) as tools for influence diagnostic. We illustrate our idea via a numerical example.

Keywords: Partial least squares; Influence function; Measurement error.

1 Introduction

Partial least squares (*PLS*) regression (Wold and Krishnaiah, 1966; Helland, 2001; Wold and Trygg, 2001) is one of the most widely used chemometrical tools to estimate concentrations from measured spectra. In recent years partial least square regression has been extended to the measurement error models (see Faber and Kowalski, 1997). As it is mostly a chemometrical tool, it has hitherto only been granted little attention in the statistical literature. A consequence thereof is that some properties of partial least squares regression have never been investigated. One of these properties is the influence function (*IF*) (Hampel et al., 1986), which is of widespread use in the literature on regression models.

In this paper we consider influence diagnostic in the measurement error models. We derive the influence function for partial least squares regression as a diagnostic tool to assess the influence of individual

Corresponding author: Tel: 0098 611 3331043; fax: 0098 611 3331043. E-mail address: Rasekh_a@scu.ac.ir

calibration samples on prediction in this context. In Section 2, we introduce the reader to the notation used throughout this article. In Section 3, we provide a short introduction to partial least squares regression. In Section 4, we introduce the reader to the population version of *PLS*, since this insight is needed for a correct computation of the influence function. In Section 5, we introduce the reader to the basic concept of the influence function, as well as the results that can be deduced thereof. In Section 6, we propose algorithms which allow efficient calculation of the influence function, derived from different *PLS* algorithms. In Section 7, this leads to a sample specific prediction interval in *PLS*, as well as to novel diagnostic plots. In section 8 we illustrate our idea via a numerical example.

2 Notation and definitions

Before we can give an introduction to partial least squares regression, we first need to define the notation used. The calibration matrix X is a matrix of size $n \times p$ in which the rows are n spectra of standard samples, measured at p channels. The corresponding n concentrations of these standard samples constitute the response vector y. Throughout this paper, we will assume both calibration and response matrices to be mean-centered. When calibration is completed, a spectrum of a new sample is denoted as ξ. The corresponding (mostly unknown) concentrations and their estimates are consistently denoted as y_ξ and $\hat{y}_\xi$, respectively. In case of measurement error models we assume that y and X are unobservable and they can be observed through $\tilde{y}$ and $\tilde{X}$ with measurement errors Δy and ΔX, respectively. We also assume that Δy and $vec(\Delta X)$ are uncorrelated with covariance matrices

$$var(\Delta y) = E(\Delta y \Delta y') = \sigma^2_{\Delta y} I$$
$$var(vec \Delta X) = E(vec \Delta X (vec \Delta X)') = \sigma^2_{\Delta X} I.$$

Through out this paper expected values with respect to a distribution G will be denoted as $EG(\ )$ and argmax(F) denotes the point (argument) that maximizes F. Finally, T and $IF(\ )$ denote the transposition

and the influence function, respectively.

3 Partial least squares regression

Partial least squares regression can be seen as a way to estimate a regression vector β in a linear model $\tilde{y} = X\beta + \Delta y$. In this equation, the elements of $\Delta\, y$ are identically and independently distributed errors with zero expectation and constant variance $\sigma^2_{\Delta y}$. In case of measurement error models we have

$$\begin{aligned} &\tilde{y} = y + \Delta\, y \\ &\tilde{X} = X + \Delta\, X \\ &\tilde{y} = (\,\tilde{X} - \Delta X\,)\,\beta + \Delta\, y = \tilde{X}\,\beta - \Delta\, X\,\beta + \Delta\, y. \end{aligned} \tag{1}$$

PLS is a latent variable regression technique. This means that *PLS* extracts independent latent variables from the original set of p (often correlated) variables. The regression vector is calculated from these latent variables, hence overcoming difficulties in ordinary least squares such as multicollinearity. In a spectrometric context, one can intuitively see that this is a correct way to proceed, as pointed out by Wold (1993).

In *PLS*, the latent variables are computed in such a way that they contain a maximum of relevant information concerning the relation between X and y. mathematically, this is expressed by the following objective function (Ter Braak and De Jong, 1998) in which the hth weighting vector ($\hat{a}_h$) is defined as: $\hat{a}_h = \arg\max_{a}\ Cov(Xa_h,\ \tilde{y})$, under the constraints that $\|a_h\| = 1$ and $a_h' X' X a_i = 0$ for $1 \leq i < h$.

This objective is a maximization problem under two constraints, which can be solved by dint of the Langrange multiplier method. All univariate *PLS* algorithms share the same objective function. However, different algorithms have been proposed to accomplish the same objective in which different scaling conventions are used. For example, in SIMPLS (De Jong, 1993), the convention is to re-scale the estimated weighting and score vectors (i.e.,$\hat{t}_h = X\hat{a}_h$) in such a way that the score vectors ultimately carry unit variance. In any algo-

rithm, the first weighting vector must be the dominant eigenvector of the matrix $X'\tilde{y}\tilde{y}'X$, which will then be or be not scaled, according to the convention imposed. From the second latent variable on, the second constraint becomes important. It requires the following latent variables to be orthogonal (uncorrelated) to the previous ones. Hence, the following weighting vectors will be dominant eigenvectors of the matrix $X'\tilde{y}\tilde{y}'X$, multiplied by a projection matrix which projects onto the orthogonal complement of the subspace spanned by the previous score vectors. Hence, before scaling, the hth weighting vector will be equal to:

$\hat{a}_h = X'(I_n - \sum_{i=1}^{h-1} \frac{\hat{t}_i\hat{t}_i'}{\hat{t}_i'\hat{t}_i})\tilde{y}$.

4 *PLS* at the population level

We first give a short description of the distinction between *PLS* at the population level and *PLS* at the sample level. We start with the case in which the true spectral matrix X is exists. Then we extend our discussion to the measurement error case. Individual experiments are samples taken from a certain population. For example, the data matrix X corresponds to a p-variate vector x of which n samples are drawn from the population. In chemometrics, this discrepancy is most currently disregarded, as the theoretical background is of minor importance to the analytical chemist. However, computation of the influence function requires prior definition of *PLS* at the population level. Let $(x', \tilde{y})$ be centered and distributed with given distribution G, then the objective for *PLS* is:

$$\hat{a}_h(G) = \arg\max_a E_G[a'x\tilde{y}] \tag{2}$$

under the constraints that

$$\|a_h(G)\| = 1 \quad \text{and} \quad a_h(G)'S(G)a_i(G) = 0 \quad \text{for} \quad 1 \leq i < h. \tag{3}$$

The only difference to the objective function stated in the previous section is the explicit dependence on the distribution G. As this

does not change the maximization problem, the exact solutions of Eqs. 2 and 3 are known and can be copied to the population level from the aforementioned algorithms. Hence, the SIMPLS (De Jong, 1993) algorithms also hold at the population level, if one does not omit the fact that all vectors are population quantities and thus dependent on the distribution G. The starting values should in this case be $s(G) = E_G[x\,\tilde{y}]$ and $S(G) = E_G[x'x]$. The population quantities corresponding to SIMPLS are defined as follows:

$$\begin{aligned}
a_h &= \begin{cases} s(G) & h = 1 \\ (I_m - \ddot{V}_{h-1}(G))\, a_{h-1}(G) & h > 1 \end{cases} \\
r_h(G) &= \frac{a_h(G)}{\sqrt{a_h'(G)\, S(G)\, a_h(G)}} \\
P_h(G) &= S(G)\, r_h(G) \\
v_h(G) &= \left(I_m - \textstyle\sum_{i=1}^{h-1} \ddot{V}_i(G)\right) P_h(G) \\
\ddot{V}_h(G) &= \frac{v_h(G)\, v_h'(G)}{v_h'(G)\, v_h(G)} \\
\beta_h(G) &= R_h(G)\, R_h'(G)\, s(G) \qquad 1 \le h \le p
\end{aligned} \tag{4}$$

Recall that $R_h(G)$ is a matrix containing $r_1(G)$, ..., $r_h(G)$ as its columns. Remark that the above equations hold for any given distribution G. The only condition is that the starting quantities $s(G)$ and $S(G)$ need to exist. The above equations define the statistical functionals a_h,r_h, P_h, v_h, $\ddot{V}_h$ and β_h all being defined as mappings sending distributions G to vector or matrix valued quantities. To return to the sample level, the empirical distribution G_n may be plugged in for G into the above expressions to yield the well-known *PLS* algorithms. The empirical distribution functionG_n is a discrete distribution giving mass $\frac{1}{n}$ to each of the n measured data points, and can be shown that will converge to G. It is therefore the sample-based analogue of G. Starting from

$$s(G_n) = E_{G_n}\,[x'\,\tilde{y}] = X'\,\tilde{y}/n \quad \text{and} \quad S(G_n) = E_{G_n}\,[x'\,x] = X'\,X/n \tag{5}$$

one finds all other quantities, now based on the sample, by applying Eqs. 4. Let ξ be a new observation (spectrum), and denote h the selected number of latent variables. Then the functional $\hat{y}_{h,\xi}$ corresponding to the predicted value based on ξ is defined as $\hat{y}_{h,\xi}(G) = \xi'\beta_h(G)$

for any distribution G. At the sample level, this corresponds to predicting a concentration of a (possibly new) sample on the basis of the calibration matrix.

If instead of having the true spectral matrix X, one measures $\tilde{X} = X + \Delta X$, then the measurement error ΔX will reflected in weighting vectors and other steps of PLS algorithm through fluctuation of covariance matrices $s(G)$ and $S(G)$. In this case the covariance matrices are $\tilde{S}(G) = E_G[\tilde{x}'\,\tilde{x}]$ and $\tilde{s}(G) = E_G[\tilde{x}'\,\tilde{y}]$. The population quantities corresponding to SIMPLS are defined as follows (see Faber and Kowalski, 1997):

$$\begin{aligned}
\tilde{a}_h &= \begin{cases} \tilde{s}(G) & h = 1 \\ (I_m - \tilde{\ddot{V}}_{h-1}(G))\,\tilde{a}_{h-1}(G) & h > 1 \end{cases} \\
\tilde{r}_h(G) &= \frac{\tilde{a}_h(G)}{\sqrt{\tilde{a}'_h(G)\,\tilde{S}(G)\,\tilde{a}_h(G)}} \\
\tilde{P}_h(G) &= \tilde{S}(G)\,\tilde{r}_h(G) \\
\tilde{v}_h(G) &= \left(I_m - \textstyle\sum_{i=1}^{h-1} \tilde{\ddot{V}}_i(G)\right)\,\tilde{P}_h(G) \\
\tilde{\ddot{V}}_h(G) &= \frac{\tilde{v}_h(G)\,\tilde{v}'_h(G)}{\tilde{v}'_h(G)\,\tilde{v}_h(G)} \\
\tilde{\beta}_h(G) &= \tilde{R}_h(G)\,\tilde{R}'_h(G)\,\tilde{s}(G) \qquad 1 \le h \le p
\end{aligned} \tag{6}$$

in which $\tilde{r}_1,\ ...,\ \tilde{r}_h$ are columns of $\tilde{R}_h$ matrix.

5 The notion of the influence function

The influence function has been introduced by Hampel et al. (1986) in order to theoretically assess the influence that an observation $\tilde{z}$ has on the value that a statistical functional T takes. This observation $\tilde{z}$ may be an observed data point or a potential outlier. When applied to PLS, the observation $\tilde{z}$ will be a couple ($\tilde{x}'$,$\tilde{y}$) containing a spectrum and its corresponding concentration. One supposes that a small fraction ε of the data is placed at the point $\tilde{z}$, whilst the other fraction $(1-\varepsilon)$ is coming from the population distribution G. Hence, the distribution becomes: $G_\varepsilon = (1-\varepsilon)G + \varepsilon\delta_{\tilde{z}}$ where $\delta_{\tilde{z}}$ is the point mass distribution at $\tilde{z}$. The influence function is then defined as

$$IF(\tilde{z},\ T,\ G) = \lim_{\varepsilon \to 0} \frac{T[(1-\varepsilon)G + \varepsilon\delta_{\tilde{z}}] - T(G)}{\varepsilon}. \tag{7}$$

In this paper, we are interested in assessing the influence of an observation on the *PLS* procedure in measurement error models, hence, we will compute the *IF* for *PLS* in these models. Actual computation of the *IF* makes use of derivation of the statistical functions, since Eq.8 yields: $IF(\tilde{z},\ T,\ G) = \frac{\partial}{\partial\varepsilon} T(G_\varepsilon)|_{\varepsilon=0}$.

6 Algorithms for the influence function

We derive the influence function when both dependent and independent variables are measured with error. First the influence functions for the starting values of the algorithm should be found. Let $\tilde{z}' = (\tilde{x}', \tilde{y})$ be an arbitrary point in the $(p+1)$-dimensional space. We will make usage of the shorthand notations $IF(T)$ instead of $IF(\tilde{z},\ T,\ G)$ and hence drop the dependence on the distribution G and on $\tilde{z}$. It is easy to check that the influence functions of covariance matrices are $IF(\tilde{S}) = \tilde{x}'\tilde{x} - \tilde{S}$ and $IF(\tilde{s}) = \tilde{x}'\tilde{y} - \tilde{s}$ where $\tilde{s}$ and $\tilde{S}$ are shorthand notations for $\tilde{s}(G)$ and $\tilde{S}(G)$. With these starting values, the influence functions can be computed recursively. The influence functions of weighting factors are:

$$IF(\tilde{a}_h) = \begin{cases} IF(\tilde{s}) & h = 1 \\ IF(\tilde{a}_{h-1}) - IF(\tilde{\ddot{V}}_{h-1})\,\tilde{a}_{h-1} - \tilde{\ddot{V}}_{h-1}\,IF(\tilde{a}_{h-1}) & h > 1 \end{cases}. \tag{8}$$

Furthermore, the influence functions of rescaled weighting factors and loading factors are

$$\begin{aligned} IF(\tilde{r}_h) &= \frac{IF(\tilde{a}_h)}{(\tilde{a}_h'\tilde{S}\tilde{a}_h)} - \frac{1}{2}\frac{\tilde{a}_h(IF(\tilde{a}_h')\tilde{S}\tilde{a}_h + \tilde{a}_h' IF(\tilde{S})\tilde{a}_h + \tilde{a}_h'\tilde{S}\,IF(\tilde{a}_h)}{(\tilde{a}_h'\tilde{S}\tilde{a}_h)^{3/2}} \\ IF(\tilde{P}_h) &= IF(\tilde{S})\,\tilde{r}_h + \tilde{S}\,IF(\tilde{r}_h) \end{aligned} \tag{9}$$

while the influence functions of $\tilde{v}_h$ and $\tilde{\ddot{V}}_h$ can be derived as:

$$\begin{array}{l} IF\,(\,\tilde{v}_h\,) = IF\,(\,\tilde{P}_h\,) - \sum_{i=1}^{h-1}\,[\tilde{\ddot{V}}_i\,IF\,(\,\tilde{P}_h\,) + IF\,(\,\tilde{\ddot{V}}_i\,)\,\tilde{P}_h] \\ IF\,(\,\tilde{\ddot{V}}_h\,) = \frac{IF\,(\,\tilde{v}_h\,)\,\tilde{v}'_h + \tilde{v}_h\,IF'\,(\,\tilde{v}_h\,)}{\tilde{v}_h\,\tilde{v}'_h} - 2\frac{\tilde{v}'_h\,\tilde{v}_h}{(\,\tilde{v}_h\,\tilde{v}'_h\,)^2} IF'\,(\,\tilde{v}_h\,)\,\tilde{v}_h \end{array} \tag{10}$$

Finally, with the aid of above relations, the influence function of regression coefficients will be

$$IF\,(\,\beta_h\,) = IF\,(\,\tilde{R}_h\,)\,\tilde{R}'_h\,\tilde{s} + \tilde{R}_h\,IF^T\,(\,\tilde{R}_h\,)\,\tilde{s} + \tilde{R}_h\,\tilde{R}'_h\,IF\,(\,\tilde{s}\,) \tag{11}$$

in which $\tilde{R}_h = \{\,\tilde{r}_1\,,\,...\,,\,\tilde{r}_h\,\}$.

A proof of the expressions in Eqs. 11–14 is straightforward by applying the functionals on G_ε, using Eq. 9 and standard differentiation rules, applied to the *PLS* algorithms at the population level. The equations are valid at any given step h of the iteration. Furthermore, note that the influence function for the prediction $\hat{y}_{h,\xi}$ is immediately obtained as $IF(\hat{y}_{h,\xi}) = \xi' IF(\beta)$. The expressions found for the *IF* are valid for any distribution G for which $\tilde{s}$ and $\tilde{S}$ are existing. In practice, the population distribution G is unknown but can be estimated by the empirical distribution function G_n. Hence, in the practical applications, the *IF* will always be evaluated for G taken to be G_n. This implies that all quantities $\tilde{s}$, $\tilde{S}$, $\tilde{a}_h$, β_h, $\tilde{r}_h$,... appearing in Eqs. 11-14 are taken to be the sample estimates as obtained by plugging G_n in Eqs. 7.

7 Applications of the influence function

Most commonly, the influence of individual samples is shown using the so-called Cook's Squared Distance (Massart et al., 1997). Roughly, it measures the change in the regression coefficients if the ith observation is omitted from the data. It is computed as follows:

$$CD(\tilde{z}_i) = \frac{1}{p\sigma_e^2}\sum_{j=1}^{n}(\hat{y}_j - \hat{y}_j^i)^2$$

In this equation, σ_e^2 is the residual variance and $\hat{y}_j^i$ denotes the predicted concentration for sample j based on a regression vector computed from calibration matrices from which sample i has been deleted. A large value for the *CD* is an indication that an observation is an influential observation or outlier. It should be clear that the influence function is apt to be a suitable measure for the influence of a sample on prediction. An analogous approach as in the Cook's Squared Distance leads to a diagnostic plot based on the influence function which is a viable alternative to the existing approaches. The measure of influence of sample i on prediction is the sum of squared influence functions for sample i on the predicted concentrations of all other samples, i.e:

$$SID\,(\,\tilde{z}_i\,) = \frac{1}{n}\sum_{j=1}^{n} IF\,(\,\tilde{z}_i\,,\,\hat{y}_j\,,\,G_n\,)^2$$

Where $\hat{y}_j$ stands for the statistical functional $\hat{y}_{h,\tilde{x}_i'}$. Both *CD* and *SID* should theoretically lead to the detection of the same influential samples, as both approaches are very closely related. For ordinary least squares regression, the relation between residual based methods (such as the *CD* here) and methods based on the influence function have been described in detail in literature (Cook and Weisberg, 1982).

8 Egyptian pottery data

Now we give an illustrative example of the methodology developed. Computations were performed by specific purpose programs written in *Microsoft Visual C++6.* These programs are available from the first author upon request. we consider a set of data, which we refer to as the Egyptian pottery data. Briefly, this data set arises from an extensive archaeological survey of pottery production and distribution in the ancient Egyptian city of Al-Amarna. The data consist of measurements of chemical contents (mineral elements) made on many samples of pottery using two different techniques, *NAA* and *ICP*. The set of pottery has been collected from different locations around the city. The group structure among the objects arises from two main

Table 1: Partial least square estimate of regression parameters

$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$
7.278×108	-3.066×108	4.530×109	1.0761×108	1.2152×107	-1.579×1011

sources, fabric code and location of pottery. Both of these subdivisions are important to the archaeologists. Consequently, according to this group structure the selected vessels have been divided into 27 groups. In each group there are different numbers of vessels from the same fabric code and provenance which can essentially be regarded as replicated observations. Among all mineral elements our interest is in the relation between *Na* measured with *NAA* versus six elements *Na, Al, K, V, Cr* and *Mn* measured with *ICP*. The data set is available at *http://www.cua.ac.ir/math/sta/rasekh/data.txt.*

Because there is error in measurements with both *NAA* and *ICP* techniques, Rasekh (2001) analyzed this data set and fitted a functional measurement error model. He showed that there is moderate collinearity among some of the independent variables. He derived the ridge estimate with ridge parameter 0.15 and showed that this estimate will improve the matters. In this section we re-analyse this data set and we compute the partial least squares measurement error model. Then we consider the influential observations. As there are replicated observations in each group, therefore we have used one observation for prediction and the remaining observations for estimation of the regression parameters. Table 1 shows the *PLS* estimate of parameters of the measurement error model according to the SIMPLES algorithm. In the next step we evaluated the predicted values, residuals and *SID* for each observation. The results are given in table 2. Table 2 show that observations 6, 13, 21 and 23 have more influence on regression coefficients than other observations, respectively.

Table 2: The dependent variables, Predicted values, residuals and *SID* values

i	y	$\hat{y}$	Residuals	*SID*	i	y	$\hat{y}$	Residuals	*SID*
1	-0.26	-0.61	0.362	4.357	15	0.82	0.02	0.800	14.960
2	0.40	0.12	0.295	-17.596	16	0.47	0.07	0.406	-11.722
3	0.75	0.39	0.368	22.333	17	-0.42	-0.17	-0.248	54.623
4	0.00	-0.09	0.095	57.373	18	-0.86	-0.19	-0.668	64.072
5	0.31	0.09	0.222	16.574	19	0.15	0.05	0.100	-60.874
6	-0.32	-0.32	-0.012	**-166.83**	20	0.57	0.20	0.374	25.145
7	-0.02	-0.03	0.148	-1.113	21	-0.06	-0.25	0.191	**-200.14**
8	0.04	0.03	0.015	73.041	22	0.53	0.08	0.449	11.968
9	-0.33	-0.08	-0.252	11.483	23	-0.57	-0.05	-0.515	**-110.35**
10	-0.11	-0.10	-0.012	55.695	24	-0.06	-0.16	0.104	-30.687
11	-0.78	-0.05	-0.726	-27.825	25	-0.32	0.33	-0.012	23.344
12	0.85	0.10	0.781	-23.003	26	-0.28	-0.28	0.002	5.513
13	-0.17	-0.02	-0.146	**-220.25**	28	0.08	0.10	-0.017	41.186
14	-.032	-0.28	-0.041	74.492					

References

[1] Cook, R.D., Weisberg, S., 1982. *Residuals and Influence in Regression*, 2nd edition, Chapman and Hall, London.

[2] De Jong, S., 1993. SIMPLS: An alternative approach to partial least squares regression, *Chemometrics and Intelligent Laboratory System s*, **18**, 251-263.

[3] Faber, K. and Kowalski, B.R., 1997. Propagation of measurement error for the validation of predictions obtained by principal component regression and partial least squares, *J. Chemometrics,* **11**, 181-238.

[4] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A., 1986. *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York.

[5] Helland, I.S., 2001. Some theoretical aspects of partial least squares regression, *Chemometrics and Intelligent Laboratory Systems,* 58, 131–150.

[6] Rasekh, A.R., 2001. Ridge estimation in functional measurement error models. *Annales DE L'I.S.U.P., Pub. De L'Institut de Statistique de L'Universite de Paris, XXXXV, Fascicule 2-3, 47-59.*

[7] Ter Braak, C.J.F. and De Jong, S., 1998. The objective function of partial least squares regression, *J. Chemometrics*, 12, 41–54.

[8] van der Voet, H., 1999. Pseudo-degrees of freedom for complex predictive models: the example of partial least squares, *J. Chemometrics*, 13, 195– 208.

[9] Wold, S., 1993. PLS in chemical practice, *Technometrics*, 35, 136–139.

[10] Wold, S., Trygg, J., Berglund, A. and Antti, H., 2001. Some recent developments in PLS modelling, *Chemometrics and Intel. Lab. Systems* 58, 131–150.

Latend transition models for analyzing longitudinal ordinal response data with random dropout

Z. Rezaei Ghahroodi[1], and M. Ganjali[2]

[1]*Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University*
[2]*Statistical Research and Training Center No. 52, Sartip Fakouri St., Babataher St., Fatemi Ave. Tehran 1413717911*

Abstract: Many methods are available for analyzing incomplete longitudinal ordinal response data when repeated measurements occur at different time points for each subject. Since in these studies there are a sequence of ordinal responses for the same individual, we have to consider the correlation between responses. Different models can be used to take into account this correlation. In this paper, in order to examine the influence of the Markovian assumption on the covariate effects, a latend transition model which given each level of previous response, has its own dynamic parameterizations for all covariates, is presented. One can use available software to find the parameter estimates of the model. However, for testing the stationarity of covariate effects or the similarity of covariate effects at different levels of previous outcome, optimization of the likelihood by numerical methods is necessary. The presented model is applied to Fluvoxamine (a treatment for deregulation of serotonin in the brain) data.

Keywords: Longitudinal data; Ordinal response; Latend transition model; Random dropout.

1 Introduction

Longitudinal ordinal categorical data are frequently collected in medical science where responses of interest for each subject are recorded in a categorical ordinal scale and observed repeatedly at several times or under various conditions. For the analysis of these studies we need to accommodate the statistical dependence between repeated measurements of the same individual. A variety of approaches to the analysis of dependent have been proposed to handel such correlation between

responses. One such possibility is marginal modelling, which can be used to make inferences about population-averaged parameters [Ten Have et al. (1996); Kim (1995); Liang et al. (1992)]. A second possibility is random effects modelling, which makes inferences about variability between respondents, [Harvile and Mee (1984); Verbeke and Lesaffre (1996); Tutz (2005)]. The third approach would be to use Markov (transition) models [see Anderson and Goodman (1957); Francom et al. (1989); Ganjali and Rezaee (2007)]. For reviews of transition and other models for ordinal longitudinal responses, see McCullagh (1980), Diggle et al. (2002) and Agresti (2002).

In longitudinal data, it is not unusual in practice for some sequences of measurements to terminate early for reasons outside the control of the investigator, and any unit so affected is often called a drop-out. Little and Rubin (2002) and Diggle and Kenward (1994) make important distinctions between the various types of drop-out mechanisms. A drop-out process is said to be a completely random drop-out (CRD) if the drop-out mechanism is independent of both unobserved and observed data, and random drop-out (RD) if it is not dependent on the current value of the response given values of the previous responses. A drop-out mechanism that is dependent on the current value of response is termed as a non-random drop-out (NRD)(vide also little, 1995).

In this paper we focus on random dropouts which occur when the missingness depends only on the observed responses and fully observed covariates in the model. We shall apply our transition model to Fluvoxamine data where the response of interest involves the side effects of using Fluvoxamine.

Several methods exist for the analysis of such data when the missing mechanism is classified as RD. In the analysis of the Fluvoxamine data, Molenberghs et al. (1997) found that the probability of dropping out at a specific time depends strongly on the previous observed level of the side effect, such that the assumption of RD is

reasonable for this data set. They used a marginal model for the responses with a multivariate Dale distribution to take into account the correlation between the responses of the same individual, and logistic regression for the dropout process in a nonignorable dropout model. Kenward et al. (1994) and Molenberghs and Lesaffre (1994) make RD assumption in their analyses of the Fluvoxamine data. The former fitted a marginal proportional odds model that relates the degree of side effects to various covariates and used two methods of estimation: generalized estimating equations (GEE) and maximum likelihood. The latter used marginal modeling of correlated ordinal data by using a multivariate Placket distribution. Hines and Hines (2005) introduced a number of methods, including two maximum likelihood approaches which use marginal global odds ratios to model the association between responses, weighted and unweighted generalized estimating equations by using Cholesky-decomposed standardized residuals to model the association structure, and another three extended models developed for longitudinal binary data in which the association structures are modeled using either Gaussian estimation, multivariate normal estimating equations or conditional residuals. However, the above mentioned models use the marginal approach and most of them need sophisticated statistical computational methods, which are not easy to use for medical practitioners to estimate the model parameters. Hence, there are many studies of the Fluvoxamine data but as yet transition models have not been used to analyze these data.

In this paper to examine the influence of Markovian assumption on the covariate effects, we present a latend transition models which given each level of previous response, has its own dynamic parametrization. We can test stationarity effect of some covariates as well as having the same effect of some others through different levels of previous outcome. We use this model instead of marginal or random effects models because, having conditioned on previous responses, this model would remove the effects of cluster-level covariates and can provide more insight into the data generating processes than the other models can do. The software R is used to optimize the like-

lihood and to find the most parsimonious model. However, for the general form of the model (dynamic parametrization of the model for each level of previous response), one can use available software such as SPSS or STATA to find the parameter estimates.

In Section 2, Fluvoxamine data (Molenberghs et al., 1997) are discussed. In Section 3 an ordered transition latend variable model with missing at random dropout is given and a maximum likelihood approach for estimating parameters is discussed. In Section 4 the results of using a transition model for the Fluvoxamine data, are presented. In Section 5 some conclusion are discussed.

2 The Fluvoxamine data

These data have been extracted from a multicentre study in which the drug Fluvoxamine was used for 315 subjects who were initially recruited into study. The data are discussed in Molenberghs and Lesaffre (1994) and Molenberghs et al. (1997). This example is a 4-wave study. At an initial visit baseline covariate information such as the sex, age, initial severity (scale 1 to 7), and duration of actual mental illness were measured for each subject. After recruitment, the individuals made four visits at weeks 2, 4, 8 and 12. At each visit side effects is measured as an ordinal response with categories: (0) no side effect, (1) no significant interference with functionality of patient, (2) significant interference with functionality of patient and (3) side-effect surpasses therapeutic effect. Similar to Hines and Hines (2005), and for assuming the assumption of random dropout, we shall use the data on three visits at weeks 2, 4 and 12. From the initially recruited subjects, 14 were not observed at all after the start, 2 had non-monotone missing responses and 12 more were missed for at least one of the covariates age, duration or initial severity of the mental illness, leaving 287 in the analysis. Since category 3 of the response variable (side effect) occurred only 2 percent of the time, we decided to combine it with category 2 for computational simplicity. Three of different patterns of missing data that occur in these data are presented in Table I.

Table I. Different patterns of missing data for longitudinal 3-period Fluvoxamine data ("O"=observed, "M"= missed).

Pattern	Frequency
OOO	216
OOM	43
OMM	28

3 Ordered transition latend variable model

Consider the following general latend variable model for analyzing Fluvoxamine data:

$$Y_{i1}^* = X_{i1}'\beta_1 + \varepsilon_{i1} \quad (1)$$

$$Y_{it}^*|y_{it-1} = X_{it}'\beta_{y_{it-1},t} + \varepsilon_{it} \quad t = 2,3, \quad y_{it-1} = 0,1,2 \quad (2)$$

where Y_{it}^* is for $t = 1,2,3$ are latend variables for responses such that if $Y_{i1}^* \leq \theta_{1,1}$ then $Y_{i1} = 0$, if $\theta_{1,1} < Y_{i1}^* \leq \theta_{2,1}$ then $Y_{i1} = 1$ and if $Y_{i1}^* > \theta_{2,1}$ then $Y_{i1} = 2$ and for $t = 2,3$

$$Y_{it} = 0, if$$

$Y_{it}^* \leq \theta_{y_{it-1},1,t}$;1, if $\theta_{y_{it-1},1,t} < Y_{it}^* \leq \theta_{y_{it-1},2,t}$,2, if $Y_{it}^* > \theta_{y_{it-1},2,t}$.

$$where$$

$\theta_{1,1}$ and $\theta_{2,1}$ $(\theta_{1,1} < \theta_{2,1})$ and $\theta_{y_{it-1},1,t}$ and $\theta_{y_{it-1},2,t}$ $(\theta_{y_{it-1},1,t} < \theta_{y_{it-1},2,t}$ for $t = 2,3)$ are cutpoints to be estimated. Y_{it} indicates a response variable which can take C ordered categorical values, labelled $0,1,...,C-1$ ($C = 3$ for Fluvoxamine data). X_{it}'s for $t = 1,2,3$ are vectors of explanatory variables. $\beta_1 = (\beta_{1,1},...,\beta_{k,1})$ and $\beta_{y_{it-1},t} = (\beta_{1,y_{it-1},t},...,\beta_{k,y_{it-1},t})$ for $y_{it-1} = 0,1,2$ and $t = 2,3$ are the vector of regression coefficients for the explanatory variables which must be estimated. The independent error terms can be chosen to have any distribution. We shall assume independent logistic distributions for them.

We shall use a first order transition model based on the cumulative logit framework. The form of this model for three response variables and by using model (1) and (2) is:

$$\begin{aligned} logit[P(Y_{i1} \leq y_{i1})] &= \theta_{y_{i1}+1,1} - X'_{i1}\beta_1 \quad y_{i1} = 0, 1 \qquad (3) \\ logit[P(Y_{it} \leq y_{it}|Y_{it-1} = y_{it-1})] &= \theta_{y_{it-1},y_{it}+1,t} - X'_{it}\beta_{y_{it-1},t}, \\ & \quad y_{it-1} = 0, 1, 2,\ y_{it} = 0, 1,\ t = 2, 3 \quad (4) \end{aligned}$$

We can rewrite (3) and (4) to have some covariate effects similar in different times or in different previous state of response. In this paper we would like to test such stationarities. The general model given in (1) and (2) can be analyzed using existing software. However, considering the same parameters in different times, existing software can no more be used. In this situation, by considering the obtained likelihood, parameters shall be estimated by using optimizing algorithms. A likelihood ratio test (in view of deviance) can be used to test for different assumptions on parameters in models (1) and (2).

The general form of the likelihood function that can be partitioned to estimate model parameters is:

$$L = \prod_{i=1}^{n} P(Y_{i1} = y_{i1}|x_{i1}) \prod_{i=1}^{n_1} P(Y_{i2} = y_{i2}|Y_{i1} = y_{i1}, x_{i2})$$
$$\prod_{i=1}^{n_2} P(Y_{i3} = y_{i3}|Y_{i2} = y_{i2}, x_{i3}) = \prod_{i=1}^{n} P(Y_{i1} = y_{i1}|x_{i1})$$
$$\prod_{a=0}^{C-1} \prod_{i \epsilon A} P(Y_{i2} = y_{i2}|Y_{i1} = a, x_{i2}) \prod_{b=0}^{C-1} \prod_{i \epsilon B} P(Y_{i3} = y_{i3}|Y_{i2} = b, x_{i3})$$

where $A = \{i, Y_{i1} = a, i = 1, ..., n_1\}$, $B = \{i, Y_{i2} = b, i = 1, ..., n_2\}$, n_1 is the number of individuals with all y_1 and y_2 observed, n_2 is the number of individuals with all y_1, y_2 and y_3 observed and n is the

number of all individuals. We have

$$\begin{aligned} P(Y_{i1}=y_{i1}) &= P(\theta_{y_{i1},1}<Y_{i1}^{*}\leq\theta_{y_{i1}+1,1}) \\ &= P(Y_{i1}^{*}\leq\theta_{y_{i1}+1,1})-P(Y_{i1}^{*}\leq\theta_{y_{i1},1}) \\ &= \frac{e^{\theta_{y_{i1}+1,1}-X_{i1}'\beta_1}}{1+e^{\theta_{y_{i1}+1,1}-X_{i1}'\beta_1}}-\frac{e^{\theta_{y_{i1},1}-X_{i1}'\beta_1}}{1+e^{\theta_{y_{i1},1}-X_{i1}'\beta_1}} \quad y_{i1}=0,1 \end{aligned}$$

where $\theta_{0,1}=-\infty$ and $P(Y_{i1}=2)=1-P(Y_{i1}=0)-P(Y_{i1}=1)$. For $t=2,3$, $y_{it-1}=0,1,2$ and $y_{it}=0,1$

$$\begin{aligned} P(Y_{it}=y_{it}|Y_{it-1}=y_{it-1}) &= P(\theta_{y_{it-1},y_{it},t}<Y_{it}^{*}\leq\theta_{y_{it-1},y_{it}+1,t}|Y_{it-1}=y_{it-1}) \\ &= P(Y_{it}^{*}\leq\theta_{y_{it-1},y_{it}+1,t}|Y_{it-1}=y_{it-1}) \\ &\quad -P(Y_{it}^{*}\leq\theta_{y_{it-1},y_{it},t}|Y_{it-1}=y_{it-1}) \\ &= \frac{e^{\theta_{y_{it-1},y_{it}+1,t}-X_{it}'\beta_{y_{it-1},t}}}{1+e^{\theta_{y_{it-1},y_{it}+1,t}-X_{it}'\beta_{y_{it-1},t}}}-\frac{e^{\theta_{y_{it-1},y_{it},t}-X_{it}'\beta_{y_{it-1},t}}}{1+e^{\theta_{y_{it-1},y_{it},t}-X_{it}'\beta_{y_{it-1},t}}} \end{aligned}$$

Where for $y_{it-1}=0,1,2$, $\theta_{y_{it-1},0,t}=-\infty$ and for $t=2,3$, $P(Y_{it}=2)=1-P(Y_{it}=0)-P(Y_{it}=1)$. For general form of the model which takes into account different parameters, the above likelihood function makes possible the use of existing software to estimate model parameters, but for other models the likelihood can be programmed in R software to estimate model parameters. In the following we shall present two forms of model in equations (3) and (4) which we shall call hereafter model I and model II for the Flovoxamine data. model I for the Flovoxamine data is:

$$\begin{aligned} logit[P(Y_{i1}\leq y_{i1})] &= \theta_{y_{i1}+1,1}-Age\times\beta_{1,1}-Duration\times\beta_{2,1} \\ &\quad -Severity\times\beta_{3,1}-Gender\times\beta_{4,1} \quad y_{i1}=0,1 \\ logit[P(Y_{it}\leq y_{it}|Y_{it-1}=y_{it-1})] &= \theta_{y_{it-1},y_{it}+1,t}-Age\times\beta_{1,y_{it-1},t} \\ -Duration\times\beta_{2,y_{it-1},t}-Severity\times\beta_{3,y_{it-1},t}-Gender\times\beta_{4,y_{it-1},t} & \\ y_{it-1}=0,1,2,\ y_{it}=0,1,\ t=2,3 & \end{aligned}$$

and model II is:

$$logit[P(Y_{i1} \leq y_{i1})] = \theta_{y_{i1}+1,1} - Age \times \beta_{1,1} - Duration \times \beta_{2,1} - Severity \times \beta_{3,1} - Gender \times \beta_{4,1} \qquad y_{i1} = 0, 1$$

$$logit[P(Y_{it} \leq y_{it} | Y_{it-1} = y_{it-1})] = \theta_{y_{it-1},y_{it}+1,t} - Age \times \beta_{1,1} - Duration \times \beta_{2,t} - Severity \times \beta_{3,t} - Gender \times \beta_{4,1}$$

$$y_{it-1} = 0, 1, 2, \; y_{it} = 0, 1, \; t = 2, 3$$

where we assume the same stationary effect of age and gender on time. In the next section results, using the most parsimonious forms of these two models, will be compared.

4 Results of using two models for Fluvoxamine data

In this section we analyze the Fluvoxamine data using the two first order latend transition models I and II. Table 1 gives the results of marginal modeling for the response at week 2 (Y_1) based upon model I. Results of model I with marginal modeling in week 2 reveal that there is no effect of gender on side effects while age, duration and severity have significant effects. Thus gender can be exclude from the model. Results from marginal modeling of Y_1 show that the probability of having more severe side effects is higher for older patients and patients with longer durations of illness. These results also show that the more severe the illness, the higher the probability of having low severe side effects in week 2. The results of the transition modeling of Y_2 given Y_1 and of Y_3 given Y_2 in model I, are presented in Table 2. We start with age, duration, severity and gender as covariates in these models. Backward variable selection is used for the transition model of Y_2 given Y_1, and we find no significant effects of severity and gender on side effects in week 4; see Table 2. In Table 2, for different values of Y_1 , $\theta_{y_{i2},2}$'s are intercepts indicating the log odds of less severe, rather than more severe side effects in week 4, assuming there is no effect of any other covariates. For example, when there are no side effects in week 2 ($Y_1 = 0$), the log odds of having no side effects in week 4 (

Table 1. Results of using marginal cumulative logit model for Y_1 in model I (bold numbers significant at % 5 level).

	(Y_1)	
Par	Est.	S.E.
$\theta_{1,1}$	-0.832	0.820
$\theta_{2,1}$	1.356	0.824
$\beta_{1,1}(Age)$	**0.019**	**0.009**
$\beta_{2,1}(Duration)$	**0.016**	**0.005**
$\beta_{3,1}(Severity)$	**-0.301**	**0.144**

Table 2. Results for transition model of Y_2 given different values of Y_1 and Y_3 given different values of Y_2 in model I(bold numbers significant at % 5 level).

	$Y_2\|Y_1=0$			$Y_2\|Y_1=1$			$Y_2\|Y_1=2$	
Par	Est.	S.E.	Par	Est.	S.E.	Par	Est.	S.E.
$\theta_{0,1,2}$	**1.976**	**0.285**	$\theta_{1,1,2}$	0.828	0.699	$\theta_{2,1,2}$	**-1.435**	**0.498**
$\theta_{0,2,2}$	-	-	$\theta_{1,2,2}$	**4.908**	**0.912**	$\theta_{2,2,2}$	-0.154	0.393
$\beta_{1,0,2}(Age)$	-	-	$\beta_{1,1,2}$	**0.037**	**0.016**	$\beta_{1,2,2}$	-	-
$\beta_{2,0,2}(Duration)$	-	-	$\beta_{2,1,2}$	**0.027**	**0.009**	$\beta_{2,2,2}$	-	-
	$Y_3\|Y_2=0$			$Y_3\|Y_2=1$			$Y_3\|Y_2=2$	
$\theta_{0,1,3}$	**2.126**	**0.293**	$\theta_{1,1,3}$	-0.323	0.250	$\theta_{2,1,3}$	-	-
$\theta_{0,2,3}$	**4.796**	**1.004**	$\theta_{1,2,3}$	**3.440**	**0.601**	$\theta_{2,2,3}$	1.253	0.802
$\beta_{2,0,3}(Duration)$	-	-	$\beta_{2,1,3}$	**0.028**	**0.014**	$\beta_{2,2,3}$	-	-

$Y_2 = 0$) is 1.976, which translates into a probability of 0.878 for having no side effects in week 4. In other words, if a patient shows no sign of side effects in week 2, (s)he will have a high probability (0.878) of having no side effects in week 4.

When there is significant interference with functionality in week 2 ($Y_1 = 2$), the log odds of having no side effects in week 4 ($Y_2 = 0$) are reduced to -1.435, translating into a low probability of 0.192 for having no side effects in week 4. When a patient has no side effects ($Y_1 = 0$) or has significant interference with functionality ($Y_1 = 2$) in week 2, there are no significant effects of age, sex, duration and initial severity on side effects in week 4. But, for no significant interference with functionality in week 2 ($Y_1 = 1$), there are significant effects of age and duration of illness on side effects in week 4. In this

Table 3. Results of using marginal cumulative logit model for Y_1 in model II (bold numbers significant at % 5 level).

	(Y_1)	
Par	Est.	S.E.
$\theta_{1,1}$	-0.959	0.800
$\theta_{2,1}$	1.221	0.800
$\beta_{1,1}(Age)$	**0.016**	**0.006**
$\beta_{2,1}(Duration)$	**0.016**	**0.005**
$\beta_{3,1}(Severity)$	**-0.301**	**0.145**

Table 4. Results for transition model of Y_2 given different values of Y_1 and Y_3 given different values of Y_2 in model II (bold numbers significant at % 5 level).

	$Y_2\|Y_1=0$			$Y_2\|Y_1=1$			$Y_2\|Y_1=2$	
Par	Est.	S.E.	Par	Est.	S.E.	Par	Est.	S.E.
$\theta_{0,1,2}$	**2.792**	**0.404**	$\theta_{1,1,2}$	-0.153	0.499	$\theta_{2,1,2}$	-0.482	0.584
$\theta_{0,2,2}$	-	-	$\theta_{1,2,2}$	**3.597**	**0.485**	$\theta_{2,2,2}$	0.800	0.505
$\beta_{1,1}(Age)$	**0.016**	**0.006**	$\beta_{1,1}$	**0.016**	**0.006**	$\beta_{1,1}$	**0.016**	**0.006**
$\beta_{2,2}(Duration)$	**0.014**	**0.007**	$\beta_{2,2}$	**0.014**	**0.007**	$\beta_{2,2}$	**0.014**	**0.007**
	$Y_3\|Y_2=0$			$Y_3\|Y_2=1$			$Y_3\|Y_2=2$	
$\theta_{0,1,3}$	**2.787**	**0.397**	$\theta_{1,1,3}$	0.133	0.356	$\theta_{2,1,3}$	-	-
$\theta_{0,2,3}$	**5.459**	**1.039**	$\theta_{1,2,3}$	**3.729**	**0.591**	$\theta_{2,2,3}$	**2.048**	**0.870**
$\beta_{1,1}(Age)$	**0.016**	**0.006**	$\beta_{1,1}$	**0.016**	**0.006**	$\beta_{1,1}$	**0.016**	**0.006**

case, the odds of observing more severe side effects increases with age and duration.

In this Table severity, gender and age were not significant in the transition model of Y_3 given different values of Y_2. When Fluvoxamine is used in week 4 for a patient with no significant effect on functionality ($Y_2 = 1$), there is a significant effect of duration of illness on the response in week 12. In this case, the odds of observing more severe side effects increases with duration of illness.

In Table 2 the probability of $Y_3 = 0$ given $Y_2 = 0$ is 0.893. This means that if a patient has no side effects in week 4, we expect with high probability (0.893) that (s)he will have no side effects in week 12. The probability of $Y_3 \leq 1$ given $Y_2 = 0$ is 0.992 which means that if a patient has no side effects in week 4 , then with 0.992 chance, (s)he would have no side effects at all or no significant interference with functionality.

Table 3 gives the results of marginal modeling for the response at week 2 (Y_1) based upon model II. Also the results of the transition modeling of Y_2 given Y_1, and Y_3 given Y_2 are presented in Tables 4. In this model we assume that parameters of age and gender have time-invariant effects on side effects and their effects are invariant on each state of previous response (for instance for age, in equations for model II we impose $\beta_{1,1} = \beta_{1,0,2} = \beta_{1,1,2} = \beta_{1,2,2} = \beta_{1,0,3} = \beta_{1,1,3} = \beta_{1,2,3}$) and the effects of duration and severity of illness at each level of previous response are the same but may change on time (for instance for duration of illness we impose $\beta_{2,0,2} = \beta_{2,1,2} = \beta_{2,2,2} = \beta_{2,2}$). By considering these assumptions, we see that there is no significant effect of gender in model II and so it can be removed from the model. Also in this model backward variable selection is used and we find that there is no significant effect of severity on side effects in week 4 and 12 and no significant effect of duration in week 12.

A comparison of results in Tables 2 and 4 reveals that by using model I, age and duration have significant effect on side effects at time 2 only when $Y_1 = 1$. But model II shows that duration and age are significant for all levels of Y_1. So model I gives more insight into the processes that generated data. Hence, model II misleads us to think that some covariates, which are not significant considering the general model, are significant.

Comparing the results in Table 2 with the corresponding results of Table 4 we can also see that by model I the duration effect is weekly significant on side effects at time 3 when $Y_2 = 1$, While, model II shows that this effect is not significant and only age is significant.

Results show that although the model I has more parameters than model II but the effects of covariates were determined in a right way in this model. For testing the same effect of covariates at different levels of previous outcome or having stationary effect of some covariates we compare two models by using $G^2 = -2(logL_{Full_{II}} - logL_{Full_I}) = 4638.874 - 4626.508 = 12.366$ (where

$L_{Full_{II}}$ and L_{Full_I} are likelihoods evaluated at the ML estimates for model II and model I, respectively) which has an approximate chi-square distribution (under the assumption that model II is the true model) with d.f.=2. All this shows that model I is a good fit to these data.

5 Discussion

In this paper a latend transition model was presented for ordinal categorical longitudinal data with random dropout over short periods. Two versions of This model (model I and model II) were applied to the Fluvoxamine data and their results are compared. In model I the effects of parameters are different in marginal and transition models in each time as well as at each level of previous outcomes to make possible the use of existing software to estimate model parameters. In model II some parameters are assumed to be similar in different times or in each previous state of response. In model II the software R is used to optimize the obtained likelihood for the model while in model I existing software such as SAS and STATA were used to estimate parameters. By comparing two models we conclude that model II is not good enough for fitting to these data.

References

[1] Agresti, A. (2002). Analysis of categorical data. John Wiley and Sons. New York.

[2] Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about markov Chains. *Ann. Math. Stat.,* **28**,89-110.

[3] Diggle, P. J., Heagerty, P., Liang, K. Y., Zeger, S. L. (2002). Analysis of longitudinal data. Oxford, University Press.

[4] Diggle, P. J. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Appl. Statist.,* **43**,49-93.

[5] Francom, S. F., Chuang-Stein, C., Landis, J. R. (1989). A log-linear model for ordinal data to characterize differential change among treatments. *Stat. Med.,* **8**,571-582.

[6] Ganjali, M. and Rezaee, Z. (2007)A transition model for analysis of repeated measure ordinal response data to identify the effects of different treatments. *Dr. Inf. J.*, **41**, 527-534.

[7] Harvile, D. A. and Mee, R. W. (1984). A mixed model procedure for analyzing ordered categorical data. *Biometrics,* **40**,393-408.

[8] Hines, R. J. and O'Hara Hines, W. G. S. (2005). An appraisal of methods for the analysis of longitudinal categorical data with MAR drop-outs. *Stat. Med.,* **24**,3549-3563.

[9] Kenward, MG., Lesaffre, E., Molenberghs, G. (1994). An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random.*Biometrics*, **50**, 945-953.

[10] Kim, K. (1995). A bivariate cumulative probit regression model for ordered categorical data. *Stat. Med.,* **14**,1341-1352.

[11] Liang, KY., Zeger, SL., Qaqish, BF. (1992). Multivariate regression analyzes for categorical data (with discussion). *J. Roy. Statist. Soc. B,* **54**,3-40.

[12] Little, R. J. A. (1995). Modelling the drop-out mechanism in repeated-measures studies. *J. Am. Statist. Assoc.,* **90**,1112-1121.

[13] Little, R. J. A and Rubin, D. B. (2002). Statistical analysis with missing data. New York. John Wily.

[14] McCullagh, P. (1980). Regression models for ordinal data (with discussion). *J. Roy. Statist. Soc. B,* **42**,109-142.

[15] Molenberghs, G., Kenward, M.G., Lesaffre, E. (1997). The analysis of longitudinal ordinal data with nonrandom drop-out. *Biometrika,* **84**,33-44.

[16] Molenberghs, G. and Lesaffre, E. (1994). Marginal modelling of correlated ordinal data using a multivariate Placket distribution. *J. Am. Statist. Assoc.,* **89**,633-644.

[17] Ten Have, T. R., Landis, J. R., Hartzel, J. (1996). Population-average and cluster specific models for clustered ordinal response data. *Stat. Med.,***15**,2573-2588.

[18] Tutz, G. (2005). Modelling of repeated ordered measurements by isotonic sequential regression. *Stat. mod.,* **5**,269-287.

[19] Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *J. Am. Statist. Assoc.,* **91**,217-221.

Archimedean Copula and Multivariate Hazard rate Ordering

M. Rezapour, and M. H. Alamatsaz

Department of Statistics, University of Isfahan

Abstract: Structure analysis of a distribution function with respect to the notions of dependence is a desirable problem, but this task maybe impossible in general cases. Some papers have considered relation between two archimedean copulas with respect to multivariate stochastic ordering. In this paper we present some new version of archimedean copula which is constructed through the mixing of previous version. These copulas use more generator functions than previous copulas and thus they are more efficient in complex models.

Dynamic version of hazard rate ordering is an important concept that is used in reliability problems. We discuss some results that compare the new version of Archimedean Copulas with respect to Dynamic version of multivariate hazard rate ordering under some suitable conditions. Finally, we shall provide some illustrative examples.

Keywords: Multivariate aging, MTP2, Multivariate hazard rate ordering, Multivariate stochastic ordering

1 Introduction

Multivariate concept of Copula is an important notion which is used in finance and insurance analysis. Particular highlighting is cited on Archimede-
an Copulas, since it is convenient to put them into practice, and fairly flexible and appropriate for a variety of distributions. Archimedean Copulas have recently been the point of focus for their convenient application in fitting some models. Some notions of ordering and aging are also discussed in this class. This class has a very close relation with laplace transforms which was noticed by Marshall et al.(1988). They considered association and total positivity for this class. More precise statement about Archimedean Copulas have been obtained by Genest and Rivest (1993) and Nelsen (1999). Various relations between Archimedean Copula and positive dependence were

discussed by Muller and Scarsini (2003). Suppermodular ordering was described under condition of Archimedean Copula by Hu and Wei (2005). Some relations between bivariate aging and generating functions of an Archimedean Copula were found by Averous and Dortet-Bernadet (2004). Notion of aging is an important Notion of Reliability. Two famous notion of aging are Hazard rate and Likelihood ratio. Dynamic version of multivariate hazard rate ordering considered by Shaked and Shantikumar (1994) another extension to multivariate case considered by Hu et. al. (2003). In this paper we obtain some results concerning relations between two random vectors with respect to dynamic version of multivariate hazard rate ordering. Then, using these result, we extend Hu and Wei (2005)'s result to compare the new version of Archimedean copula with respect to dynamic version of multivariate hazard rate ordering .

Let us first define some concepts of ordering and the notion of Multivariate Aging. Recall that a *Copula* is a multivariate distribution function with all univariate margins being standard uniform distributions. Let $\hbar$ be the class of all multivariate distribution functions s.t. if $F \in \hbar(F_1, ..., F_n)$ then F is a n-variate distribution with i*th* univariate marginal F_i for $1 \leq i \leq n$. The *Copula Associated* with F is a distribution function $C : [0,1]^n \mapsto [0,1]$ that satisfied:

$$F(\mathbf{x}) = C(F_1(x_1), ..., F_n(x_n)) \tag{1}$$

If F is a continuous distribution function, $F \in \hbar(F_1, ..., F_n)$ and quintile function $F_1^{-1}, ..., F_n^{-1}$, then the above equality becomes :

$$C(\mathbf{u}) = F(F_1^{-1}(u_1), ..., F_n^{-1}(u_n)), \qquad \mathbf{u} = (u_1, ..., u_n) \in [0,1]^n$$

Now we make the first step on this section with defining a concept that has very intimate relation with laplace transforms. A function $\psi : \Re_+ \mapsto [0,1]$ is called *d-alteranating* if $(-1)^k \psi^{(k)} \geq 0$ for $k \in \{1, ..., d\}$. If a function is d-alternating for all $d \in \aleph$, we call it *completely monotone* . If a Copula C_ψ has the following form, it is called Archimedean Copula :

$$C_\psi(x_1, ..., x_d) = \psi\left(\sum_{i=1}^{n} \psi^{-1}(x_i)\right), \tag{2}$$

where $\psi : \Re_+ \mapsto [0,1]$ is a d-alternating($d \geq 2$) function such that $\psi(0) = 1$, and $\lim_{x\to\infty} \psi(x) = 0$. We call the ψ the generator function of the Copula. (For more details see Joe(1997) or Nelson(1999))

We could extend the above definition for general cases. For example if we suppose that ϕ_1 , ϕ_2 ,ϕ_3 , $\phi_1^{-1}o\phi_2$ and $\phi_1^{-1}o\phi_3$ were completely monotone functions, then

$$C(\mathbf{u}) = \phi_1 \left(\phi_1^{-1} o \phi_2 \Big(\sum_{i=1}^{k} \phi_2^{-1}(u_i) \Big) + \phi_1^{-1} o \phi_3 \Big(\sum_{i=k+1}^{n} \phi_3^{-1}(u_i) \Big) \right) \quad (3)$$

is an Archimedean Copula. In addition if we assume ϕ_4 , and $\phi_3^{-1}o\phi_4$ are completely monotone functions, then

$$\begin{aligned} C(\mathbf{u}) &= \phi_1 \Big(\phi_1^{-1} o \phi_2 \Big(\phi_2^{-1} o \phi_3 \Big(\phi_3^{-1} o \phi_4 \Big(\sum_{i=1}^{k_1} \phi_4^{-1}(u_i) \Big) \\ &+ \sum_{i=k_1+1}^{k_2} \phi_3^{-1}(u_i) \Big) + \sum_{i=k_2+1}^{k_3} \phi_2^{-1}(u_i) \Big) + \sum_{i=k_3+1}^{n} \phi_1^{-1}(u_i) \Big) \quad (4) \end{aligned}$$

is an Archimedean Copula. Also

$$\begin{aligned} C(\mathbf{u}) &= \phi_1 \Big(\phi_1^{-1} o \phi_2 \Big(\sum_{i=1}^{k_1} \phi_2^{-1}(u_i) \Big) \\ &+ \phi_1^{-1} o \phi_3 \Big(\phi_3^{-1} o \phi_4 \Big(\sum_{i=k_1+1}^{k_2} \phi_4^{-1}(u_i) \Big) + \sum_{i=k_2+1}^{k_3} \phi_3^{-1}(u_i) \Big) \Big) \quad (5) \end{aligned}$$

is an Archimedean Copula if we assume that $\phi_2^{-1}o\phi_3$ is a completely monotone function instead of $\phi_1^{-1}o\phi_3$ in the above assumption. Obviously above definition of copula has more parameters than the previous version. Therefore, these copulas are very useful in modelling dependent data. On this framework, we desire to investigate their ordering in the sense of aging concept. Consequently, we need to introduce dynamic ordering concepts and then uncover some relations between these notions and Archimedean Copulas.

We say that $\phi : \Re^n \to \Re$ is a suppermodular function if for every $\mathbf{x}, \mathbf{y} \in \Re^n$ it satisfies

$$\phi(\mathbf{x}) + \phi(\mathbf{y}) \leq \phi(\mathbf{x} \vee \mathbf{y}) + \phi(\mathbf{x} \wedge \mathbf{y})$$

where

$$(\mathbf{x} \vee \mathbf{y}) = (max(x_1, y_1), max(x_2, y_2), ..., max(x_n, y_n))$$
$$(\mathbf{x} \wedge \mathbf{y}) = (min(x_1, y_1), min(x_2, y_2), ..., min(x_n, y_n)).$$

Let X be a random variable with density function f. Then we say X is multivariate totally positive of order 2 (MTP2) if $log(f)$ is a suppermodular function. The concept of multivariate aging plays an important roll in the Theory of Reliability . In this paper, we consider a kind of extension of the univariate concept that has been described by Shaked and Shanthikumar (1991). Suppose that X and Y are two vectors. We say $X \leq Y$ iff $x_i \leq y_i$ for $i = 1, ..., n$. Let $\ell = (\ell_1, ..., \ell_n)$ be a nonnegative random vector with an absolutely continuous distribution function. The coordinations of ℓ will be contemplated of the lifetimes of n devices. To define the concept of aging, we need a history of the life time. We exhibit this history with $\Im_t$ which demonstrates the list of devises that fail and their failure time as below:

$$\Im_t = \{\ell_I = x_I, \ell_{\overline{I}} > te\}$$

where $I = \{i_1, ...i_r\} \subset \{1, ..., n\}$, $\overline{I}$ is the complement, e is vector of units with suitable size. Put $X^+ = max(0, X)$. Occasionally two histories $\Im_t$, and $\Im'_{t'}$, (where $t \leq t'$) might be compared by their 'severeness'. We say that $\Im$ is less severe than $\Im'$ (indicated by $\Im \prec \Im'$) if in excess of the interval time $[0, t]$, the set of elements which fail in $\Im_t$, is a subset of the set of elements which fail in $\Im'_{t'}$ and for elements which fail by time t in both histories, the failures in $\Im_t$, happened not before the failures in $\Im'_{t'}$, More clearly, $\Im_t \prec \Im'_{t'}$, for $t \leq t'$, iff

$$\Im_t = \{\ell_I = x_I, \ell_{\overline{I}} > te\} \quad , \qquad \Im'_{t'} = \{\ell_I = x'_I, \ell_J = x_J, \ell_{\overline{I \cup J}} > t'e\} \quad (6)$$

for two disjoint subset I and J of $\{1, ..., n\}$, $0 \leq x'_I \leq x_I \leq te$ and $x_J \leq t'e$.. We define the multivariate hazard rate at time t as below.

$$\zeta_i(\Im_t) = lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P[t < T_i \leq t + \Delta t | \Im_t]$$

The above term is the probability of instantaneous failure of element i, given the history $\Im_t$. By above definition we could define the multivariate hazard rate order as below.

Let X and Y be two n-variate random vectors with cumulative distribution function F and G and hazard rate functions $\zeta_.(.), \xi_.(.)$. Then, we say that X is greater than Y with respect to Multivariate hazard rate ordering(denoted by $X \succeq_h Y$ or $F \succeq_h G$) if

$$\zeta_i(\ell_t) \leq \xi_i(\ell'_t), \qquad whenever \quad \ell_t \leq \ell'_t \tag{7}$$

for all $t > 0$ and component k which has not failed by time t in ℓ'_t.

2 Archimedean Copula and aging

In this section we describe the relations between notions of ordering that were initiated previously and Archimedean Copulas. Our result could be used for classification and characterization of some Copulas with respect to notions of stochastic ordering and its relation with a certain class of Copulas. We also reveal a relation between Archimedean Copula and multivariate aging concept. Hu et. al.(2003) considered the relation between two vectors with respect to Hazard Rate Ordering. We represent the following theorem with respect to Dynamic version of Multivariate Hazard Rate Ordering.

Theorem 1 *Let $X = (X_1, X_2, ..., X_n)$ be random a vector and $Y = (Y_1, Y_2, ..., Y_n)$ be a vector of independent random variables such that $X_i =^d Y_i; i = 1, 2, ..., n$. If X is MTP2 then $Y \prec_h X$.*

Proof. Let $\Im_t$ and $\Im'_t$ be as in (6). According to (7) we should prove:

$$\begin{aligned} &\frac{Pr(X_I = x_I, X_i = t, X_{L_1} > te)}{Pr(X_I = x_I, X_{\bar{I}} > te)} \\ \leq &\frac{Pr(Y_I = x'_I, Y_J = x_J, X_i = t, X_{L_2} > te)}{Pr(Y_I = x'_I, Y_J = x_J, X_{\overline{I \cup J}} > te)} \end{aligned} \tag{8}$$

where $I = \{i_1, ..., i_r\}, J = \{j_1, ..., j_p\}$ are two disjoint subsets of $\{1, ..., n\}$, $L_1 = \{\bar{I} - \{i\}\} = \{l_1, ..., l_q\}$ and $L_2 = \{\overline{I \cup J} - \{i\}\} = \{k_1, ..., k_g\}$. But the right hand side of (8) equals to:

$$\frac{\prod_{j \in I} Pr(Y_j = x'_j) \prod_{j \in J} Pr(Y_j = x_j) Pr(X_i = t) \prod_{j \in L_2} Pr(X_j > t)}{\prod_{j \in I} Pr(Y_j = x'_j) \prod_{j \in J} Pr(Y_j = x_j) \prod_{j \in \overline{I \cup J}} Pr(X_j > t)}$$

thus we must have

$$\frac{Pr(X_I = x_I, X_i = t, X_{L_1} > te)Pr(X_i > t)}{Pr(X_I = x_I, X_{\bar{I}} > te)Pr(X_i = t)} \leq 1$$

or

$$\frac{Pr(X_I = x_I, X_{L_1} > te|X_i = t)}{Pr(X_I = x_I, X_{L_1} > te|X_i > t)} \leq 1$$

Now, if we let $\varpi = \{X_I = x_I, X_{L_1} > te\}$ then we should prove that

$$\int_t^{+\infty} [Pr(\varpi, X_i = t')Pr(X_i = t) - Pr(\varpi, X_i = t)Pr(X_i = t')]dt' \geq 0$$

But the integral can be written as bellow:

$$\int_t^{+\infty} \int_t^{+\infty} ... \int_t^{+\infty} [Pr(X_{i_1} = x_{i_1}, ..., X_{i_r} = x_{i_r},$$
$$X_i = t', X_{j_1} = t'_1, ..., X_{j_p} = t'_p)$$
$$\times lim_{a\to+\infty} Pr(X_1 = a, ..., X_{i-1} = a, X_i = t, X_{i+1} = a, ..., X_n = a)$$
$$-Pr(X_{i_1} = x_{i_1}, ..., X_{i_r} = x_{i_r}, X_i = t, X_{j_1} = t'_1, ..., X_{j_p} = t'_p)$$
$$\times lim_{a\to+\infty} Pr(X_1 = a, ..., X_{i-1} = a, X_i = t', X_{i+1}$$
$$= a, ..., X_n = a)]dt'dt'_1...dt'_k$$

Thus, since X is MPT2, the interior of the integral is positive and hence we have the result.•

Now we need to reveal some properties of multivariate hazard rate ordering that we will use in the next section.

Theorem 2 *(I)(concatenation). Let* $(X_1, ..., X_m) \prec_h (Y_1, ..., Y_m)$ *and* $(X'_1, ..., X'_n) \prec_h (Y'_1, ..., Y'_n)$. *Suppose that* $(Y_1, ..., Y_m)$ *is independent of* $(Y'_1, ..., Y'_n)$, *and* $(X_1, ..., X_m)$ *is independent of* $(X'_1, ..., X'_n)$. *Then*

$$(X_1, ..., X_m, X'_1, ..., X'_n) \prec_h (Y_1, ..., Y_m, Y'_1, ..., Y'_n) .$$

(II) (closure under increasing transform). $X \prec_h Y$ *implies that* $(g_1(X_1), ...$ $,g_m(X_m)) \prec_h (g_1(Y_1), ..., g_m(Y_m))$ *for all increasing functions* $g_i : \Re \mapsto \Re; i = 1, ..., m$.

(III) (closure under mixture). Suppose that F_θ and G_θ are m-variate distribution functions for all $\theta \in \Theta$, and that $F_\theta \prec_h G_\theta$ for all $\theta \in \Theta$. If $H(\theta)$ is a distribution on Θ, and F, G are given by

$$F(x) = \int_\Theta F_\theta(x) dH(\theta) \qquad G(x) = \int_\Theta G_\theta(x) dH(\theta)$$

then $F \prec_h G$.

Proof. For part (I) if $\zeta(.), \zeta_1(.), \zeta_2(.)$ and $\xi(.), \xi_1(.), \xi_2(.)$ are the hazard rate functions of $(X_1, ..., X_m, X'_1, ..., X'_n)$,$(X_1, ..., X_m)$, $(X'_1, ..., X'_n)$ and $(Y_1, ..., Y_m$
$,Y'_1, ..., Y'_n)$, $(Y_1, ..., Y_m), (Y'_1, ..., Y'_n)$ respectively, then we have $\zeta(.) = \zeta_1(.) \times \zeta_2(.)$ and $\xi(.) = \xi_1(.) \times \xi_2(.)$.
Part (II) obviously holds since $g_i(.)$ is increasing and preserves the inequality.
For part (III) suppose $(X|\theta)$ and $(Y|\theta)$ have hazard rate functions $\zeta(.)$ and $\xi(.)$ respectively, then under assumption of (7) we could write $\zeta_i(\ell_t) = lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} E[I_{(t<X_i \leq t+\Delta t)}(X_i)|\ell_t]$ and $\xi_i(\ell'_t) = lim_{\Delta t \downarrow 0} \frac{1}{\Delta t}$
$E[I_{(t<X_i \leq t+\Delta t)}(Y_i)|\ell'_t]$. Now let X and Y have hazard rate functions $\zeta^*(.)$ and $\xi^*(.)$ thus we have

$$\begin{aligned} \zeta_i^*(\ell_t) &= E(lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} E[I_{(t<X_i \leq t+\Delta t)}(X_i)|\ell_t]|\theta) \\ &\leq E(lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} E[I_{(t<X_i \leq t+\Delta t)}(Y_i)|\ell'_t]|\theta) = \xi_i^*(\ell'_t). \bullet \end{aligned}$$

Now, we state a result due to Muller and Scarsini (2003) that characterize MTP2 dependence on families of Archimedean Copulas.

Theorem 3 *If ϕ is completely monotone, then $C_\phi : [0,1]^d \mapsto [0,1]$, constructed like in (2), is MTP2 for any d.*

From theorems 2.1,2.2 and 2.3 we obtain the following corollary.

Corollary 1 *If W_i and Z_i be two sequences of independent random variables, $Z_1, ..., Z_i$ and $Z_{i+1}, ..., Z_n$ be MPT2. Suppose that $g_i : \Re^2 \to \Re$ is an increasing function on Z for constant value of W and*

$$X = (g_1(W_1, Z_1), g_2(W_2, Z_2), ..., g_n(W_n, Z_n))$$

$$T = (g_1(W_1, Z_1), g_2(W_2, Z_1), ..., g_n(W_n, Z_1))$$
$$Y = (g_1(W_1, Z_1), g_2(W_2, Z_1), ..., g_i(W_i, Z_1), g_{i+1}(W_{i+1}, Z_j), ..., g_n(W_n, Z_j))$$

where $j > i$, then we have $Y \prec_h X$ and $T \prec_h X$.

Theorem 4 *With the notions in (4), let $X \sim C_1$ and $Y \sim C_2$ where $C_1(U) = \psi_1\Big(\psi_1^{-1} o \phi_2\Big(\sum_{i=1}^{k} \phi_2^{-1}(u_i)\Big) + \psi_1^{-1} o \phi_3\Big(\sum_{i=k+1}^{n} \phi_3^{-1}(u_i)\Big)\Big)$ and $C_2(U) = \psi_1\Big(\psi_1^{-1} o \psi_2\Big(\sum_{i=1}^{k} \psi_2^{-1}(u_i)\Big) + \psi_1^{-1} o \psi_3\Big(\sum_{i=k+1}^{n} \psi_3^{-1}(u_i)\Big)\Big)$. Suppose $\psi_2^{-1} o \phi_2$ and $\psi_3^{-1} o \phi_3$ are completely monotone. Then we have $Y \prec_h X$.*

Proof. Let $u^* = (e^{x_1}, e^{x_2}, ..., e^{x_n})$, the stochastic representation of C_1 and C_2 are:

$$X_i = (g(U_{i1}, \gamma_{i1}^*), ..., g(U_{ik}, \gamma_{ik}^*), g^*(U_{i,k+1}^*, \gamma_{i,k+1}^{***}), ..., g^*(U_{im}^*, \gamma_{im}^{***}))$$

$$Y_i = (g(U_{i1}, \gamma_{i1}), ..., g(U_{ik}, \gamma_{i1}), g^*(U_{i,k+1}^*, \gamma_{i,k+1}^{**}), ..., g^*(U_{in}^*, \gamma_{i,k+1}^{**}))$$

where $g(u, \lambda) = -log\phi_2(-\frac{1}{\lambda}\log u)$, $g^*(u, \lambda) = -log\phi_3(-\frac{1}{\lambda}\log u)$ and $\{U_{ij}, V_{ij}, W_{ij}, j = 1, ..., m\}$ are random variables with standard uniform distributions. Let A_i be a random variable with distribution function $M_{\psi_1}(.)$

$$B_{ij} := M^{-1}_{\psi_1^{-1} o \psi_2}(W_{ij}; A_i) \quad , \quad \gamma_{ij}^* := M^{-1}_{\psi_2^{-1} o \phi_2}(V_{ij}; \beta_{ij})$$
$$\gamma_{ij} := M^{-1}_{\psi_2^{-1} o \phi_2}(U_{ij}; \beta_{ij}) \quad , \quad B_{ij}^* := M^{-1}_{\psi_1^{-1} o \psi_3}(W_{ij}; A_i)$$
$$\gamma_{ij}^{**} := M^{-1}_{\psi_3^{-1} o \phi_3}(W_{ij}; \beta_{ij}^*) \quad , \quad \gamma_{ij}^{***} := M^{-1}_{\psi_3^{-1} o \phi_3}(U_{ij}; \beta_{i,k+1})$$

Then, using above representation and corollary 2.4 the proof is complete.•

Remark 1 *With the assumption of (4). Let $X \sim C_1$ and $Y \sim C_2$*

where

$$\begin{aligned} C_1(U) &= \psi_1\Big(\psi_1^{-1}o\psi_2\Big(\psi_2^{-1}o\phi_3\Big(\phi_3^{-1}o\phi_4\Big(\sum_{i=1}^{k_1}\phi_4^{-1}(u_i)\Big) \\ &+ \sum_{i=k_1+1}^{k_2}\phi_3^{-1}(u_i)\Big) + \sum_{i=k_2+1}^{k_3}\psi_2^{-1}(u_i)\Big) + \sum_{i=k_3+1}^{n}\psi_1^{-1}(u_i)\Big) \\ C_2(U) &= \psi_1\Big(\psi_1^{-1}o\psi_2\Big(\psi_2^{-1}o\psi_3\Big(\psi_3^{-1}o\psi_4\Big(\sum_{i=1}^{k_1}\psi_4^{-1}(u_i)\Big) \\ &+ \sum_{i=k_1+1}^{k_2}\psi_3^{-1}(u_i)\Big) + \sum_{i=k_2+1}^{k_3}\psi_2^{-1}(u_i)\Big) + \sum_{i=k_3+1}^{n}\psi_1^{-1}(u_i)\Big) \end{aligned} \tag{9}$$

Suppose $\psi_2^{-1}o\phi_2$,$\psi_3^{-1}o\phi_3$ *and* $\psi_4^{-1}o\phi_4$ *are completely monotone. Then, we have* $Y \prec_h X$.

Remark 2 *With the assumption of (5). Let* $X \sim C_1$ *and* $Y \sim C_2$ *where*

$$\begin{aligned} C_1(U) &= \psi_1\Big(\psi_1^{-1}o\phi_2\Big(\sum_{i=1}^{k_1}\phi_2^{-1}(u_i)\Big) \\ &+ \psi_1^{-1}o\psi_3\Big(\psi_3^{-1}o\phi_4\Big(\sum_{i=k_1+1}^{k_2}\phi_4^{-1}(u_i)\Big) + \sum_{i=k_2+1}^{n}\psi_3^{-1}(u_i)\Big)\Big) \\ C_2(U) &= \psi_1\Big(\psi_1^{-1}o\psi_2\Big(\sum_{i=1}^{k_1}\psi_2^{-1}(u_i)\Big) \\ &+ \psi_1^{-1}o\psi_3\Big(\psi_3^{-1}o\psi_4\Big(\sum_{i=k_1+1}^{k_2}\psi_4^{-1}(u_i)\Big) + \sum_{i=k_2+1}^{n}\psi_3^{-1}(u_i)\Big)\Big) \end{aligned} \tag{10}$$

Suppose $\psi_2^{-1}o\phi_2$,$\psi_3^{-1}o\phi_3$ *and* $\psi_4^{-1}o\phi_4$ *are completely monotone. Therefore, we have* $Y \prec_h X$.

3 Examples

Example 3.1. A multivariate generalization of the Galambos bivariate copula is

$$C(\mathbf{u};\theta) = \exp\left\{\sum_{S\in\mathcal{F}}(-1)^{|S|}\left[\sum_{i\in S}\tilde{u}_i^{-\theta}\right]^{-1/\theta}\right\}, \theta > 0, \mathbf{u} \in (0,1)^m$$

where $\tilde{u}_i = -logu_i$ Therefore, by Theorem 2.1, $C(.;\theta_1) \prec_h C(.;\theta_2)$ for all $\theta_1 < \theta_2$.

Example 3.2. A multivariate Gumbel copula

$$C(\mathbf{u};\theta) = \exp\left\{-\left[\sum_{i=1}^{m}\tilde{u}_i^{\theta}\right]^{1/\theta}\right\}$$

which can be obtained as Archimedean copula corresponding to LT1(Laplace Transform 1)(joe 1997). It is known from Joe (1997 Theorems 4.6 and 4.7) that $C(.;\theta)$ is increasing in the sense of multivariate concordance and positive function dependence orderings with respect to θ. By Theorem 2.1, we conclude that $C(.;\theta_1) \prec_h C(.;\theta_2)$ for all $\theta_1 < \theta_2$.

References

[1] Averous J. and Dortet-Bernadet J.L. (2004) Dependence for Archimedean Copulas and Aging Properties of Their Generating Functions *Indian Statist. Institute* 66, 4,607-620.

[2] Genest C. and Rivest L.P. (1993) Statistical inference procedures for bivariate Archimedean copulas,*J. Amer. Statist. Assoc.* 88 10341043.

[3] Hu T., Khaledi B.-E.,Shaked M. (2003) Multivariate hazard rate orders *J. Multivariate Anal.* 84 173189.

[4] Hu T., Wei G. (2005) Supermodular dependence ordering on a class of multivariate copulas *Statist. Prob. Lett.* 57 , 375385.

[5] Joe, H. (1997)*Multivariate Models and Dependence Concepts.* Chapman & Hall, London.

[6] Marshall A.W. and Olkin I. (1988) Families of multivariate distributions,*J. Amer. Statist. Assoc.* 83 834841.

[7] Nelsen R.B. (1999) *An Introduction to Copulas,* Springer,New York.

[8] Shaked, M. and Shanthikumar, J. G. (1994) *Stochastic Orders and their Applications,* Academic Press, London.

[9] Muller A. and Scarsini M.(2003) Archimedean Copula and positive dependence *J. Multivariate Anal.* 28,1-12.

Point and interval estimators of fuzzy numbers

A. Saeidifar

Department of Statistics, Islamic Azad University

Abstract: In this paper, the point and interval estimators of fuzzy numbers and their important properties are introduced. This estimators play an important role for solving of fuzzy systems problems, e.g. decision-making to fuzzy systems manner. We obtain the classical forms of estimators and their possibilistic limits to the trapezoidal fuzzy numbers. Also the interval-approximations of a fuzzy number can be consider as a confidence interval, that gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample fuzzy data. Therefore we want to obtain the point and interval approximations for the fuzzy parameter by using a defuzzification of the fuzzy numbers with a deterministic possibilistic limits e.g., the mean value, median, mode, interval-valued mean, interval-valued median. The point and interval estimators of the parameters to form fuzzy numbers play an important role in the statistics distributions. In the statistics, the center of distribution of a quantitative variable is determined besides the mean value, median value and the mode value; furthermore, much information are expressed in the central and interval indices of a distribution. However, to find out these indices to approximate the fuzzy parameter of a known population in a fuzzy environment are very important and useful.

Keywords: Fuzzy numbers, Point estimation, Interval, Mean value Median; Mode.

1 Introduction

The point and interval estimators of the parameters to form fuzzy numbers play an important role in the statistics distributions. Therefore we will find a efficiency and good estimators to fuzzy parameters of a distribution in parametric and non-parametric statistics inference.

In the real world, sometimes the observations are described by qualitative terms instead of quantitative values. For example, the management achievement of a company may be qualitatively evaluated

by an expert as of extremely high standard, high standard, average standard, low standard, or extremely low standard. There is no well-defined boundary between two linguistic levels. It is inappropriate to use crisp values to represent different levels. In this case, the fuzzy set theory (Zadeh 1978) has been demonstrated to be a suitable methodology to describe the uncertain situation. In majority of practical case decision are made after the evaluation of existing information pertaining to the consider problem . This informations are in point and interval estimators of parameter so that they can help to decision-maker's for solving of fuzzy problems. Let us assume that a phenomenon of interest is described by a random variable X distributed according to a certain probability distribution belonging to a family of probability distributions indexed by a parameter . In a classical statistical setting we also assume that decisions depend entirely on the value of this parameter. If the value of were known we might take an appropriate decision without any problem. However, the value of is usually not known, and we could only formulate a point or interval estimators or a respective hypothesis about it. We assume that parameter is a fuzzy number and the main aim in this work is estimate of . Carlsson and Fuller and Majlender (2001, 2003) introduced the interval-valued possibilistic mean and variance of fuzzy number. Dubios and Bodjanova (1987, 2005) introduced interval-valued median , interval-valued mean and the central interval of a fuzzy number that can be used as a crisp approximation of a fuzzy number. Chanas (2001) has introduced the interval approximation of a fuzzy number with respect to the Haming distance. In the statistics, center of distribution of a quantitative variable is determined besides the mean value, median value and the mode value, furthermore, much informations are expressed in central and interval indexes of a distribution. Therefore we will find some of these indexes for approximate fuzzy parameter of a known population.

Let $\mathbf{R}$ be the set of all real numbers. We assume that the fuzzy

number A can be expressed for all $x \in \mathbf{R}$ in the following form:

$$A(x) = \begin{cases} g(x) & x \in [a, b), \\ 1 & x \in [b, c), \\ h(x) & x \in [c, d), \\ 0 & otherwise, \end{cases} \tag{1}$$

where a, b, c, d are real numbers so that $a < b \leq c < d$, g is a real valued function that is increasing and right continues, and h is a real valued function that is decreasing and left continues. Notice that (1) is an **L-R** fuzzy number with a strictly monotonic shape function, as proposed by Dubois and Prade in 1981 and also described in [4]. A fuzzy number A with shape functions g and h is given by:

$$g(x) = (\frac{x-a}{b-a})^n, \qquad h(x) = (\frac{d-x}{d-c})^n, \tag{2}$$

will be denoted by $A = (a, b, c, d)_n$, $n > 0$. Each fuzzy number A described by (1) has the following α-level sets$(\alpha - cuts)$:
(**a.**)$A_\alpha = [g^{-1}(\alpha), h^{-1}(\alpha)]$ for all $\alpha \in (0, 1)$,
(**b.**)$A_0 = [a, d]$,
(**c.**)$A_1 = [b, c]$.
If $A = (a, b, c, d)_n$, then for all $\alpha \in (0, 1)$.

$$A_\alpha = [a + \alpha^{1/n}(b-a), d - \alpha^{1/n}(d-c)] = [a_1(\alpha), a_2(\alpha)]. \tag{3}$$

2 Mean value and interval-valued mean

Definition 1 *(Carlson and Fuller 2003) The possibilistic mean value of fuzzy number A is define as*

$$M(A) = \int_0^1 \alpha(a_1(\alpha) + a_2(\alpha))d\alpha. \tag{4}$$

We show this point is the nearest point of fuzzy number A with respect to the following distance quantity

$$D(A, C(A)) = \left[\int_0^1 \alpha[(a_1(\alpha) - C(A))^2 + (a_2(\alpha) - C(A))^2]d\alpha\right]^{1/2}. \quad (5)$$

By minimize the above distance quantity with respect to $C(A)$ ($C(A)$ is a point of support function) we get $C(A) = M(A)$. Therefore, the following theorem holds.

Theorem 1. Let A be a fuzzy number. Then $M(A)$ is the nearest point to fuzzy number A with respect to relation (5), which is unique.

Now we rewrite eqnarray (5) as follows

$$D(A, C_1(A), C_2(A)) = \left[\int_0^1 \alpha[(a_1(\alpha) - C_1(A))^2 + (a_2(\alpha) - C_2(A))^2]d\alpha\right]^{1/2}. \quad (6)$$

By minimize the above expression with respect to $C_1(A), C_2(A)$ we get the following interval

$$[C_1(A), C_2(A)] = \left[\int_0^1 \alpha a_1(\alpha)d\alpha, \int_0^1 \alpha a_1(\alpha)d\alpha\right], \quad (7)$$

this interval is the nearest interval of fuzzy number A. we replace

$$M_1(A) = C_1(A), M_2(A) = C_2(A), IM(A) = [M_1(A), M_2(A)],$$

where $IM(A)$ is called the interval-valued mean of fuzzy number A. Also, $M_1(A)$ is called the lower possibilistic mean value and $M_2(A)$ is called the upper possibilistic mean value of a fuzzy number A (more see Carlsson and Fuller 2001). Dubois and Prade (1987) have introduced the expected interval of fuzzy number A as

$$EI(A) = [E_1(A), E_2(A)] = \left[\int_0^1 a_1(\alpha)d\alpha, \int_0^1 a_2(\alpha)d\alpha\right],$$

we consider the middle point of $EI(A)$ as $E(A) = \int_0^1 \frac{a_1(\alpha)+a_2(\alpha)}{2}d\alpha$, that is called the mean value of a fuzzy number A.

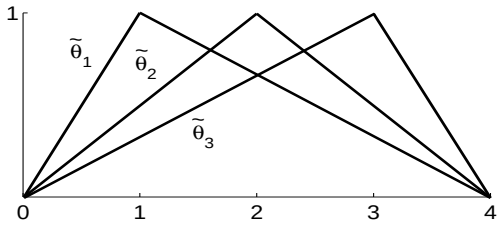


Figure 1: Fuzzy number $A = (a, b, c, d)_n$

Example 1 Let $A = (a, b, c, d)_n$ be a fuzzy number. Then we obtain:

$$EI(A) = [E_1(A), E_2(A)] = \left[a + (b-a)\frac{n}{n+1}, d - (d-c)\frac{n}{n+1}\right].$$

$$IM(A) = [M_1(A), M_2(A)] = \left[a + (b-a)\frac{2n}{2n+1}, d - (d-c)\frac{2n}{2n+1}\right].$$

$$E(A) = \frac{a+d}{2} + \frac{n}{2(n+1)}(b+c-a-d),\ M(A) = \frac{a+d}{2} + \frac{n}{2n+1}(b+c-a-d).$$

3 Median value and interval-valued median

The median value of a fuzzy number was defined by Dubois and Prade (1987) as the numerical value from the support function, it that equally, divides the area under the membership function.

Definition 2 *The median value of a fuzzy number A is the numerical value $Me(A)$ of support function so that*

$$\int_a^{Me(A)} A(x)dx = \frac{1}{2}\int_a^d A(x)dx. \tag{8}$$

Definition 3 *Let A be a fuzzy number with membership function $A(x)$ so that $x_0 = min\{x|A(x) \geq 1\}$ and $y_0 = max\{x|A(x) \geq 1\}$. Then*

interval $IMe = [Me_1(A), Me_2(A)]$, *from support function is define as the interval-valued median of a fuzzy number* A, *where* $Me_1(A)$ *and* $Me_2(A)$ *are obtains by solving*

$$\int_a^{Me_1(A)} A(x)dx = \frac{1}{2}\int_a^{x_0} A(x)dx,$$
$$\text{and} \quad \int_{y_0}^{Me_2(A)} A(x)dx = \frac{1}{2}\int_{y_0}^{d} A(x)dx. \tag{9}$$

For example let $A = (a, b, c, d)_n$. Then

$$IMe = [Me_1(A), Me_2(A)] = \left[a + (b-a)2^{-1/(n+1)}, d - (d-c)2^{-1/(n+1)}\right],$$

one can see that for fuzzy number $A = (a, b, c, d)_n$, $IMe = [A]_\alpha$, so that $\alpha = 2^{-1/(n+1)}$, this means that at least membership grade for the interval-valued median fuzzy number A is $2^{-1/(n+1)}$.

4 Mode values and interval-valued mode

In this section we introduce several mode value of a fuzzy number. These points of the support function of a fuzzy number A are point estimations with more membership grade. Let A be a normal fuzzy number with continues membership function. The mode value of A is define as $Mode(A) = \theta_0 = \{x|A(x) \geq 1\}$. A fuzzy number my have been one or more than mode value. Let $A = (a, b, c, d)$ be a normal trapezoidal fuzzy number with continues membership function $A(x)$, we study four the mode values for A.

(1) The first mode value of A is: $FMo(A) = min\{x|A(x) \geq 1\} = b$.

(2) The last mode value of A is: $LMo(A) = min\{x|A(x) \geq 1\} = c$.

(3) The third mode value of A is middle point of interval $[FMo(A), LMo(A)]$, this means that, $MMo(A) = \frac{FMo(A)+LMo(A)}{2} = \frac{b+c}{2}$. We can rewrite the above items as follows:

Let $supp(A) = \{x \in X|A(x) > 0\}$ and $Core(A) = \{x \in X|A(x) = \sup_{x\in X} A(x)\}$, where X is universe set and restric to a bounded subset of the real line e.g. $[a, b]$. Then for a fuzzy number A we have

Figure 2: Fuzzy number A and and mode different values

$FMo(A) = minCore(A), FMo(A) = maxCore(A)$,
$MMo(A) = \frac{minCore(A)+maxCore(A)}{2}$.

(4) By an analytic geometry, we propose a new method to find the mode value of a trapezoidal fuzzy number. Let $A = (a, b, c, d)$ be a trapezoidal fuzzy number, and let $P_1 = (a, 0), P_2 = (c, 1), P_3 = (b, 1), P_4 = (d, 0)$ be four points on A. Also suppose (x_{Mo}, y_{Mo}) be point of intersection between lines $\overline{P_1P_2}$ and $\overline{P_3P_4}$ (see Fig. 2), then we get
$\overline{P_1P_2} : y = \frac{x-a}{c-a}$ and $\overline{P_1P_2} : y = \frac{d-x}{d-b}$; therefore,

$$(x_{Mo}, y_{Mo}) = \left(\frac{cd - ab}{(c + d) - (a + b)}, 1 \right).$$

Definition 4 *Let A = (a, b, c, d) be a trapezoidal fuzzy number. We define the new value for the mode of A as* $NMo(A) = \frac{cd-ab}{(c+d)-(a+b)}$ *so that* $A(NMo(A))$ *=1.*

Theorem 1 *Let B = (a, b, c, d) n . Then the new value of the mode for fuzzy number B is:*
$NMo(A) = \frac{cd-ab}{(c+d)-(a+b)}$ *and* $B(NMo(B)) = 1$.
This theorem shows that if K be the new value of the mode for a

Figure 3: Symmetric Fuzzy number A

trapezoidal fuzzy number $A = (a, b, c, d)$, then K is the new value of the mode for fuzzy number $A = (a, b, c, d)_n$.

Theorem 2 *Let A be a symmetric fuzzy number with $[A]_\alpha = [k - s(\alpha), k + s(\alpha)]$. Then*

$$E(A) = M(A) = Me(A) = Mo(A) = k. \tag{10}$$

Note that for any $\alpha \in [0, 1], k = \frac{a_1(\alpha)+a_2(\alpha)}{2}$.

Example 2 Let $A = (0, 2, 4)$ be a triangular fuzzy number. Then we obtain
$[A]_\alpha = [2\alpha, 4 - 2\alpha]$, $k = \frac{2\alpha+4-2\alpha}{2} = 2$. Therefore, $E(A) = M(A) = Me(A) = Mo(A) = 2$.

5 Fuzzy estimators based on expert opinion

In this section we want to estimate the value for a certain parameter θ. First assume that we have only expert. Let θ_1 be the smallest possible value for θ, let θ_3 be the largest possible value for, and let θ_2 be the most likely value. We can ask the expert to give values for θ_1,θ_2,θ_3 and we construct the triangular fuzzy estimator $\tilde{\theta_1} = (\theta_1, \theta_1, \theta_1)$ for θ . Now suppose we have N expert. We still want to construct a

triangular fuzzy estimator $\tilde{\theta_1} = (\theta_1, \theta_1, \theta_1)$. The easiest way to do this is to ask the experts for their $\theta_{1i}, \theta_{2i}, \theta_{3i}$, for all $1 \leq i \leq N$, and then take average of each component. For find the confidence interval we use from the intersection of confidence intervals for all fuzzy numbers $\tilde{\theta_i}$. let $\theta_{1i} = \theta_1$ and $\theta_{3i} = \theta_3$, for any $1 \leq i \leq N$, this means that for all $\tilde{\theta_i}$ support functions are equals. Hence, we introduce the following intervals and point estimators for fuzzy parameter $\tilde{\theta}$.

$$EI(\hat{\tilde{\theta}}) = \left[E_1(\hat{\tilde{\theta}}), E_2(\hat{\tilde{\theta}})\right], \tag{11}$$

where

$$E_1(\hat{\tilde{\theta}}) = max\{E_1(\tilde{\theta_i}), 1 \leq i \leq N\}. \tag{12}$$

$$E_2(\hat{\tilde{\theta}}) = min\{E_2(\tilde{\theta_i}), 1 \leq i \leq N | E_2(\tilde{\theta_i}) \geq max\{E_1(\tilde{\theta_i}), 1 \leq i \leq N\}\}, \tag{13}$$

$$E(\hat{\tilde{\theta}}) = \frac{E_1(\hat{\tilde{\theta}}) + E_2(\hat{\tilde{\theta}})}{2} \tag{14}$$

Similarly, we can obtain these estimators for median value of a fuzzy number.

Example 3 Let $\tilde{\theta_1} = (0, 1, 4)$,$\tilde{\theta_2} = (0, 2, 4)$,$\tilde{\theta_3} = (0, 3, 4)$ be three triangular fuzzy numbers (see Figure 4). we get results as the following table:

	EI	IM	IMe	E	M	Me
$\tilde{\theta_1}$	[0.5,2.5]	[0.667,2.000]	[0.707,1.879]	1.5	1.333	2.449
$\tilde{\theta_2}$	[1.0,3.0]	[1.333,2.667]	[1.414,2.586]	2	2	2
$\tilde{\theta_3}$	[1.5,3.5]	[2.000,3.333]	[2.121,3.293]	2.5	2.667	1.551
$\hat{\tilde{\theta}}$	[1.5,2.5]	[2.000,2,000]	[1.414,2.586]	2	2	2

Figure 4: Fuzzy numbers $\tilde{\theta_1}, \tilde{\theta_2}, \tilde{\theta_3}$

When one of methods for the interval estimators is not proper, we use from an other method, e.g. $IM(\hat{\tilde{\theta}}) = [2, 2]$, that can not be as a suitable interval-valued mean. If we apply simple method of take the average of each component, then we have $\hat{\tilde{\theta}} = (0, 2, 4)$.

We can also apply an other method to find the interval estimators of fuzzy numbers as:

$$EI^*(\hat{\tilde{\theta}}) = \left[\frac{\sum_{i=1}^{N} E_1(\tilde{\theta_i}) A_{\tilde{\theta_i}}(E_1(\tilde{\theta_i}))}{\sum_{i=1}^{N} A_{\tilde{\theta_i}}(E_1(\tilde{\theta_i}))}, \frac{\sum_{i=1}^{N} E_2(\tilde{\theta_i}) A_{\tilde{\theta_i}}(E_1(\tilde{\theta_i}))}{\sum_{i=1}^{N} A_{\tilde{\theta_i}}(E_2(\tilde{\theta_i}))} \right]. \quad (15)$$

$$IM^*(\hat{\tilde{\theta}}) = \left[\frac{\sum_{i=1}^{N} M_1(\tilde{\theta_i}) A_{\tilde{\theta_i}}(M_1(\tilde{\theta_i}))}{\sum_{i=1}^{N} A_{\tilde{\theta_i}}(M_1(\tilde{\theta_i}))}, \frac{\sum_{i=1}^{N} M_2(\tilde{\theta_i}) A_{\tilde{\theta_i}}(M_1(\tilde{\theta_i}))}{\sum_{i=1}^{N} A_{\tilde{\theta_i}}(M_2(\tilde{\theta_i}))} \right]. \quad (16)$$

The above intervals are the weighted interval-valued estimators of intervals for a fuzzy number.

5 Conclusion

This paper reviewed some intervals and approximately points of a fuzzy number. We have developed this intervals and points for a sample of fuzzy data and the new methods to obtain the mode value,

expected interval, interval-valued mean and median of fuzzy numbers are introduced. These indexes can apply to estimate fuzzy parameters unknown of a population in a fuzzy environment.

References

[1] Bodjanova S. (2005). Median value and median interval of a fuzzy number. *Information Sciences*, **172**, 73-89.

[2] Carlsson C., Fuller R. (2001). On possibilistic mean value and variance of fuzzy numbers. *Fuzzy Sets and Systems*, **122**, 315-326.

[3] Chanas S. (2001). On the interval approximation of a fuzzy number. *Fuzzy Sets and Systems*, **122**, 353-356.

[4] Dubois D., Prade H. (1987). The mean value of a fuzzy number. *Fuzzy Sets and Systems*, **24**, 279-300.

[5] Dubois D., Prade H. (2000). Fundamental of fuzzy sets. Handbook of Fuzzy Sets . vol.1. *Kluwer, Dordrecht*.

[6] Heilpern S. (1992). The expected value of a fuzzy number. *Fuzzy Sets and Systems*, **47**, 81-86.

[7] Grzegorzewski P. (2002). Nearest interval approximation of a fuzzy number. *Fuzzy Sets and Systems*, **130**, 321-330.

[8] Saiedifar A, Pasha E. (2007). The percentiles of trapezoidal fuzzy numbers and their applications. *ICREM3. Kuala Lumpur, Malaysia. Proceedings (Pure Mathematics Statistics)*, 61-66.

[9] Fuller R., Majlender P. (2003). On weighted possibilisitic mean and variance of fuzzy numbers. *Fuzzy sets and Systems*, **136**, 363-374.

[10] Zadeh L.A , (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, **1**, 3-28.

A Characterization and Generalization for Two Sided Power Distributions and Adjusted Method of Moments

A. R. Soltani* **and H. Homei**

Kuwait University and Shiraz University

Abstract: A characterization is provided for the two sided power distributions, introduced by van Dorp and Kotz (2002). The characterization naturally leads us to a new class of generalized two sided power distributions which appears to be rich. In this class, the symmetric distributions are supported by finite intervals and have normal density shape densities. Our study on two sided power distributions also leads us to a new class of discrete distributions on $\{0, 1, \cdots, k\}$. We also give a new numerical method for parameter estimation using moments.

Keywords: Two sided power distributions, Generalized two sided power distributions, Discrete distribution induced from two sided power distributions, Random weighted average, adjusted method of moments.

1 Introduction

The work of Nadarajah (1999) initiated research work on distribution functions supported by finite intervals, say $(0, 1)$, that assume different formulations on subintervals of their supports, say $(0, \theta), (\theta, 1)$. Two sided power distributions (TSP) are of this type, and were introduced by van Dorp and Kotz (2002a) as underlying statistical distributions for certain monthly interest rates. A TSP distribution is specified by four parameters a, m, b, n, where $-\infty < a < b < +\infty$. The interval $[a, b]$ supports the density functions; m, $a < m < b$, is a, say , side parameter, one side of the distribution is supported by $[a, m]$, the other side by [m,b]. The parameter $n > 0$ gives the smoothness of the spectral density. The density is continuous. It is increasing on $[a, m]$, decreasing on $[m, b]$ and differentiable at every point in (a, b) except at m. The potential, flexibility and applicability of TSP distributions in applied fields have been examined and explored in a series of papers

Invited speaker

by van Dorp and Kotz, (2002a, 2002b, 2003b), Nadarajah, (2005) and Perez et al., (2005), among others.

In this articles we use a bivariate Dirichlet distribution and provide a characterization for TSP random variables, for $n > 1$. Based on our characterization, for integer n, a TSP random variable can be deduced from n i.i.d uniform $(0, 1)$ random variables. We also derive equivalent forms for the k-th moment of the TSP distributions about zero. An interesting and new family of discrete distributions on $\{0, 1, 2, 3, ...k\}$, k an integer, is captured from these equivalent forms. Our characterization also naturally leads to a generalization of TSP distributions (GTSP). A generalizations for the TSP distribution is produced in the literature, van Dorp and Kotz (2003a), see also Ozlem and Bairamov (2005). The GTSP distributions, introduced here, are very different from those in the cited works and are introduced by scrutinizing the basic structure of TSP random variables. Our GTSP distributions exhibit interesting features, and overcome some deficiencies of the TSP distributions, see Remark 3.1 for more details.

We were led by a referee to represent the density functions of our GTSP in terms of certain Gauss Hypergeometric Functions.

Our symmetric GTSP random variables form a rich class of symmetric distributions on finite intervals with normal density shape densities.

A new parameter estimation procedure, called "adjusted method of moments" (AMM), is established by using auxiliary parametric statistics, which are defined here (called upper and lower random Stieltjes sums), and moments of the distribution function.

This article is organized as follows. In Section 2 we give our characterization for TSP distributions, Theorem 2.1. Three different formulas for the kth moments, about zero, of TSP distributions are given, Theorem 2.2. A new class, to the best of our knowledge, of discrete distributions is given, Theorem 2.3. Section 3 is for GTSP distributions. Various features of this new class of distributions are brought into sight. Special attention is given to the symmetric case. In Section 4 we present our estimation procedure. The random Stieltjes sums are

introduced, and are compared with the sample mean. In an example, the priority of AMM estimators to other well known estimators is brought into sight through simulation.

2 A characterization for TSP Random Variables

We recall from van Dorp and Kotz (2002a) that a TSP random variable $Z \sim \text{TSP}(a, m, b, n)$ is defined on an interval (a, b) in the real line with the probability density function (p.d.f.) $f(z \mid a, m, b, n)$ given by:

$$f(z|a,m,b,n) = \begin{cases} \frac{n}{b-a}\left(\frac{z-a}{m-a}\right)^{n-1}, & a < z < m, \\ \\ \frac{n}{b-a}\left(\frac{b-z}{b-m}\right)^{n-1}, & m < z < b. \end{cases}$$

A TSP random variable with $a = 0$, $b = 1$ is called standard, and is denoted by $X \sim \text{STSP}(\theta, n)$, the parameter m in this case is denoted by θ.

Our characterization for TSP random variables is included in the following theorem. Since $Z = (b-a)X + a$, we confine ourselves to the class of STSP random variables.

Theorem 2.1. Assume that (W, V) is a bivariate standard Dirichlet random vector with parameters $\alpha = 1,\ \beta = n-1,\ \gamma = 1;\ n > 1$, i.e. (W, V) possesses the joint density function

$$f_{W,V}(w, v) = n(n-1)v^{n-2}\ I_{[0,1]}(w)I_{[0,1]}(v)\ I_{[0,1]}(w+v).$$

Then

(a) W and $S = V + W$ possess the marginal p.d.f.

$$f_W(w) = n(1-w)^{n-1}, \quad 0 < w < 1, \quad \text{and} \quad f_S(s) = ns^{n-1}, \quad 0 < s < 1,$$

respectively;

(b)

$$E(W^j(V+W)^k) = \left[\frac{j!}{n+k+j}\right]\frac{\Gamma(n+1)}{\Gamma(n+j)},$$

for nonnegative integers j, k;

(c) for given W; $T = V/(1-W)$ possesses the same p.d.f. of the form of W, replacing n by $n-1$;

(d)

$$E(W^jV^k) = \frac{n(n-1)j!}{(k+n-1)}\,\frac{\Gamma(n+k)}{\Gamma(n+j+k+1)},$$

(e) for every θ, $0 < \theta < 1$, $X = W + \theta V$ is a STSP random variable with parameters θ, n.

Proof. (a) and (b) are immediate. (c): The joint p.d.f. of $T = V/(1-W)$ and W is given by

$$f_{T,W}(t,w) = n(n-1)(t-tw)^{n-2}(1-w) = n(n-1)t^{n-2}(1-w)^{n-1},$$

$0 < t < 1$, $0 < w < 1$, giving that

$$f(t|w) = \frac{n(n-1)t^{n-2}(1-z)^{n-1}}{n(1-z)^{n-1}} = (n-1)t^{n-2},$$

$0 < t < 1$, $0 < w < 1$.

(d): This part is known. But we give a new short proof. By using (c) and (a),

$$\begin{aligned} EW^jV^k &= E(E(W^jV^k|W)) = E(W^j(1-W)^kE[\frac{V^k}{(1-W)^k}|W]) \\ &= E(W^j(1-W)^kE(T^k|W)) = E[W^j(1-W)^k]\frac{n-1}{n+k-1} \\ &= \frac{n(n-1)j!\,\Gamma(n+k)}{(k+n-1)\Gamma(n+j+k+1)}, \end{aligned}$$

(e): The joint density function of $X = W + \theta V$ and W is given by

$$f_{X,W}(x,w) = \frac{n(n-1)}{\theta}(\frac{x-w}{\theta})^{n-2}I_{[0,1]}(w)I_{[w,w+\theta(1-w)]}(x).$$

Thus

$$f_X(x) = \begin{cases} \int_0^x \frac{n(n-1)}{\theta}(\frac{x-w}{\theta})^{n-2}\,dw, & 0 < x < \theta \\ \int_{\frac{x-\theta}{1-\theta}}^x \frac{n(n-1)}{\theta}(\frac{x-w}{\theta})^{n-2}\,dw, & \theta < x < 1, \end{cases}$$

giving

$$f_X(x) = \begin{cases} n(\frac{x}{\theta})^{n-1}, & 0 < x < \theta \\ n(\frac{1-x}{1-\theta})^{n-1}, & \theta < x < 1, \end{cases}$$

which is the p.d.f. of a STSP random variable with parameters θ, n. The proof of the theorem is complete.

Remark 2.1. The characterization given in Theorem 2.1 (e), for integer n, can be read as follows. For an integer $n \geq 1$, $X \sim STSP(\theta, n)$ if and only if

$$X \stackrel{d}{=} U_{(1)} + \theta(U_{(n)} - U_{(1)}) \ , \ 0 < \theta < 1,$$

where $U_{(1)} = \min(U_1, ..., U_n)$, $U_{(n)} = \max(U_1, ..., U_n)$; and $U_1, ..., U_n$ are i.i.d uniform $(0, 1)$.

Remark 2.2. The cases that $\theta = 0$ or $= 1$ are included in Theorem 2.1(a). Therefore the conclusion of Theorem 2.1(e) is fulfilled for every $\theta \in [0, 1]$.

The following theorem gives equivalent forms for the $\mu_k' = E(X^k)$.

Theorem 2.2. Let $Z \sim \text{STSP}(\theta, n),\ n > 1,\ \theta \in [0, 1]$. Then for any integer $k \geq 0$,

(a)

$$EX^k = \frac{n\theta^{k+1}}{n+k} + \sum_{i=0}^{k}(-1)^k \binom{k}{k-i} \frac{n}{n+i}\,(1-\theta)^{i+1},$$

(b)

$$EX^k = \frac{\Gamma(n+1)}{n+k+1} \sum_{i=0}^{k} \left(\begin{array}{c} k \\ i \end{array} \right) \frac{\Gamma(i+1)}{\Gamma(n+i)} \theta^{k-i}(1-\theta)^i,$$

(c)

$$EX^k = \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) \frac{n(n-1)}{n+j-1} \beta(n+j, k-j+1)\theta^j.$$

Proof. (a): This formula was derived by van Dorp and Kotz (2002).

(b): By using Theorem 2.1(e), we obtain that

$$\begin{aligned} E(X)^k &= E(W+\theta V)^k \\ &= \sum_{i=0}^{k} \left(\begin{array}{c} k \\ i \end{array} \right) (1-\theta)^i \theta^{k-i} E[W^i (V+W)^{k-i}]. \end{aligned}$$

The result will follow from Theorem 2.1(b).

(c): We let $T = V/(1-W)$, and use parts (e), (c) and (d) of Theorem 2.1 to obtain that

$$\begin{aligned} E(E(W+\theta V)^k|W) &= E(E((1-W)^k(\frac{W}{1-W} + \theta \frac{V}{1-W})^k|W) \\ &= E(E(\sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) (1-W)^j W^{k-j} T^j|W)\theta^j \\ &= \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) E(1-W)^j W^{k-j} E(T^j|W)\theta^j \\ &= \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) \frac{n(n-1)}{n+j-1} \frac{\Gamma(j+n)\Gamma(k-j+1)}{\Gamma(n+k+1)} \theta^j. \end{aligned}$$

The proof of the theorem is complete.

As an application of Theorem 2.2 we give the following theorem which gives a new family of discrete distributions supported by $\{0, 1, 2, ..., k\}$; k is a positive integer.

Theorem 2.3. For every real number $n > 1$,

$$g(j : k, n) = \binom{k}{j} \frac{(n-1)(n+k)}{n-1+j} \beta(n+j, k-j+1), \quad j = 0, ..., k,$$

are probability masses of a discrete random variable on $\{0, ..., k\}$ parameterized by n.

Proof. By equating the formulas (a) and (c) for $\theta = 1$ in Theorem 2.2, we get

$$\sum_{j=0}^{k} \binom{k}{j} \frac{(n-1)(n+k)}{n-1+j} \beta(n+j, k-j+1) = 1.$$

The proof of the theorem is complete.

Remark 2.3. For integer $n \geq 2$,

$$g(j; k, n) = \frac{\binom{n+j-2}{j}}{\binom{n+k-1}{k}}, \ j = 0, \cdots, k.$$

It would be interesting to directly verify that the masses add to 1, in this case. Indeed

$$\begin{aligned} \sum_{j=0}^{k} \binom{n+j-2}{j} &= \sum_{j=1}^{k} [\binom{n+j-1}{j} - \binom{n+j-2}{j}] \\ &= \binom{n+k-1}{k} - \binom{n-1}{0} + 1 \\ &= \binom{n+k-1}{k}, \end{aligned}$$

giving that $\sum_{j=0}^{k} g(j;k,n) = 1$. Only for n integer we could find the trace of this distribution in Wilks (1962) in the case that $k = N - m,\ N \geq m,\ n = m+1$, where $g(j;k,n)$ gives probability masses of $\max\{J_j\}$; $J_1, \cdots, J_m$ are m without replacement draws from the set $\{1, \cdots, N\}$.
The probability masses in Theorem 2.3 are depicted in Fig. 1.

3 Generalized TSP random variables

The characterization given in Theorem 2.1(e) leads us to a natural generalization of TSP random variables. The details are included in the following theorem and its immediate proceeding definition and remark.

Theorem 3.1. Assume that (W,V) is a bivariate standard Dirichlet random vector with parameters $\alpha = k_1,\ \beta = k_2 - k_1,\ \gamma = n - k_2 + 1;\ 0 < k_1 < k_2 < n+1$, i.e. (W,V) possesses the joint density function

$$f(w,v) = \frac{\Gamma(n+1)}{\Gamma(k_1)\,\Gamma(k_2-k_1)\,\Gamma(n-k_2+1)} w^{k_1-1} v^{k_2-k_1-1} (1-w-v)^{n-k_2}, \tag{3.1}$$

$0 < w < 1,\ 0 < v < 1,\ \ 0 < w+v < 1$. Then

(a) W and $S = V + W$ are $\beta(k_1, n-k_1+1)$ and $\beta(k_2, n-k_2+1)$ respectively;

(b)

$$E(W^j(V+W)^k) = \frac{\Gamma(n+1)\Gamma(k+j+k_2)\Gamma(j+k_1)}{\Gamma(k_1)\Gamma(k+j+n+1)\Gamma(j+k_2)};$$

(c) for given $W,\ T = V/(1-W)$ possesses the same p.d.f. as W;

(d)

$$E(W^j V^k) = \frac{\Gamma(n+1)\Gamma(k_1+j)\Gamma(k_2-k_1+j)}{\Gamma(k_1)\Gamma(k_2-k_1)\Gamma(n+j+k+1)}.$$

Proof. The proof of this theorem is very similar to the proof of Theorem 2.1, and therefore is omitted here.

Definition 3.1. A random variable $X = W + \theta V$, where (W, V) possesses the joint p.d.f. given in (3.1) is called standard generalized two sided power random variable with parameters θ, n, k_1, k_2; $X \sim SGTSP(\theta, n, k_1, k_2)$. Note that $0 < \theta < 1$, n, k_1, k_2 are real numbers subject to $0 < k_1 < k_2 < n + 1$.

The p.d.f. of X can be expressed in terms of certain Gauss Hypergeometric Functions. Indeed the joint p.d.f. of (W, X) is given by

$$f_{W,X}(w, x) = \kappa\, w^{k_1-1}\, (x - w)^{k_2-k_1-1}\, (\theta - x + w(1 - \theta))^{n-k_2}, \quad (3.2)$$

where $0 < w < 1$ and $w \le x \le w + \theta(1 - w)$, and

$$\kappa = \frac{\Gamma(n + 1)}{\Gamma(k_1)\, \Gamma(k_2 - k_1)\, \Gamma(n - k_2 + 1)\, \theta^{n-k_1}}. \quad (3.3)$$

Therefore

$$f(x) = \begin{cases} \int_0^x f(w, x)\, dw, & 0 < x < \theta, \\ \int_{\frac{x-\theta}{1-\theta}}^x f(w, x)\, dw, & \theta < x < 1. \end{cases} \quad (3.4)$$

Although the integrations in (3.4) can not be taken, but the density function $f(x)$ can be expressed in term of the Gauss Hypergeometric Function $F(a, b, c; z)$, which is a well known special function. Indeed according to the Euler's Formula, the Gauss Hypergeometric Function assumes the integral representation

$$F(a, b, c; z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c - b)} \int_0^1 t^{b-1}(1 - t)^{c-b-1}(1 - tz)^{-a} dt,$$

where a, b, c are parameters subject to $-\infty < a < +\infty$, $c > b > 0$, whenever they are real, and z is the variable. For more on the Gauss

Hypergeometric Function see Zayed (1996). By using the Euler's Formula, the density function in (3.4) can be expressed as follows.
For $0 < x < \theta$, $f(x)$ is given by

$$\kappa_2\, \theta^{k_1-1}(\frac{x}{\theta})^{k_2-1}(1-\frac{x}{\theta})^{n-k_2}F(k_2-n,k_1,k_2;-\frac{x(1-\theta)}{\theta-x}); \qquad (3.5a)$$

for $\theta < x < 1$, $f(x)$ is given by

$$\kappa_1\,(1-\theta)^{n-k_2}(\frac{1-x}{1-\theta})^{n-k_1}(1-\frac{1-x}{1-\theta})^{k_1-1}$$
$$F(1-k_1,n-k_2+1,n-k_1+1;-\frac{\theta(1-x)}{x-\theta}), \qquad (3.5b)$$

where $\kappa_i = \frac{\Gamma(n+1)}{\Gamma(k_i)\Gamma(n-k_i+1)},\ i = 1, 2.$
A close examination of the integrand in (3.4) reveals that the density function exists at every $x \in (0,\theta) \cup (\theta, 1)$. The density f at θ exists if $k_2 < n + k_1$ only. In this case the integration in (3.4) for $x = \theta$ can be easily performed, giving

$$f(\theta) = \frac{n\Gamma(n+k_1-k_2)}{\Gamma(k_1)\Gamma(n-k_2+1)}\theta^{k_1-1}(1-\theta)^{n-k_2}, \quad k_2 < n+k_1. \qquad (3.6)$$

If $k_1 = 1, k_2 = n$, then $F(0, 1, n, z) = 1$, and the density function $f(x)$ given by (3.5a) and (3.5b) readily reduces to the STSP density function given in Section 2.

Definition 3.2. A random variable Z with values in a given interval $[a, b]$ is said to be GTSP if $Z = (b-a)X + a$, where $X \sim$ SGTSP(θ, n, k_1, k_2).

Remark 3.1. For the case that k_1, k_2, and n are integers, a GTSP random variable can be expressed as

$$Z = bU_{(k_1)} + m(U_{(k_2)} - U_{(k_1)}) + a(1 - U_{(k_2)}), \qquad (3.7)$$

where $m = (b-a)\theta + a$, and $U_{(1)}, \cdots, U_{(n)}$ is the order statistic of n i.i.d uniform $(0, 1)$ random variables. Note (3.7) for the case that

$k_1 = 1$, $k_2 = n$ gives a TSP random variable. random weighted averages of the form $\sum x_j U_{(k_j)}$, $\{k_1, \cdots, k_n\} \subset \{1, \cdots, n\}$, $x_j > 0, \forall j$, are considered in the literature. It is well known that these random variables do not assume densities in a pleasant form. Even the form derived for their cumulative distributions has complexities. Articles are produced for numerical evaluation of cumulative probabilities, see [15, 1].

Remark 3.2 As we noticed in Remark 3.1, a TSP random variable Z, for integer n, is the random weighted average of b, m, and a, $[b > m > a]$, with random masses $1 - U_{(1)}$, $U_{(n)} - U_{(1)}$ and $1 - U_{(n)}$ respectively. The intermediate point m receives more mass than the ending points b, a, which statistically receive equal masses. For large n, the intermediate point receives substantially larger weight than the ending points; and the density is too tall at θ, $f(\theta) = n$. The GTSP distribution overcome this illusion. By putting more random weights on the ending points, $f(\theta)$ will be of reasonable size, even for large n. Indeed it is evident from (3.2) that $f(\theta) \to 0$, as $n \to \infty$ for fixed k_1, k_2.

The probability density functions of some GTSP distributions are plotted and illustrated in Fig. 2.

The moments of GTSP distributions can be derived by using the following theorem.

Theorem 3.2. Let $X \sim SGTSP(\theta, n, k_1, k_2)$, and let $n_1 = k_1$, $n_2 = k_2 - k_1$, $n_3 = n - k_2 + 1$. Then for any integer $k \geq 0$ and any real number $n \geq 0$, $n \neq 1$,

(a):

$$EX^k = \frac{\Gamma(k + k_2)\,\Gamma(n+1)}{\Gamma(k_1)\,\Gamma(n+k+1)} \sum_{i=0}^{k} \binom{k}{i} \frac{\Gamma(k_1 + i)}{\Gamma(k_2 + i)} \theta^{k-i}(1-\theta)^i$$

(b):

$$EX^k = \sum_{j=0}^{k} \left(\begin{smallmatrix} k \\ j \end{smallmatrix} \right) \frac{\beta(j+n_2, n-n_2+1)}{\beta(n_2, n_3+n_1)} \frac{\beta(k-j+n_1, n-n_1+k-j+1)}{\beta(n_1, n-n_1+1)} \theta^j.$$

Proof. (a): By using part (b) of Theorem 3.1, we obtain that

$$\begin{aligned}
E(X)^k &= E(W+\theta V)^k \\
&= E\sum_{i=0}^{k} \left(\begin{smallmatrix} k \\ i \end{smallmatrix} \right) ((1-\theta)W)^i ((V+W)\theta)^{k-i} \\
&= \sum_{i=0}^{k} \left(\begin{smallmatrix} k \\ i \end{smallmatrix} \right) (\frac{1-\theta}{\theta})^i \, \theta^k E\, W^i S^{k-i} \\
&= \theta^k \sum_{i=0}^{k} \left(\begin{smallmatrix} k \\ i \end{smallmatrix} \right) (\frac{1-\theta}{\theta})^i \frac{\Gamma(n+1)\Gamma(k_1+i)\Gamma(k+k_2)}{\Gamma(k_2+i)\Gamma(k_1)\Gamma(n+k+1)} \\
&= \frac{\theta^k \, \Gamma(k+k_2) \, \Gamma(n+1)}{\Gamma(k_1) \, \Gamma(n+k+1)} \sum_{i=0}^{k} \left(\begin{smallmatrix} k \\ i \end{smallmatrix} \right) (\frac{1-\theta}{\theta})^i \, \frac{\Gamma(k_1+i)}{\Gamma(k_2+i)},
\end{aligned}$$

(b): By using parts (c) and (d) of Theorem 3.1, we obtain that

$$\begin{aligned}
& E(E(W+\theta V)^k|W) = E(E((1-W)^k(\frac{W}{1-W} + \theta\frac{V}{1-W})^k|W) \\
= \; & E((1-W)^k(\sum_{j=0}^{k} \left(\begin{smallmatrix} k \\ j \end{smallmatrix} \right) E[(\frac{W}{1-W})^{k-j}(\frac{V}{1-W})^j|W]\theta^j \\
= \; & E(E(\sum_{j=0}^{k} \left(\begin{smallmatrix} k \\ j \end{smallmatrix} \right) (1-W)^j W^{k-j} T^j|W)\theta^j \\
= \; & \sum_{j=0}^{k} \left(\begin{smallmatrix} k \\ j \end{smallmatrix} \right) E(1-W)^j W^{k-j} E(T^j|W)\theta^j.
\end{aligned}$$

But

$$E(T^j|W) = \frac{\beta(j+n_2, n_1+n_3)}{\beta(n_2, n_1+n_3)},$$

and

$$E(1-W)^j W^{k-j} = \frac{\beta(k-j+n_1, n-n_1+k-j+1)}{\beta(n_1, n-n_1+1)}.$$

Therefore we arrive at formula (b). The proof of the theorem is complete.

Theorem 3.3.. Let $n_1 > 0,\ n_2 > 0,\ n_3 > 0$. Then

$$a_j = \binom{k}{j} \frac{\beta(j+n_2, n_1+n_3)}{\beta(n_2, n_3+n_1)} \frac{\beta(k-j+n_1, n_2+n_3+k-j)}{\beta(n_1, n_2+n_3)}$$

$$\times \frac{\Gamma(n_2+n_1)\Gamma(n_1+n_2+n_3+k)}{\Gamma(k+n_2+n_1)\Gamma(n_1+n_2+n_3)},\quad j = 0, ..., k,$$

are probability masses of a discrete random variable on $\{0, ..., k\}$ parameterized by n_1, n_2 and n_3.

Proof. The result will follow by equating the formulas (a) and (b) for $E(X^k)$ given in Theorem 3.2 evaluated at $\theta = 1$. The proof is complete.

Let us record that if $X \sim SGTSP(\theta, n, k_1, k_2)$, then

$$E(X) = \frac{k_1 + \theta(k_2 - k_1)}{n+1}$$

and

$$V(X) = \frac{k_1(n-k_1+1)(1-\theta)^2 + k_2(n-k_2+1)\theta^2 + 2k_1(n-k_2+1)\theta(1-\theta)}{(n+1)^2(n+2)}. \tag{3.8}$$

In the special case that $k_1 = 1,\ k_2 = n$, it gives the variance for STPD derived by van Dorp and Kotz (2002a).

Let us pay a special attention to the symmetric SGTSP distributions and random variables. A SGTSP random variable is said to be symmetric if $X \stackrel{d}{=} 1 - X$. This immediately implies that $E(X) = 1/2$. This is necessary and sufficient condition for a TSP random variable to be symmetric, as can be easily seen by examining the p.d.f. But this is not the case for SGTSP random variables. Details are given in the following theorem.

Theorem 3.4. Let $X \sim SGTSP(\theta, n, k1, k2),\ 0 < k_1 < k_2,\ k_2 < \min\{n+1,\ n+k_1\}$, then X is symmetric if and only if $\theta = \frac{1}{2}$ and $k_2 = n - k_1 + 1$.

Proof. The "if" part is straight forward. For the "only if" part, we notice that for $x = \theta$, from (3.6), $f(\theta) = f(1-\theta)$ implies that $k_2 = n - k_1 + 1$. From the symmetry $E(X) = 1/2$, which gives $k_1(1-\theta) + k_2\theta = (n+1)/2$. But for a given $k_1 < (n+1)/2$, the line $k_1(1-\theta) + (n - k_1 + 1)\theta$ joins the points $(0, n-k_1+1)$ and $(1, k_1)$ as θ varies from 0 to 1. Hence it intersects the line $(n+1)/2$ at exactly one point attained at $\theta = 1/2$. The proof of the theorem is complete.

If X is a symmetric SGTSP random variable with parameters (n, k), then it follows from Theorem 3.4 that $\frac{1}{2} < k < \frac{n+1}{2}$. The density function of X can be deduced from (3.5a) and (3.5b), namely,

$$f(x) = \begin{cases} 2^{n-2k+1}\kappa_3\, x^{n-k}(1-2x)^{k-1}G(-\frac{x}{1-2x}), & 0 < x < \frac{1}{2} \\ 2^{n-2k+1}\kappa_3\,(1-x)^{n-k}(2x-1)^{k-1}G(-\frac{1-x}{2x-1}), & \frac{1}{2} < x < 1, \end{cases}$$

where $G(z) = F(1-k, k, n-k+1; z),\ \kappa_3 = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)}$. Also

$$f(1/2) = n\frac{\Gamma(2k-1)}{[\Gamma(k)]^2}(1/4)^{k-1}.$$

Clearly $f(x) = f(1-x),\ 0 < x < 1$, as expected. Let us also record from (3.8) that

$$E(X) = \frac{1}{2} \quad V(X) = (1/2)\left[\frac{k}{(n+1)(n+2)}\right]. \tag{3.9}$$

Plots of the density of X for some different values of the parameters are depicted in Fig. 2[bottom]. As it is evident from the density plots, a symmetric GTSP random variable Z has normal density shape densities. On a given interval $[a, b]$, the parameters n, k give the desired variance, $V(Z) = (b-a)^2 V(X)$, which can be any value in the interval $(0, (b-a)^2)/[4(n+2)]),\ n > 0$.

4 Adjusted Method of Moments

In this section we introduce a new parameter estimation procedure, that is called here "adjusted method of moments, (AMM). Let $X_1, \cdots, X_m$ be a random sample. The distribution of X_i is denoted by F, and is assumed to be arbitrary, determined by certain parameters, that are assumed to be unknown. In order to establish AMM, we first introduce following random expressions.

Since $\mu'_k = E(X)^k = \int_\Omega X^k dP$, the classical integration theory reveals that μ'_k can be approximated by

$$\widetilde{X^k} = \sum_{i=1}^{m} X_{(i)}^k [F(X_{(i+1)}) - F(X_{(i)})], \quad k = 1, 2, ..., \tag{4.1}$$

or by

$$\widetilde{\widetilde{X^k}} = \sum_{i=1}^{m} X_{(i)}^k [F(X_{(i)}) - F(X_{(i-1)})], \quad k = 1, 2, ..., \tag{4.2}$$

where $F(X_{(0)}) = 0,\ F(X_{(m+1)}) = 1$, and $X_{(1)}, ..., X_{(m)}$ is the order statistic of $X_1, \cdots, X_m$. We refer to the terms given in (4.1) and (4.2) as lower random Stieltjes sum (LRSS) and upper random Stieltjes sum (URSS). We note that if in (4.1) and (4.2) F is replaced by its empirical distribution function, then $\hat{F}(X_{(i+1)}) - \hat{F}(X_{(i)}) = 1/n$, and $\widetilde{X^k}$ and $\widetilde{\widetilde{X^k}}$ will become $(1/m)\sum_{i=1}^{m-1} X_{(i)}^k$, $\overline{X^k}$, respectively. In contrast to $\overline{X^k}$, the values of $\widetilde{X^k}$ and $\widetilde{\widetilde{X^k}}$ can not be fully determined by $X_1, \cdots, X_m$, due to the presence of F, specially when F has unknown parameters. But when F is given, we expect (4.1) and (4.2) estimate

$E(X^k)$ more accurately than $\overline{X}$. As an illustration, for the standard normal distribution, through the simulation, the absolute error of $\widetilde{X^1}$, $\widetilde{\widetilde{X^1}}$ and $\overline{X}$ from the actual mean zero are plotted against m, Fig. 2. It is evident from Fig. 2 that the terms in (4.1) and (4.2), for $k = 1$, give more accurate values than $\overline{X}$ for the actual mean.

AMM is the same as method of moments (MM), the k-th sample moments are replaced by the corresponding LRSS or URSS. The idea is to use $\widetilde{X^k}$ or $\widetilde{\widetilde{X^k}}$ instead of $\overline{X^k}$ in the method of moments (MM). Thus if the distribution F has say L unknown parameters, then the parameters can be estimated, in the AMM, by solving the following system of nonlinear equations.

$$\mu'_k = \widetilde{\widetilde{X^k}}, \quad k = 1, \cdots, L,$$

or

$$\mu'_k = \widetilde{X^k}, \quad k = 1, \cdots, L.$$

The resulting equations in AMM are more complicated than those in MM. This is not a serious disadvantage if numerical derivation in the parameter estimation is concerned. The Implicit Function Theorem and the Inverse Function Theorem together with advanced mathematical softwares can be applied to solve AMM equations. Since, compare to the sample moments, LRSS and URSS give more accurate values for the population moments, AMM is expected to give more accurate estimates.

It is hard in general to derive analytical expressions for AMM estimators; and thus establishing statistical properties of AMM estimators is a challenging task. As an illustration for the performance of AMM estimators, let F be the uniform distribution on $(0, a)$, then AMM with using URSS gives the following estimator for the parameter a.

$$\hat{a} = \sqrt{2}\left[\sum_{i=1}^{m} X_{(i)}(X_{(i)} - X_{(i-1)})\right]^{1/2}.$$

By using simulated samples [$a = 4$, $m = 50$], the absolute errors of $\hat{a}$, $\overline{X}$ and $\max\{X_i\}$ from the parameter a are plotted against m, the sample size, and depicted in Fig. 5. Indeed for a given sample size, the value which is given for each estimator by the graph is the average of the values for the deviation of that estimator from a for 50 replications of the sample. Interestingly, according to Fig 3, $\hat{a}$ shows better performance, even better than the MLE.

References

[1] Diniz, M.C., de Souza e Silva, E., Gail, H.R. (2002). Calculating the distribution of a linear combination of uniform order statistics. *Informs J. Computing*,**14**, 0124-0131.

[2] Johnson, N. L., Kotz, S., Balakrishnan, N. (1994), *Continuous Univariate Distributions1* (2nd ed.), Wiley, New York.

[3] Johnson, N. L., Kotz, S., Balakrishnan, N. (1995), *Continuous Univariate Distributions2* (2nd ed.), Wiley, New York.

[4] Johnson N.L., Kotz, S. (1990). Randomly weighted averages: Some aspects and extensions. *Amer. Statist.* **44**, 245-249.

[5] Kotz, S., Kozubowski, T. J., Podgorski, K. (2001), *The Laplace Distribution and Generalizations, A Revisit with Applications to Communications, Economics, Engineering and Finance*. Boston: Birkhauser.

[6] Nadarajah, S.(2005). On the two-sided power distribution. *Metrika*, **61**, 309-321.

[7] Nadarajah, S. (1999). A polynomial model for bivariate extreme value distributions. *Statist. Probab. Lett.*, **42**, 15-25.

[8] Ozlem, E. O., Bairamov, I. (2005). On the general class of two sided power distributions. *Comm. Statist. Theory Methods*, **34**, 1009-1017.

[9] Perez, J.G., Rambaud, S.C., Garcia, L.G. (2005). The two-sided power distribution for the treatment of the uncertainty in PERT. *Stat. Methods Appl.*, **14**, 209-222.

[10] van Dorp J.R., Kotz, S. (2003a). Generalizations of two sided power distributions and their convolution. *Comm. Stat. Theory Methods*, **32**, 17031723.

[11] Van Dorp, J.R., Kotz, S. (2003b). Generalized trapezoidal distributions. *Metrika*, **58** (1), 85-97.

[12] van Dorp J.R., Kotz, S. (2002a). The standard two-sided power distribution and its properties: with applications in financial engineering. *The Amer. Statist.*, **56**, 9099.

[13] van Dorp J.R., Kotz, S. (2002b). A novel extension of the triangular distribution and its parameter estimation. *The Statistician* **51**, 63-79.

[14] van Dorp J.R., Kotz, S. (1990). A novel extension of the triangular distribution and its parameter estimation. *The Statistician.*, **51**, 6379.

[15] Weisberg, H. (1971). The distribution of linear combinations of order statistics from the uniform distribution. *Ann. Math. Statist.*, **42**, 704-709.

[16] Wilks, S. S. (1962). *Mathematical Statistics*, Wiley, New York.

[17] Zayed, A. I. (1996). *Function and Generalized Function Transformations*. CRC Press, New York.

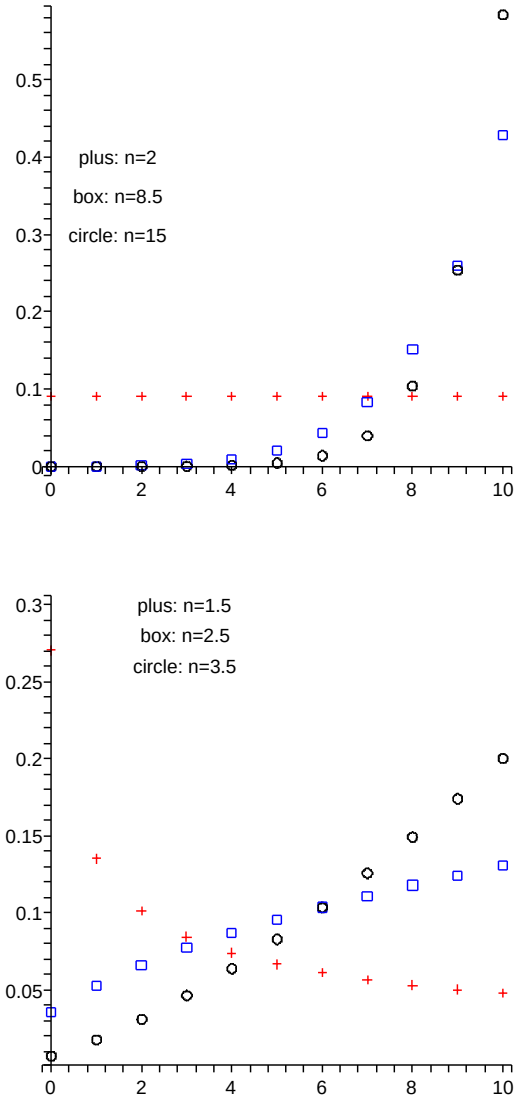


Figure 1: Discrete probability masses $g(j;k,n),\ j=0,\cdots,k;\ k=10.$

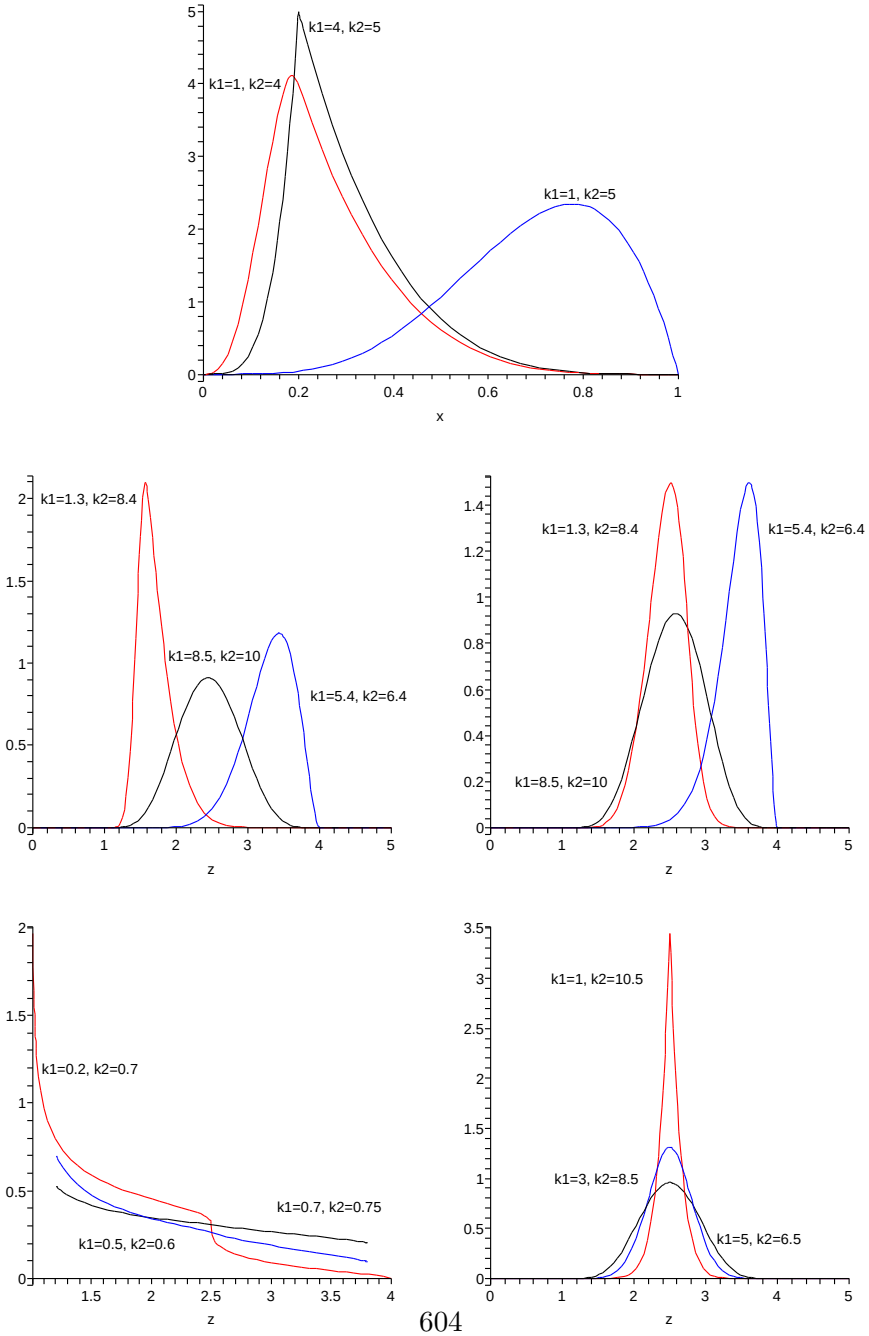


Figure 2: GTSP density functions; upper-left: standard $\theta = 0.2$, $n = 5$ upper right: on $[1, 4]$, $m = 1.6$, $n = 10.5$; middle-left: on $[1, 4]$, $m = 2.8$, $n = 10.5$; middle-right: on $[1, 4]$, $m = 2.5$, $n = 0.8$; bottom: symmetric on $[1, 4]$, $n = 10.5$.

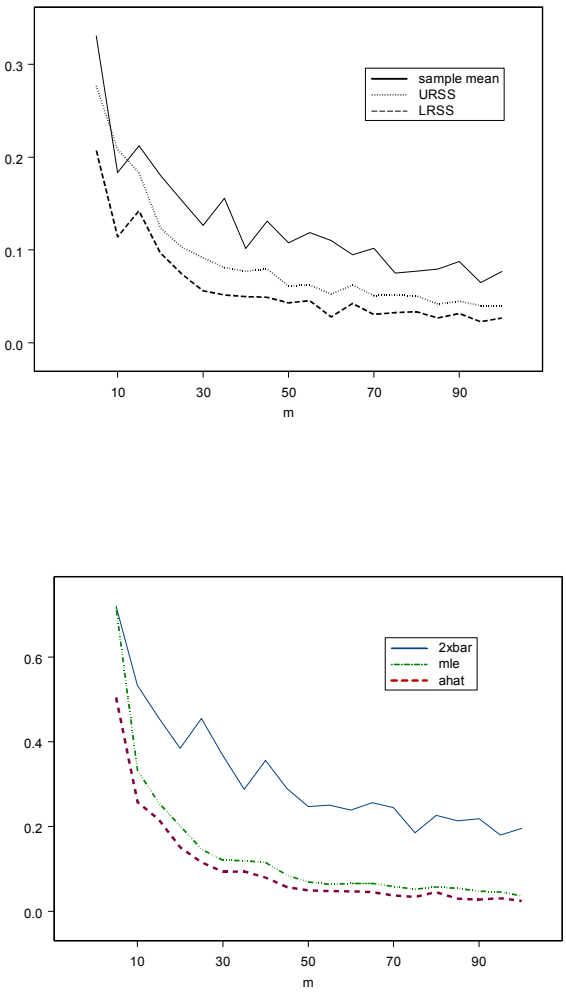


Figure 3: Top: absolute deviation from the mean for LRSS, URSS and sample mean; Below: absolute deviation from $a = 4$ for $2\overline{x}$, $\hat{a}$ and MLE.

Minimum-Kolmogorov-distance estimator of the tail index of heavy- tailed distributions using V-statistics

L. Tafakori, A. R. Nematollahi

Department of Statistics, University of shiraz

Abstract: In this paper, an estimator of the stability parameter α when the parent distributions are in a special class of heavy-tailed distributions is proposed through the idea of minimizing the Kolmogorov distance of distribution functions, which can also be explained as a V-statistic. The advantage of this method, apart from being simple to calculate, is that the estimator can be shown to be uniformly strongly consistent and asymptotically normal.

Keywords: Stable distribution, Heavy-tailed distribution, Minimum-distance estimator, V-statistic.

1 Introduction

Heavy-tailed distributions have been observed in many natural phenomena including telecommunications, hydrology, biology, sociology, geology, and insurance and environmental sciences show large variability. The family of stable distributions have been shown to be very useful and appropriate in modeling such heavy-tailed distributions (Zolotarev, 1986 and Samorodintsky and Taqqu, 1994).The high variability of the stable distributions is one of the reasons they play an important role in modeling. Stable distributions have been used to model such diverse phenomena as gravitational fields of stars, temperature distributions in nuclear reactions, stresses in crystalline lattices, stock market prices and annual rainfall. However, the stable distribution suffers from two major drawbacks. It lacks a simple closed from expression for the probability density function except in a few cases like the Gaussian (with α =2), Cauchy(with α =1), and Levy distribution (with α =0.5). The second problem in using stable distributions with index of stability$\alpha \in (0\,,2\,)$, is that it possesses absolute moments only of order $p < \alpha$. In particular, the variance of these

random variables is infinite except for the case of the normal distribution. This leads to a serious problem in estimation, providing confidence intervals or carrying out inference becomes very difficult. Estimation of tail indices of heavy-tailed distributions has been an active field of research for quite along time, known statistical procedures such as estimation method of Hill (1975). He constructed a simple estimator for the tail index by maximizing the conditional likelihood function. It depends on the number k of largest observations chosen to calculate the estimates. It vary considerably with the choice of k. Dumouchel (1973) proposed a maximum likelihood type algorithm, which although theoretically superior, based on order statistics and are computationally inexpensive. Fama and Roll (1971) proposed the first quantile method which was improved later by McCulloch (1986). For further estimation methods, we consider the empirical characteristic function as an estimate of the theoretical one. Clearly, this is an unbiased and consistent estimator. It is possible to calculate estimators of the unknown parameters from the analytical form of the characteristic function. Since the empirical characteristic function is consistent the derived estimators are easily seen to be consistent as well. This kind of estimator is due to Press (1972). In this article, we propose an estimate of the stability parameter α when the parent distributions are in a special class of heavy-tailed distributions, strictly stable distributions through the idea of minimizing the Kolmogorov distance of distribution functions, which can also be explained as a V-statistic. The advantage of this method, apart from being simple to calculate, is that this estimator is uniformly strongly consistent and asymptotically normal.

2 Preliminaries

We denote an α-stable random variable by $S_\alpha(\sigma\,,\,\beta\,,\,\mu)$, where $-1 \leq \beta \leq 1$, $\sigma \geq 0$, $0 < \alpha \leq 2$ are the indexes of stability, skewness, scale and shift, respectively. The problem of not having a closed form for the density is somewhat mitigated by the presence of a closed form

for the characteristic function given by

$$E(\exp i\theta X) = \begin{cases} \exp\left\{-\sigma^{\alpha}|\theta|^{\alpha}\left(1 - i\beta\left((sign\theta)\tan\frac{\pi\alpha}{2}\right) + i\mu\theta\right\}\right., & \alpha \neq 1 \\ \exp\left\{-\sigma|\theta|\left(1 + i\beta\frac{2}{\pi}(sign\theta)Ln\ |\theta|\right) + i\mu\theta\right\}\right., & \alpha = 1. \end{cases}$$

One important property of stable distributions is that the family of them is closed under convolution. It means if X,Y are independent and identically distributed stable random variables, then for any given positive numbers A and B , there exist positive C and real D, such that

$$AX + BY \stackrel{d}{=} CX + D\,,$$

where $C^{\alpha} = A^{\alpha} + B^{\alpha}$and $\alpha \in (0\,,2\,)$ is the exponent of the stable distribution. The case $D = 0$ corresponds to strictly stable distribution of X.

3 The construction of the estimator

When the involved stable distribution F is supposed to be strict, we know from the definition, that , if $X_1, X_2, ..., X_m$ are independent random variables with the same distribution F ,then the sum of them is also stably distributed that is

$$\frac{X_1 + X_2 + ... + X_m}{m^{1/\alpha}} \xrightarrow{d} X_1,$$

where $\xrightarrow{d}$ means equivalence in distribution.
Suppose now $X_1, X_2, ..., X_n$ are i.i.d copies of a strictly distributed stable random variable $X$with exponent α_0 and let h denote the kernel

$$h_t(x_1, x_2, ..., x_m) = I_{(X_1+X_2+...+X_m \leq m^{1/\alpha} \times t)} - \frac{1}{m}(I_{(X_1 \leq t)} + I_{(X_2 \leq t)} + ... + I_{(X_m \leq t)})$$

where $t \in R$ and $0 < \alpha < 2$. A V-statistic with kernel h can be defined by

$$V_n(\alpha, t) = \frac{1}{n^m} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} ... \sum_{i_m=1}^{n} h_t(X_{i_1}, X_{i_2}, ..., X_{i_m}).$$

is a strongly consistent $V_n(\alpha, t)$ The properties of V-statistic guarantee that
estimator of

$$E(h_t(X_1, X_2, ...X_m)) = P(X \leq m^{\frac{\alpha_0 - \alpha}{\alpha_0 \alpha}} \times t) - p(X \leq t)$$

Let

$$A(\alpha) := \sup_{t \in R} |E h_t(X_1, X_2, ..., X_m)|$$

We know that $A(\alpha)$ is monotone in $\frac{1}{\alpha} - \frac{1}{\alpha_0}$ and attains its minimum only when $\alpha = \alpha_0$. Now we have the following estimator of α based on the above arguments,

$$\hat{\alpha}_{md} = \arg \min_{\alpha \in [0,2]} \sup_{t \in R} |V_n(\alpha, t)| . \tag{1}$$

The proof of the Theorems 3.1 and 3.2 can be found in Nematollahi and Tafakori , 2007.

Theorem 3.1 Suppose $X_1, X_2, ..., X_n$ are i.i.d. copies of X with strictly α_0- stable distribution. Let $\hat{\alpha}_{md}$ given by 2.1. Then $\hat{\alpha}_{md}$ is a strongly consistent estimator of α_0.

Theorem 3.2 Suppose $X_1, X_2, ..., X_n$ are i.i.d. observations with distribution F, which is attracted to someα-stable law, and let $\hat{\alpha}_{md}$ be defined by (1). We have

$$\sqrt{n}\, S_g^{-1}(\hat{\alpha}_{md} - E\hat{\alpha}_{md}) \xrightarrow{d} N(0, 1), \quad n \to \infty \ .$$

where

$$S_g^2 := \frac{1}{n-1}\sum_{i=1}^{n}\left(U_{n-1}^i - \bar{U}_{n-1}\right)^2 \xrightarrow{P} m^2\xi_1\left[g'(\eta)\right]^2$$

and

$$\xi_1 = Var\left(E\left(h\left(X_1,\ldots,X_m\right)|X_1\right) - E\,h\left(X_1,\ldots,X_m\right)\right)$$

Note that the estimator $\overset{\wedge}{\alpha}_{md}$ is in fact the minimum- distance estimator with respect to a special distance Kolmogorov –distance:

$$d_k(F,G) = Sup_{t\in R}\left|F(t) - G(t)\right|,$$

where F,G are two distribution functions.
Suppose now that $X_1, X_2, ..., X_n$ are i.i.d copies of the parent population X, with distribution function $F \in$ F, a class of distribution functions. Let

$$G_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(x_i\leq x)}$$

be the empirical distribution function, then the corresponding minimum-distance of F, based on Kolmogorov-distance is the function $F_0 \in F$, such that

$$d_k(F_0, G_n) = \min_{F\in F} d_k(F, G_n).$$

The consistency of the minimum-distance estimator under some conditions holds according to the following proposition.
Proposition (Parr and Schucany, 1980**).** Let T_n be the asymptotic minimum-distance estimator based on the empirical distribution G_n with respect to Kolmogorov –distance

$$d_k(F,G) = \sup_{x\in R}\left|F(x) - G(x)\right|.$$

If

(i) there is a point $\theta_0 \in \Omega$, the parameter space, such that

$$\inf_{\theta\in\Omega} d_k(F_\theta, G) = d_k(F_{\theta_0}, G),$$

and
(ii) $\lim_{l\to\infty} d_k(F_{\theta_L}, G) = d_k(F_{\theta_0}, G)$ implies $\lim_{l\to\infty} \theta_l = \theta_0$,
then

$$P(\lim_{n\to\infty} T_n = \theta_0) = 1.$$

In our assumptions here, the sample comes from some stable population F with exponent $\alpha_0 \in [0, 2]$.
The sum-preserving property of the stable distributions guarantees that,

$$\frac{X_1 + X_2 + ... + X_m}{m^{1/\alpha}} \stackrel{d}{=} X_1,$$

hence for $\alpha_0 \in [0, 2]$, we have also a corresponding empirical distribution function of V-statistic structure,

$$\widetilde{F}_n(\alpha, t) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} I(\frac{X_1 + X_2 + \ldots + X_m}{m^{\frac{1}{\alpha_0}}} \leq t)$$

If α_0 is the exponent, then

$$EI(\frac{X_1 + X_2 + \ldots + X_m}{m^{\frac{1}{\alpha_0}}} \leq t) = P(X \leq t) = F(t),$$

and the boundedness of the kernel guarantees the strong consistency of the empirical distribution function $\underset{n}{\widetilde{F}}(\alpha_0, .)$ with respect to $F(.)$.
Finally, we consider the class of distributions

$$\mathrm{F} = \{\underset{n}{\widetilde{F}}(\alpha, .) \quad : \alpha \in [0, 2]\}.$$

A distribution or equivalently a value of $\alpha_0 \in [0, 2]$ is to be chosen by minimizing the Kolmogorov-distance between $\underset{n}{\widetilde{F}}(\alpha, .)$ and the empirical distribution function $G_n(.) = \frac{1}{n}\sum_{i=1}^{n} I_{(x_i \leq .)}$. In the other words, we are choosing the value of $\underset{md}{\hat{\alpha}}$, such that

$$\min_{\alpha \in [0,2]} d_k(\underset{n}{\widetilde{F}}(\alpha, .), G_n(.)) = d_k(\underset{n}{\widetilde{F}}(\underset{md}{\hat{\alpha}}, .), G_n(.))$$

From the properties of minimum- distance estimators, this estimator is consistent and also it is robust, in the sense of M-estimators.

References

[1] Dumouchel, W. H. (1973) On the asymptotic normality of the maximum likelihood estimate when sampling from a stable distribution, Ann.Statist.1, pp. 957.

[2] Fama, E. F and Roll, R. (1971) Parameter estimates for symmetric stable distributions. J. Amer. Statist. Assoc. 66. pp. 331-338.

[3] Hill, B.M. (1975) A simple general approach to inference about the tail of a distribution, Ann. Statist, 3, 1163-1174.

[4] McCulloch, J. H. (1986) Simple consistent estimators of stable distribution parameters. Commun. Statist. Simula., 15, No, 4, pp. 1109-1136.

[5] Nematollahi A.R.and Tafakori L. (2007) On the minimum-distance estimation of the tail index of heavy-tailed distributions using v-statistics, submitted for publication.

[6] Press S. J. (1972) Estimation in univariate and multivariate stable distributions, JASA, 67, No.340, 842-846.

[7] Samorodnitsky, G. and Taqqu, M. S. (1994) Stable non-Gaussian random processes, Chapman & Hall.

[8] Parr W. C. and Schucancy W. R. (1980) Minimum-distance and robust estimation, JASA.75, no.371, 616-624.

[9] Zolotarev, V. M. (1986) One-dimensional stable distributions, Translations of mathematical Monographs, 65.

On the Inactivity Time of Lower k-Records and Associated Stochastic Orders

Mahdi Tavangar, and Majid Asadi

Department of Statistics, University of Isfahan

Abstract: In the present study we will propose a concept of mean inactivity time based on lower k-record values and investigate some results concerning this notion. Let $Z_1^{(k)}, Z_2^{(k)}, \ldots$ denote the lower k-record values based on a sequence of i.i.d. random variables with common distribution function F. Consider the *general inactivity time* of lower k-records $(t - Z_n^{(k)} \mid Z_m^{(k)} \leq t)$ and denote its expectation as $M_{n,m}^{(k)}(t)$. We obtain a representation of F in terms of $M_{n,1}^{(k)}(t)$ and $M_{n-1,1}^{(k)}(t)$. Hence the parent distribution function can be recovered perfectly from $M_{n,1}^{(k)}(t)$. More generally we show that $M_{n,m}^{(k)}(t)$ and $M_{n-m,1}^{(k)}(t)$, for some value of m and n $(m < n)$, uniquely identify F. It is shown that $M_{n,m}^{(k)}(t)$ is an increasing function of t when the reversed hazard rate is decreasing. When the reversed hazard rates of two distributions are ordered, then the corresponding mean inactivity times are also ordered, but in the opposite sense. In this paper, some new characterization results for the power function distribution based on lower k-record values are presented. Also some stochastic orders for the inactivity time of lower k-records in one and two sample problems are studied.

Keywords: Characterization, Ordered Random Variables, Power Function Distribution, Reversed Hazard Rate, Stochastic Ordering.

1 Introduction

Because of importance of record values in many field of applications these kind of ordered data have been extensively studied in the literature. Among the books which have devoted to the theory and applications of record values one can refer to Arnold et al. (1998) and Ahsanullah (1995). Consider a sequence $X_1, X_2, \ldots$ of i.i.d. random variables with common continuous distribution function F and probability density function f. Denote by $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ the order statistics corresponding to observations $X_1, X_2, \ldots, X_n$. For

fixed $k \geq 1$, the nth lower k-record time $L_k(n)$, $n \geq 1$, with $L_k(1) = 1$ is defined to be

$$L_k(n+1) = \min\{j > L_k(n) : X_{k:L_k(n)+k-1} > X_{k:j+k-1}\}, \quad n = 1, 2, \ldots,$$

and the nth lower k-record value is defined as $Z_n^{(k)} = X_{k:L_k(n)+k-1}$, $n \geq 1$. For $n = 1$, we have $Z_1^{(k)} = X_{k:k} = \max\{X_1, X_2, \ldots, X_k\}$. The probability density function of $Z_m^{(k)}$ is given by

$$f_{Z_m^{(k)}}(u) = k^m \frac{\{-\log F(u)\}^{m-1}}{(m-1)!}\{F(u)\}^{k-1} f(u), \quad -\infty < u < \infty, \quad (1)$$

and the joint probability density function of $Z_m^{(k)}$ and $Z_n^{(k)}$, $1 \leq m < n$, is given by

$$f_{Z_m^{(k)}, Z_n^{(k)}}(u, v) = k^n \frac{\{-\log F(u)\}^{m-1}}{(m-1)!} \frac{\{-\log F(v) + \log F(u)\}^{n-m-1}}{(n-m-1)!} \times \frac{f(u)}{F(u)}\{F(v)\}^{k-1} f(v), -\infty < v < u < \infty.$$

Using (1) the survival function of $Z_m^{(k)}$ at $t > 0$ can be written as

$$\begin{aligned} P(Z_m^{(k)} > t) &= \frac{1}{\Gamma(m)} \int_0^{-k \log F(t)} z^{m-1} e^{-z} dz \\ &= \frac{1}{\Gamma(m)} \gamma(m; -k \log F(t)) \\ &= \sum_{j=m}^{\infty} \frac{1}{j!}\{-k \log F(t)\}^j \{F(t)\}^k = P(Y_t \geq m), \quad (2) \end{aligned}$$

where $\Gamma(m) = (m-1)!$, $\gamma(m; x) = \int_0^x z^{m-1} e^{-z} dz$ is the incomplete Gamma function and Y_t is a Poisson random variable with mean $-k \log F(t)$. See, for example, Abramowitz and Stegun (1964) and Arnold et al. (1998).

In the present paper, the inactivity time of lower k-records has been studied. We define its mean as a reliability measure for the system of generating data. Also we provide some results on characterization and stochastic ordering based upon lower k-records.

2 The mean inactivity time of lower k-records

To introduce our new concept let us look at the following example. Consider the 100 meters, which is one of the most prestigious events in athletics. The current word record is 9.72 seconds. Now what is the average of the difference between the next record and the current record given that current record is 9.72? More generally, let $Z_\ell^{(1)}$ be the ℓth lower record and it is known that $Z_\ell^{(1)} = t$. Then the quantity of interest is the average of the conditional random variable $(t - Z_n^{(1)} \mid Z_\ell^{(1)} = t)$, $n > \ell$. Usually in any sport competition, a list of the top m scores is updated throughout. Let participants with records $Z_{\ell-m+1}^{(1)} = t_1, Z_{\ell-m+2}^{(1)} = t_2, \ldots, Z_{\ell-1}^{(1)} = t_{m-1}, Z_\ell^{(1)} = t$ have obtained top m scores. Then from the Markovian property of a sequence of record values (see, for example, Arnold et al., 1998) we may conclude that

$$\left(t - Z_n^{(1)} \mid Z_{\ell-m+1}^{(1)} = t_1, Z_{\ell-m+2}^{(1)} = t_2, \ldots, Z_{\ell-1}^{(1)} = t_{m-1}, Z_\ell^{(1)} = t\right) \stackrel{d}{=} \left(t - Z_n^{(1)} \mid Z_\ell^{(1)} = t\right).$$

A similar result holds based on lower k-records. Hence to study distribution of the nth lower k-record, one can only use the information in highest score of the list. Collins et al. (2007) studied, among other things, the number of times the list changes and the waiting times between changes.

Let the conditional random variable $(t - Z_n^{(k)} \mid Z_\ell^{(k)} = t)$ is denoted by $R_{n,\ell}^{(k)}(t)$. We also define the *general inactivity time* of lower k-records as $S_{n,m}^{(k)}(t) = (t - Z_n^{(k)} \mid Z_m^{(k)} \leq t)$. The following result indicates that distributional properties of the $R_{n,\ell}^{(k)}(t)$ depend on n and ℓ only through the difference between n and ℓ.

Lemma 1 *For $1 \leq \ell < n$, we have $R_{n,\ell}^{(k)}(t) \stackrel{d}{=} S_{n-\ell,1}^{(k)}(t)$.*

In view of Lemma 1, one can study properties of the general inactivity time $S_{n,m}^{(k)}(t)$ and then apply them for $R_{n,\ell}^{(k)}(t)$.

Remark 1 Note that the survival function of $S_{n,1}^{(k)}(t)$ can be expressed as

$$P(t - Z_n^{(k)} > x \mid Z_1^{(k)} \leq t) = 1 - \frac{1}{\Gamma(n)}\gamma(n; \eta_t(t-x))$$
$$= \sum_{j=0}^{n-1} \frac{1}{j!}\{\eta_t(t-x)\}^j e^{-\eta_t(t-x)} = P(Z_{t,x} \leq n-1),$$

where $Z_{t,x}$ is a Poisson random variable with mean $\eta_t(t-x)$.

The expectation of $S_{n,m}^{(k)}(t)$ is defined as $M_{n,m}^{(k)}(t) = E(t - Z_n^{(k)} \mid Z_m^{(k)} \leq t)$. Then for $m = 1$ we have

$$\begin{aligned} M_{n,1}^{(k)}(t) &= \int_0^t \frac{1}{\Gamma(n)} \int_{\eta_t(x)}^{\infty} z^{n-1} e^{-z} dz\, dx \\ &= \sum_{j=0}^{n-1} \frac{1}{j!} \int_0^t \{\eta_t(x)\}^j e^{-\eta_t(x)} dx, \end{aligned}$$

and for $m = 2, \ldots, n-1$ we get

$$M_{n,m}^{(k)}(t) = \sum_{j=0}^{m-1} p_j(t) M_{n-j,1}^{(k)}(t),$$

where

$$p_j(t) = \frac{\{-k \log F(t)\}^j / j!}{\sum_{i=0}^{m-1} \{-k \log F(t)\}^i / i!}.$$

The last representation show that $M_{n,m}^{(k)}(t)$ can be rewritten as a convex combination of the mean inactivity times $M_{n-m+1,1}^{(k)}(t), M_{n-m+2,1}^{(k)}(t), \ldots, M_{n,1}^{(k)}(t)$.

Example 1 *(Power function distribution)* Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables from a population with distribution $F(x) = (\frac{x}{\beta})^\alpha$, $x \in [0, \beta)$, $\alpha, \beta > 0$. Then

$$M_{n,1}^{(k)}(t) = t\left\{1 - \left(\frac{k\alpha}{k\alpha+1}\right)^n\right\},$$

which is linear in t. For $m = 2, \ldots, n-1$ we have

$$M_{n,m}^{(k)}(t) = \sum_{j=0}^{m-1} p_j(t) M_{n-j,1}^{(k)}(t), \tag{3}$$

where

$$p_j(t) = \frac{\frac{(k\alpha)^j}{j!}\{-\log(t/\beta)\}^j}{\sum_{i=0}^{m-1}\frac{(k\alpha)^i}{i!}\{-\log(t/\beta)\}^i}.$$

Note that $M_{n,m}^{(k)}(t)$ is non-decreasing in t.

In next theorems, we present some structural properties of $M_{n,m}^{(k)}(t)$.

Theorem 1 (Tavangar and Asadi, 2008) *We have the following:*

(i) *For fixed values of m, k and t, $M_{n,m}^{(k)}(t)$ is an increasing function of n, $n = m+1, m+2, \ldots$. Moreover $M_{n,m}^{(k)}(t)$ is strictly increasing in n if F is strictly increasing on $(0,t)$.*

(ii) *For fixed values of n, k and t, $M_{n,m}^{(k)}(t)$ is a decreasing function of m, $m = 1, 2, \ldots, n-1$. Furthermore $M_{n,m}^{(k)}(t)$ is strictly decreasing in m if F is strictly increasing on $(0,t)$.*

(iii) *For fixed values of n, m and t, $M_{n,m}^{(k)}(t)$ is a decreasing function of k, $k = 1, 2, \ldots$. Moreover $M_{n,m}^{(k)}(t)$ is strictly decreasing in k if F is strictly increasing on $(0,t)$.*

Let X be a random variable with absolutely continuous distribution function F and probability density function f. The reversed hazard rate of X is defined as $r(t) = f(t)/F(t)$. The next result shows that $M^{(k)}_{n,m}(t)$, for some values of m and n ($m < n$), uniquely identifies F through the reversed hazard rate of mth lower k-record.

Theorem 2 *Assume that F is absolutely continuous and strictly increasing on its support. Then the $M^{(k)}_{n,m}(t)$ determines the reversed hazard rate of mth lower k-record as follows:*

$$r_{Z^{(k)}_m}(t) = \frac{1 - \frac{d}{dt}M^{(k)}_{n,m}(t)}{M^{(k)}_{n,m}(t) - M^{(k)}_{n-m,1}(t)}.$$

It is known that the reversed hazard rate function univocally characterizes the distribution function. Hence from Theorem 2 we conclude that the distribution function of the mth lower k-record is given by

$$F_{Z^{(k)}_m}(t) = \exp\left\{ - \int_t^\infty \frac{1 - \frac{d}{dx}M^{(k)}_{n,m}(x)}{M^{(k)}_{n,m}(x) - M^{(k)}_{n-m,1}(x)}\, dx \right\}.$$

But the distribution function of $Z^{(k)}_m$ uniquely identifies F. (See for example (2).) Therefore knowledge of $M^{(k)}_{n,m}(t)$ for some specific values of n, m and k will be adequate to determine F. In particular, when $m = 1$ we get the following result.

Corollary 1 *Assume that F is absolutely continuous and strictly increasing on its support. Then F can be expressed as*

$$F(t) = \exp\left\{ -\frac{1}{k} \int_t^\infty \frac{1 - \frac{d}{dx}M^{(k)}_{n,1}(x)}{M^{(k)}_{n,1}(x) - M^{(k)}_{n-1,1}(x)} dx \right\},$$

for some $n > 1$.

The next theorem determines the behavior of $M^{(k)}_{n,m}(t)$ when the reversed hazard rate decreases.

Theorem 3 *Let the reversed hazard rate $r(t)$ is decreasing. Then for each n, m and k, the mean inactivity time $M_{n,m}^{(k)}(t)$ is an increasing function of t.*

The following theorem gives an ordering for the mean inactivity times of two sequences of lower k-th record values in the case where the the reversed hazard rates are ordered.

Theorem 4 *Consider two populations with absolutely continuous distributions F and G, and mean inactivity times $M_{n,m}^{(k)}(t)$ and $H_{n,m}^{(k)}(t)$, respectively. Let $r_F(t)$ and $r_G(t)$ be the reversed hazard rates of F and G, respectively. If $r_F(t) \geq r_G(t)$, then $M_{n,m}^{(k)}(t) \leq H_{n,m}^{(k)}(t)$.*

Proof. First note that $r_F(t) \geq r_G(t)$ if and only if $F(t)/G(t)$ is an increasing function of t. Thus for $x < t$, $\frac{F(x)}{F(t)} \leq \frac{G(x)}{G(t)}$ and

$$M_{n,1}^{(k)}(t) = \int_0^t \frac{1}{\Gamma(n)} \int_{-k\log\left(\frac{F(x)}{F(t)}\right)}^{\infty} z^{n-1}e^{-z}dz$$
$$\leq \int_0^t \frac{1}{\Gamma(n)} \int_{-k\log\left(\frac{G(x)}{G(t)}\right)}^{\infty} z^{n-1}e^{-z}dz = H_{n,1}^{(k)}(t).$$

Therefore using Equation (3), we get

$$M_{n,m}^{(k)}(t) \leq \sum_{j=0}^{m-1} p_j(t) H_{n-j,1}^{(k)}(t),$$

where

$$p_j(t) = \frac{u_t^j/j!}{\sum_{i=0}^{m-1} u_t^i/i!},$$

and $u_t = -k\log F(t)$. Now it is enough to show that

$$\sum_{j=0}^{m-1} p_j(t) H_{n-j,1}^{(k)}(t) \leq \sum_{j=0}^{m-1} q_j(t) H_{n-j,1}^{(k)}(t) = H_{n,m}^{(k)}(t),$$

where $q_j(t)$ is the same as $p_j(t)$ with u_t replaced by $v_t = -k \log G(t)$. But the last inequality is true since one can prove that

$$\sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \frac{u_t^i}{i!} \frac{v_t^j}{j!} \left\{ H_{n-j,1}^{(k)}(t) - H_{n-i,1}^{(k)}(t)(t) \right\} \geq 0.$$

Hence we have the result.

3 Characterizations of the power function distribution

In the literature the mean inactivity time of a random variable with distribution F is defined as $m^*(t) = E(t - X \mid X \leq t)$. One can easily show that

$$m^*(t) = \frac{\int_{-\infty}^{t} F(x)dx}{F(t)},$$

for each t with $F(t) > 0$. Next theorem gives a characterization of the power function distribution when $M_{n,1}^{(k)}(t)$ and the parent mean inactivity time $m^*(t)$ are proportional.

Theorem 5 *Consider an absolutely continuous distribution function F with the mean inactivity time $m^*(t)$ and corresponding mean inactivity time of lower k-record $M_{n,1}^{(k)}(t)$. Assume that the support of F is $[0, 1]$, and $m^*(1) = (\alpha + 1)^{-1}$. Then*

$$M_{n,1}^{(k)}(t) = c_{n,k} m^*(t), \ t > 0,$$

where $c_{n,k} = (\alpha+1) \left\{ 1 - (\frac{k\alpha}{k\alpha+1})^n \right\}$, if and only if F is a power function distribution with parameter α, for some $\alpha > 0$; i.e. $F(x) = x^\alpha$, $x \in [0, 1)$.

Proof. The 'if' part of the theorem follows from Example 1 and the fact that for the power function distribution $m^*(t) = (\alpha + 1)^{-1} t$. To prove the 'only if' part, first note that the reversed hazard rate of $Z_1^{(k)}$

is the same as $kr(t)$. Hence using the assumption $M_{n,1}^{(k)}(t) = c_{n,k}m^*(t)$, we have

$$\begin{aligned} r(t) &= \frac{1}{k}\left\{\frac{1-\frac{d}{dt}M_{n,1}^{(k)}(t)}{M_{n,1}^{(k)}(t)-M_{n-1,1}^{(k)}(t)}\right\} \\ &= \frac{1}{k}\left\{\frac{1-c_{n,k}\frac{d}{dt}m^*(t)}{(c_{n,k}-c_{n-1,k})m^*(t)}\right\}. \end{aligned} \tag{4}$$

Note that, after some simplification,

$$k(c_{n,k}-c_{n-1,k}) = \left(\frac{\alpha+1}{\alpha}\right)\left(\frac{k\alpha}{k\alpha+1}\right)^n.$$

The fact that $r(t) = \{1-\frac{d}{dt}m^*(t)\}/m^*(t)$ together with Equation (4) imply that $\frac{d}{dt}m^*(t) = (\alpha+1)^{-1}$. Since $m^*(1) = (\alpha+1)^{-1}$, we have $m^*(t) = (\alpha+1)^{-1}t$. Using the inversion formula

$$F(x) = \frac{m^*(1)}{m^*(x)}\exp\left\{-\int_x^1 \frac{1}{m^*(t)}\,dt\right\}, \quad 0 \le x < 1,$$

we have the parent distribution function F to be $F(x) = x^\alpha,\ x \in [0,1)$. The proof is complete.

In the following we characterize the power function distribution based on the property that the reversed hazard rate $r(t)$ and $M_{n,1}^{(k)}(t)$ are reciprocal. For proof we refer the reader to Tavangar and Asadi (2008).

Theorem 6 *Consider an absolutely continuous distribution function F with the reversed hazard rate $r(t)$. Assume that the support of F is $[0,1]$. Then*

$$r(t)M_{n,1}^{(k)}(t) = c_{n,k}\left(\frac{\alpha}{\alpha+1}\right),\ t \in [0,1],$$

for some $k \ge 1$ and two consecutive values of n (e.g. $n-1$ and n), if and only if F is a power function distribution with parameter α, for some $\alpha > 0$; i.e. $F(x) = x^\alpha,\ x \in [0,1)$.

4 Some stochastic orders on lower k-records

Many notions of stochastic orderings are introduced yet and they have been discussed in detail in Shaked and Shanthikumar (1994). Let X and Y be two random variables with distribution functions F and G, and density functions f and g, respectively. Let $l_X(l_Y)$ and $u_X(u_Y)$ be the left and the right extremity of support of $X(Y)$. X is said to be *stochastically* smaller than Y (denoted $X \leq_{st} Y$) if $1 - F(t) \leq 1 - G(t)$ for all t. X is said to be smaller than Y in *reversed hazard rate* ordering (denoted $X \leq_{rh} Y$) if $F(t)/G(t)$ is decreasing in $t \in (\min(l_X, l_Y), \infty)$. X is said to be smaller than Y in *likeihood ratio* order (denoted $X \leq_{lr} Y$) if $g(t)/f(t)$ is increasing in $t \in (-\infty, \max(u_X, u_Y))$. In this section some stochastic orders for the inactivity time of lower k-record $S^{(k)}_{n,m}(t) = (t - Z^{(k)}_n \mid Z^{(k)}_m \leq t)$ in one and two sample problems are presented.

In the following theorem, some comparisons of the random variables $S^{(k)}_{n,m}(t)$ have been considered in one sample problem.

Theorem 7 *(i) We have the following stochastic orders between the inactivity time of lower k-records:*

$$S^{(k)}_{n-1,m}(t) \leq_{lr} S^{(k)}_{n,m}(t) \leq_{lr} S^{(k)}_{n,m-1}(t).$$

(ii) We have $S^{(k)}_{n,m}(t) \leq_{lr} S^{(k-1)}_{n,m}(t)$.
Combining all stochastic orders, we have $S^{(k)}_{n,m}(t) \leq_{lr} S^{(k')}_{n',m'}(t)$, *where* $1 \leq m' \leq m$, $n' \geq n$ *and* $1 \leq k' \leq k$.

Proof. We only prove part (ii). Note that the probability density function of $S^{(k)}_{n,m}(t)$ has the form

$$h^{(k)}_{n,m,t}(x) = \sum_{j=0}^{m-1} p_j(t) h^{(k)}_{n-j,1,t}(x),$$

where

$$h^{(k)}_{n,1,t}(x) = \frac{k}{\Gamma(n)}\, r(t-x)\{\eta_t(t-x)\}^{n-1} e^{-\eta_t(t-x)}, \quad x \in [0,t].$$

Thus

$$\frac{h^{(k-1)}_{n,m,t}(x)}{h^{(k)}_{n,m,t}(x)} = \frac{\sum_{i=0}^{m-1}\{-k\log F(t)\}^i/i!}{\sum_{i=0}^{m-1}\{-(k-1)\log F(t)\}^i/i!}$$
$$\cdot\frac{\sum_{j=0}^{m-1}\binom{n-1}{j}\{-(k-1)\log F(t)\}^j(k-1)(\frac{F(t-x)}{F(t)})^{k-1}}{\sum_{j=0}^{m-1}\binom{n-1}{j}\{-k\log F(t)\}^j k(\frac{F(t-x)}{F(t)})^k\left\{-k\log(\frac{F(t-x)}{F(t)})\right\}^{n-j-1}}$$
$$\cdot\frac{\left\{-(k-1)\log(\frac{F(t-x)}{F(t)})\right\}^{n-j-1}}{\sum_{j=0}^{m-1}\binom{n-1}{j}\{-k\log F(t)\}^j k(\frac{F(t-x)}{F(t)})^k\left\{-k\log(\frac{F(t-x)}{F(t)})\right\}^{n-j-1}}$$
$$= \frac{\sum_{i=0}^{m-1}\{-k\log F(t)\}^i/i!}{\sum_{i=0}^{m-1}\{-(k-1)\log F(t)\}^i/i!}\left(\frac{k-1}{k}\right)\left(\frac{F(t-x)}{F(t)}\right)^{-1},$$

which is increasing in $x \in (0,t)$ and so $S^{(k)}_{n,m}(t) \leq_{lr} S^{(k-1)}_{n,m}(t)$. The proof is complete.

In the following we obtain an interesting result for comparison of two sequences of inactivity times of lower k-records. The proof can be found in Tavangar and Asadi (2008).

Theorem 8 *Let $X_1, X_2, \ldots$ be i.i.d. random variables with continuous distribution function F and $Y_1, Y_2, \ldots$ be i.i.d. random variables with continuous distribution function G. If $X_1 \leq_{rh} Y_1$ then $(t - Y^{(k')}_n \mid Y^{(k')}_m \geq t) \leq_{st} (t - Z^{(k)}_n \mid Z^{(k)}_m \leq t)$, for some $k' \geq k$, where $Z^{(k)}_n$ is the nth lower k-record corresponding to the sequence $\{X_i\}$ and $Y^{(k)}_n$ is the nth lower k-record corresponding to the sequence $\{Y_i\}$.*

References

[1] Abramowitz, M., and Stegun, I.A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables.* U.S. Government Printing Office, Washington, DC.

[2] Ahsanullah, M. (1995). *Record Statistics*. Nova Science Publishers, Commack, New York.

[3] Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1998). *Records*. Wiley, New York.

[4] Collins, D., Lengyel, T., and Tollisen, G.P. (2007). Keeping scores. *Journal of Statistical Planning and Inference*, **137**, 2208-2213.

[5] Shaked, M., and Shanthikumar, J.G. (1994). *Stochastic Orders and Their Applications*. Academic Press, New York.

[6] Tavangar, M., and Asadi, M. (2008). A note on the mean inactivity time of lower k-records and associated stochastic orders. *Submitted*.

Characterization of Pareto distribution through coincidences of semiparamtric families

Gh. Yari, and M. Rajaei Salmasi

Department of Mathematics, Iran University of Science and Technology

Abstract: The term semiparametric refers to models in which there is an unknown function in addition to an unknown finite dimensional parameter. For example, the binary response model $P(Y = 1|x) = G(\beta x)$ is semiparametric if the function G and the vector of coefficients β are both treated as unknown quantities. A number of semiparametric families suitable for lifetime data are introduced: scale, power, frailty(proportional hazards), age, moment, tilt parameter families. The coincidence of two families provides a characterization of distributions such as Pareto distribution. Characterization of the Weibull, gamma, lognormal, and Gompertz distributions are contained in Albert W.Marshal , Ingram Olkin (2007).

Keywords: Power parameter families, Scale parameter families, Proportional hazards parameter families, Moment parameter families, Age parameter families, Tilt parameter families, Pareto distributions, Characterization.

1 Introduction

Semiparametric families are families that have both a real parameter and a parameter that is itself a distribution. Detailed discussions of semiparametric families are contained in Marshal and Olkin (1997, 2007). Familiar semiparametric families are location and scale families F(ax+b). Lehmann and Casella (1998) use the term group family to represent families obtained by a transformation, that is, F(g(x)), where the function g usually containes parameters. Location and scale families are the prototype of these families. Some semiparametric families that are natural for lifetime data arising from nonnegative random variables are the following:

1. Scale parameter families $\overline{F}(\lambda x)$, $\lambda > 0$;

2. Power parameter families $F(x^\alpha)$, $\alpha > 0$;
3. Frailty parameter or proportional hazards families $\overline{F}(x)^\xi$, $\xi > 0$; Where $\overline{F} = 1 - F$ is the survival function;
4. Moment parameter families ,$\int_0^x z^\beta dF(z)/\mu_\beta$, where μ_β is normalizing constant;
5. Age parameter families $\overline{F}(x+t)/\overline{F}(x)$, $t > 0$;;
6. Tilt parameter families $\frac{\gamma\overline{F}(x)}{F(x)+\gamma\overline{F}(x)}$, $\gamma > 0$;

Albert W.Marshal, Ingram Olkin (2007) discussed: to provide some motivation for what is meant by characterization consider the example in which the underlying distribution is exponential with survival function $\overline{F}(x) = exp(-x)$. The introduction of a scale parameter yields $\overline{F}(x;\lambda) = exp(-\lambda x)$; the introduction of a frailty parameter (proportional hazards parameter) yields $\overline{F}(x;\xi) = [\overline{F}(x)]^\xi$. Clearly these families coincide. Because scale parameter families with survival function $\overline{F}(\lambda x)$ and proportional hazards families with survival function $[\overline{F}(x)]^\xi$ do not in general coincide, it is natural to ask if there are underlying distributions other than the exponential distribution for which the introduction of scale and proportional hazards parameter families lead to the same family. This coincidence of families yields a characterization of the underlying distribution.
Albert W. Marshal, Ingram Olkin (2007) showed that power and frailty (proportional hazards) parameters families coincidence if and only if there exists a function $\xi(.)$ with domain and range $(0,\infty)$ such that $F(x^\alpha) = [\overline{F}(x)]^\xi$ From that it follows that F has the form $\overline{F}(x) = \frac{1}{x}$, $x \geq 1$ is a pareto survival function.
In the present paper we obtain characterizations of pareto distributions. In particular we investigate the coincidence of the following families: power and frailty (proportional hazards) parameter families(Proposition 1), power and moment (Proposition 2), frailty or

proportional hazards and moment (Proposition 3). Each coincidence yields a characterization of the underlying distribution.

2 Main results

Proposition 1. With the underlying distribution F, power and frailty (proportional hazards) parameter families coincide if and only if there exists a function $\xi(.)$ with domain and range $(0, \infty)$ such that $F(x^{\alpha}) = [\overline{F}(x)]^{\xi}$ for all $\lambda > 0$ and all x.
From above it follows that F either has the form

$$\overline{F}(x) = exp\{-b(logx)^{c}\}, \qquad x \geq 1, \qquad b > 0, \qquad c > 0$$

$$= 1, \qquad x < 1,$$

This survival function is the survival function of Pareto (I) if $b = c = 1$.
Proof.
See Albert W. Marshal, Ingram Olkin (2007).

Proposition 2. Denote by the B the set of all β such that the βth moment of the underlying distribution F for all x¿1 is finite. Power and moment parameter families coincide if and only if there exists a function $\beta(.)$¡0 and range B such that

$$\frac{\int_0^x z^{\beta} dF(z)}{\mu_{\beta}} = F(x^{\alpha}), \qquad \alpha > 0. \tag{1}$$

Eq. (1) is satisfied if and only if F is a pareto distribution.

$$F(x) = 1 - x^{-k}, \qquad x > 1.$$

Proof.

We must show that for some $\beta = \beta(.)$ in B, Eq. (1) holds. To solve (1), the first step is to show that (1) implies that F is everywhere differentiable, with density that satisfies the functional equation

$$\frac{x^{\beta(\alpha)} f(x)}{\mu_{\beta(\alpha)}} = \alpha x^{\alpha-1} f(x^{\alpha}) \tag{2}$$

this is true because F is differentiable almost everywhere so is differentiable at x^{α}, since α is a positive and constant.
With x=1, it follows from (2) that, $\alpha\mu_{\beta(\alpha)} = 1$ use this and suppose that $g(X) = xf(x)$, rewrite (2) in the form $\frac{g(x^{\alpha})}{g(x)} = x^{\beta(\alpha)}$; It is sufficient that we solve this functional equation. There are 3 cases for $\alpha > 0$:
i) $\alpha = 1$, then we have $x^{\beta} = 1$
ii) $\alpha > 1$, then

$$g(x) = x^{\beta(\alpha)/\alpha} g(x^1/\alpha)$$
$$g(x^1/\alpha) = x^{\beta(\alpha)/\alpha^2} g(x^1/\alpha^2)$$
$$...$$
$$g(x^1/\alpha^{n-1}) = x^{\beta(\alpha)/\alpha^n} g(x^1/\alpha^n)$$

use these to show that

$$\forall n : g(x) = x^{\beta(\alpha)(\frac{1}{\alpha}+\frac{1}{\alpha^2}+\cdots+\frac{1}{\alpha^n})} g(x^{\frac{1}{\alpha^n}})$$

$$\lim_{n\to\infty}(\frac{1}{\alpha} + \frac{1}{\alpha^2} + \cdots + \frac{1}{\alpha^n}) = \frac{1}{\alpha - 1}, \qquad \lim_{n\to\infty} g(x^{\frac{1}{\alpha^n}}) = g(1) = k$$

$$\Rightarrow g(x) = kx^{\beta/\alpha - 1}$$

iii) $\alpha < 1$, with the similar manner of (ii) we obtain

$$\forall n : g(x) = \frac{g(x^{\alpha^n})}{x^{\beta(\alpha)(\alpha+\alpha^2+\cdots+\alpha^{n-1})}}$$

so

$$g(x) = kx^{\beta/\alpha - 1}$$

finally

$$f(x) = kx^{\frac{\beta}{\alpha-1}-1} \tag{3}$$

we know that f(x) is a density function;

$$\int_1^\infty kx^{\frac{\beta}{\alpha-1}-1}dx = 1 \qquad \Rightarrow \beta = k(1-\alpha);$$

use these to rewrite (3) in the form

$$f(x) = kx^{-(k+1)}, \qquad \Rightarrow F(x) = 1 - x^{-k}, \qquad x > 1.$$

Proposition 3. With the underlying distribution F, Denote by the B the set of all β such that the βth moment of the underlying distribution F is finite, moment and frailty (proportional hazards) parameter families coincidence if and only if there exists a function $\beta(.)$ with domain $(0, \infty)$ and range B such that

$$\frac{\int_0^x z^\beta dF(z)}{\mu_\beta} = \overline{F}(x)^\xi, \qquad \xi > 0, \qquad \beta > 0. \tag{4}$$

From (4) it follows that F has the form

$$F(x) = 1 - (\frac{\sigma}{x})^\alpha, \qquad x > \sigma.$$

Proof.

We require that for some $\beta = \beta(\xi)$ Eq. (4) holds. either f(x)=0 or by differentiating both sides of (4) with respect to x, we have

$$-\xi(1 - F(x))^{\xi-1} = \frac{x^\beta}{\mu_\beta}$$

$$1 - F(x) = (\frac{x}{(-\xi\mu_\beta)^{1/\beta}})^{\frac{\beta}{\xi-1}}$$

This is clearly in the form of Pareto distribution, we suppose that $\xi = \frac{\alpha-\beta}{\alpha}$ and $\mu_\beta = \frac{\alpha\sigma^\beta}{\beta-\alpha}$, then we obtain

$$F(x) = 1 - (\frac{x}{\sigma})^{-\alpha}, \qquad x > \sigma.$$

Proposition 4. With the underlying distribution F, scale and tilt parameter families coincide if and only if there exists a function$\gamma(.)$ with domain and range $(0, \infty)$ such that

$$\frac{\gamma\overline{F}(x)}{F(x) + \gamma\overline{F}(x)} = \overline{F}(\lambda x), \qquad \lambda > 0, \;\; x \geq 0. \tag{5}$$

it follows that for some $\lambda, \alpha > 0$ either

$$\overline{F}(x) = \frac{1}{1 + (\lambda x)^\alpha}, \qquad x > 0,$$
$$\overline{F}(x) = 1, \qquad x \leq 0 \tag{6}$$

that is a Pareto (III) distribution, defined and discussed by Marshal, A.W., Olkin, I.,2007 in Section 11.B.

Before proof of this proposition we discuses a functional equation.

Lemma 5. Suppose that f and g are real functions defined on $(0, \infty)$ and that f is continuous at some points. If

$$f(xy) = g(x)f(y), \qquad forallx, y > 0, \tag{7}$$

then either

$$f(x) = bx^c, \qquad g(x) = x^c, \qquad x > 0, \tag{8}$$

for some constants b and c, or f(x)=0, for all $x > 0$, and g is arbitrary. Proof of this lemma is contained in Aczel, J., 1966.

Proof of proposition 4.

Set $x = 0$ in (5), to obtain $[1 - \gamma(\lambda)][F(0)(F(0))^2] = 0$. Because γ can take any positive value, $[F(0)(F(0))^2] = 0$, and hence either $\overline{F}(0) = 0$ or $\overline{F}(0) = 1$.
Let $\overline{F}(0) = 1$, so that $\overline{F}(x) = 1, \qquad x \leq 0$. Such a survival function satisfies (6) for all $x < 0$. For $x \geq 0$, rewrite (5) in the form

$$\frac{1}{\overline{F}(\lambda x)} = \frac{F(x)}{\gamma(\lambda)\overline{F}(x)} + 1;$$

now apply this equation to obtain

$$\frac{\overline{F}(\lambda x)}{F(\lambda x)} = \gamma(\lambda)\frac{\overline{F}(x)}{F(x)}, \qquad x, \lambda > 0. \tag{9}$$

Define $G(x) = \frac{\overline{F}(x)}{F(x)}$, in terms of G(x), equation (9) becomes

$$G(\lambda x) = \gamma(\lambda)G(x) \tag{10}$$

This is a functional equation of the form of (7) and can be solved using (8). According to that, there are constants b and c such that

$$\frac{\overline{F}(x)}{F(x)} = bx^c, \qquad x > 0, \qquad \gamma(\lambda) = \lambda^c$$

this means that

$$\overline{F}(x) = \frac{bx^c}{1 + bx^c}, \qquad x > 0.$$

Because $\overline{F}$ is nonnegative, b¿0, and F is a cumulative distribution function so is nondecreasing, $\overline{F} = 1 - F$ is decreasing, c¡0. With the change of parameters c=-a, b=λ^{-a}, the survival function (6) is obtained. Because $\gamma(\lambda) = \lambda^c$, each parameter takes on all values in $(0, \infty)$ as the other parameter ranges over the same interval.

Remark. $\overline{F}(0) = 0$, so that $\overline{F}(x) = 0, \qquad x \geq 0$. For the interval x¡0, let y=-x, and rewrite (5) in the form of (9), with change of x=-y.

This equation is in the form of (7) and can be solved in the manner that (9)is solved; in this way, we obtain

$$F(x) = \frac{(-\lambda x)^{\alpha}}{1 + (-\lambda x)^{\alpha}}, \qquad x < 0,$$

$$F(x) = 0, \qquad x \geq 0. \tag{11}$$

This is the distribution of $1/(-Y)$, where Y has a Pareto (III) distribution. It can be verified directly that the distributions defined by (6) and (11) satisfy (5).

References

[1] Aczel, J., 1966. Lectures on Functional Equations and Their Applications. *Academic Press, NewYork.*

[2] A W.Marshal, I olkin 2007 "Characterization of distributions through coincidences of semiparametric families" *Journal of statistical planning and inference* 137 pp 3618-3625

[3] Castillo, E., Ruiz-Cobo, M.R., 1992. Functional Equations and Modelling in Science and Engineering. *Marcel Dekker, NewYork.*

[4] Lehmann, E.,Casella, G., 1998.Theory of point Estimation. *Springer, New York.*

[5] Marshal, A.W., Olkin, I., 1997. Semiparametric families of life distributions *Technical Report 1997-2, Department of Statistics, Stanford University.*

[6] Marshal, A.W., Olkin, I.,2007. Structure of life distributions: Structure of Nonparametric, Semiparametric and Parametric Families. *Springer, New York.*

Assessment of Digit Preference in age data of national Iranian Census, 2005

Aida Yazdanparast, Mohamad Amin Pourhoseingholi, Ghodrat ollah Roshanaee, Alireza Abadi

Department of Statistics, Allameh Tabatabaii University.

Abstract: Data on age in developing countries are subject to errors, particularly in circumstances where literacy levels are not high. Common problems include considerable age-misstatement in the form of digit preference and systematic over- or under-representation of ages in particular age- and gender-groups. A common error in the age reporting is the tendency of rounding the ages to the nearest figure ending in '0' or '5'or to a lesser extent, in even number. Because of this tendency, commonly known as "digital preference", age heaping occurs at certain ages. The aim of this study was employed Myers and Whipple's index to identify the digit preference in dataset obtained from national census in 2005. The degree of digit preference bias was assessed using a modification of Myers' index show heaping at ages ending in 0 and 5 years, the pattern of heaping is pronounced from age 20 onwards and this is true for both urban and rural population. The Whipple's index shows that males have higher tendency of age heaping than females in rural areas where it has been observed reverse in urban areas. It can be deduced from the analysis that the quality of age reporting for the 2005 census data was poor as compared to the 1995 census data. However, it was of better quality than the 1985 census data. This may suggest that both males and females tend to misreport their ages before age 60. The magnitude of digit preference bias seems to be reducing with the passage of time.

Keywords: Digit preference, Census, Myers' index, Whipple index.

1 Introduction

Age is usually defined in demographic literature as the number of completed years, which have elapsed since the birth of the individual. In other words, the census data on age should usually refer to ages of

aida.yazdanparast@yahoo.com

individuals at their last birthday. Age structure is a crucial component in demographic analysis as it provides a quick and ready tool for mapping the broad contours of demographic history of a population. Similarly, the future demographic events are influenced to a large extent by the present sex-age structure, other things being constant. Ewbank (1981) discussed at the length about the effect of age misreporting on the parental survival technique for estimating mortality. He did a simulation exercise to demonstrate the effect that age exaggeration has on estimated life expectancy. The results showed that age exaggeration of approximately 2.5 years will bias the estimated of life expectancy upward by approximately the same amount. Good age-reporting is a crucial prerequisite for accurate estimates of age-specific fertility rates, which relate births to the age of the mother at the time of the birth. If women's ages are misstated, even an accurate enumeration of the total births by each woman will result in distortions in age-specific fertility rates and, if age misreporting is systematically related in any way to marital status and/or parity, there will be systematic biases in fertility estimates.

Though conceptually, collection of information on age seems to be a simple item but age returns in the censuses were found to be far from the true ages for a large part of the population. Apart from differential under-enumeration in various ages, the age data suffers from distortion owing to preferences for certain ages and digits due to social, cultural and legal habit and norms observed in a society. A common error in the age reporting is the tendency of rounding the ages to the nearest figure ending in '0' or '5'or to a lesser extent, in even number. Because of this tendency, commonly known as "digital preference", age heaping occurs at certain ages. This error is quite common in many developing countries.

The aim of this study was employed Myers and Whipple's index to identify the digit preference in dataset obtained from national census of Iranian population in 2005.

2 Methods

The data used for preparing this paper comes from the full 2005 census. Two standard indices used for this purpose are the Whipple's and Myer's indices. Whipple's index assumes uniform distribution of population in a five-year range and aims to detect heaping on terminal digits '0' and '5' in the range 23 to 62 years. Theoretically, the index varies between 100, representing no preference for '0' or '5' and 500, indicating that only ages ending in '0' and '5' were reported. It must be noted that, Whipple's index has several limitations one of which is the assumption of uniform distribution of the population. This assumption might not always hold and as a result, computed Whipple's index could be less than 100.

The Whipple's Index is usually calculated as:

$$\begin{aligned} &The\ Whipple's\ Index = \\ &\frac{(Sum\ of\ numbers\ at\ ages\ 25{,}30{,}\ldots{,}55{,}60) \times 100 \times 5}{Total\ number\ between\ ages\ 23\ and\ 62} \end{aligned} \quad (1)$$

The value of the Whipple's Index in a population with perfect age reporting as well as no any large changes in fertility, mortality and migration for a long time would be 100.
The United Nations recommended a standard for measuring the age heaping as follows in table 1. The choice of 23 and 62 as the limits of age band to be examined in the classic Whipple's Index calculation is arbitrary but has been found most suitable for practical purpose of measuring age heaping in general in a population of all ages.

The choice of 23 and 62 as the limits of age band to be examined in the classic Whipple's Index calculation is arbitrary but has been found most suitable for practical purpose of measuring age heaping in general in a population of all ages.

The Myers' index was developed to detect preference for all terminal digits from 0 to 9. The method yields an index of reference for each terminal digit as well as a summary index of preference for terminal digits. The theoretical range of Myers' index from 0 to 90. An index of 0 represents no heaping and an index of 90 represents

a heaping of all reported ages at a single digit, say five. Secondary data from previous censuses (censuses that have been done in 1995 and 1985) are also employed to make necessary comparisons.

3 Results

Table 2 shows the degree of digit preference bias that was assessed using a modification of Myers' index for whole population in separation of Male and Female. It shows heaping at ages ending in zero and five years, for both males and females. The preference for these digits among males may be attributed to the greater tendency to overestimate the age, whilst for females; it may be due to underestimation of their age.

Table 3 gives the Myers' index for males and females, in both urban and rural population. It shows a higher index for females than males in urban population, indicating that age was more accurately reported among males than females. The pattern of heaping is pronounced from age 20 onwards and this is true for both urban and rural population. Table 4 shows the Whipple's index for whole population in separation of male and female. According to the standard for age measuring the Whipple's index for whole population shows that the quality of data is ok, this implies that the age reporting is good.

The table 5 gives the Whipple's index for urban and rural population. The figures from Whipple's index show that the reporting of age in urban is more accurate than rural population. Table 5 further shows that males have higher tendency of age heaping than females in rural areas where it has been observed reverse in urban areas. This could be due to the fact that men were often not available at the time of the census and therefore, female respondents had to report on behalf of men in the census. It is highly likely that the female respondents may not have correctly reported the age of males during the census. Myers' index for total population in 1995 Census was 2.645 which in 2005 Census were 3.2036; it implies that age reporting was better in Census 1995.

4 Discussion

Sweden is considered to be the country with the most reliable demographic data in the world. In country like India, the distortions are naturally greater because precise reporting or determination of age is not considered important in the everyday affairs and laxity is permitted on various grounds of discretion, expediency and convenience.

It can be deduced from the analysis that the quality of age reporting for the 2005 census data was poor as compared to the 1995 census data. However, it was of better quality than the 1985 and 1975 census data. This may suggest that both males and females tend to misreport their ages before age 60. Frequently, the elderly population either does not know its age at all or reports ages in big age bands such as 60-70, 70-80 etc. It is possible that often the enumerator is forced to estimate the age of a person based on physical appearance or hearsay in absence of any reliable documents or observance of socio-cultural norms which allow the individual or member of the household to know their ages precisely. There are other two groups for whom recording of age is rather difficult, women being one of them. Although, frequently, women may be in a position to recall when they were married or a child was born to them, it is difficult for them to state their own date/year of birth unless they are literate. In addition, assessing age by the enumerator may also be difficult for young women, as in certain sections of this population; they may not be permitted to appear during the enquiry, unless the enumerator is a lady. The other groups, which may suffer from these inaccuracies, are infants and children particularly those not attending school. However, the magnitude of digit preference bias seems to be reducing with the passage of time. This is especially true in the case of females. The possibility of increased female literacy as a factor underlying this reduction is pointed out. The absence of significant digit preference at ages divisible by five or ten, however, is not necessarily proof of data accuracy since other kinds of errors in age misreporting may also distort the data quality. One way

of addressing this issue is to examine the reported population at very old ages relative to the total elderly population. As shown by Coale and Kisker (1986), the proportion of those age 95 or over among people age 70 or over in 23 countries with accurate data was always less than six per thousand. This proportion in 28 countries with poor data ranged from one percent to 10 percent.

References

[1] Haaga, John G. (May 1988). Reliability of retrospective survey data on infant feeding. Demography, Vol. 25, No. 2307-14 pp. Alexandria, Virginia. In Eng.

[2] Dr. D. R. Choudhury, Deputy Registrar General (C&T) .Office of the Registrar General & Census Commissioner, India

[3] R. H. DAW, B.Sc., F.I.A. (1982). The improving accuracy of the basic data and the exposed-to-risk of the English life tables. Journal of the Institute of Actuaries [JIA] 109: 117-125

[4] Ibrahim Naseem, Bhakta Gubhaju and Hussain Niyaaz .Rapid Fertility Decline in the Mal dives: An Assessment, (Demographers' Notebook) Asia-Pacific Population Journal .Vol. 19, No. , p 57-75.

[5] Megan Beckett, Julie DaVanzo, Narayan Sastry, Constantijn Panis, Christine Peterson (September 1999) The Quality of Retrospective Reports in the Malaysian Family Life Survey ,RAND , 1700 Main Street, Santa Monica, California 90401

[6] Zeng Yi, James W. Vaupel (18 April 2003) Oldest-Old Mortality in China, DEMOGRAPHIC RESEARCH, VOLUME 8, ARTICLE 7, PAGES 215-244

[7] Sulaiman Bah ,(1999) Diagnostic tests on assessing the quality of maternal orphan hood data, Arminel Roundtable April 9th -11th, 1999

[8] Shryock, H., Siegel, J et al. (1976). The Methods and Materials of Demography. Academic Press, San Diego

[9] Ewbank, D. (1981) Age Miss reporting and Age-Selective Under enumeration: Sources, Patterns and Consequences for Demographic Analysis. National Academy Press, Washington DC

Table 1: A standard for measuring the age heaping recommended by United Nations

Whipple's Index	quality of data perfect	deviation from
<105	Very accurate	<5%
105-110	Relatively accurate	5-9.99%
110-125	OK	10-24.9%
125-175	Bad	25-74.99%
>175	Very bad	>= 75%

Table 2: Results of Myers' index for the age data of the population in the aggregate, Iranian census, 2005

Total	
3,203596	Total
3,064657	Male
3,34749759	Female

[10] United Nations (1995). Indirect Techniques for Demographic Estimation. United Nations, New York

[11] IRAN STATISTICAL YEAR BOOK 1383 (2005)

Table 3: Results of Myers' index in both urban and rural population, Iranian census, 2005

Rural	Urban	
3,8796872	2,88592	Total
3,77832143	2,7306349	Male
4,00833247	3,045577	Female

Table 4: Results of Whipple index for the age data of the population in the aggregate, Iranian census, 2005

Total Population		
Whipple's Index	Quality of Data	
111,585889	OK	Grand Total
110,444	OK	Male
112,75751	OK	Female

Table 5: Results of Whipple index for both urban and rural population, Iranian census, 2005

Rural		Urban		
Whipple's Index	Quality of Data	Whipple's Index	Quality of Data	
117,199803	OK	109,262793	Relatively accurate	Grand Total
115,22435	OK	108,504063	Relatively accurate	Male
119,16064	OK	105,8131	Relatively accurate	Female

Almost Sure Results for a linear function of the top k order statistics

Alireza Yousefi, and R. Vasudeva

Department of Statistics ,University of Mysore,Mysore,570006,India

Abstract: The difference between the average of the top k out of n order statistics and the population mean expressed in population standard deviation units is known as the selection differential.In this paper we obtain some almost sure results for selection differential in case of uniform distribution .Similar results are obtained for some continuous distributions using transformation .

Keywords: order statistics; selection differential; almost sure behavior; iterated logarithm law.

1 Introduction

Let X_1,X_2,...,X_n be a random sample from a population with continuous distribution function F,having mean μ and variance σ^2. Let $X_{1,n} \leq X_{2,n} \leq \ldots X_{n,n}$ denote the order statistics of these random observations.Then

$$V_{k,n} = \frac{\sum_{i=n-k+1}^{n} (X_{i,n} - \mu)}{k\sigma} \tag{1}$$

is called the selection differential .It represents the difference between the average of selected group and population mean expressed in standard deviation units which has long been a familiar term to genitists and breeders who often refer to it as "intensity of selection" .One of the measure of the effectiveness of selection differential is the high value it can reach ,in relation to the population mean .In knowing the effectiveness ,we obtain almost sure lower bounds for the selection differential.This is achieved by the iterated logarithm law. Nagaraja(1981),(1982) had obtained sharp bounds for expectation and

asymptotic distribution of selection differential and Suresh(1993) introduced percentile selection differential and has obtained some asymptotic results.

Throughout the paper ,except in the last section ,we assume that F is uniform(0,1).Hence we have $\mu = \frac{1}{2}$ and $\sigma = \frac{1}{\sqrt{12}}$. As such ,we discuss the behaviour of

$$D_{k,n} = \frac{\sum_{i=n-k+1}^{n} X_{i,n}}{k} \tag{2}$$

throughout the paper.

2 Almost sure results

Throughout this section ,with no loss of generality, we define selection differential for k fixed,$1 \leq k \leq n$, except for origin and scale shift as $D_{k,n} = \frac{X_{n-k+1}+\ldots+X_{n-1}+X_n}{k}$. We assume that (X_n) is defined over a probability space(Ω, F, P).

Theorem 1 *Let $X_1, X_2, \ldots, X_n$ be i.i.d random variables with common U(0,1) distribution function.Then*

$$\lim D_{k,n} = 1 \quad a.s \tag{3}$$

Lemma 1 *Let $X_{n-k+1,n}$ be the k^{th} maximum of a sample $X_1, X_2, \ldots, X_n$ of i.i.d random variables from U(0,1).Then*

$$X_{n-k+1,n} \to 1 \quad a.s \tag{4}$$

proof of lemma :

To show that $X_{n-k+1,n} \to 1 \quad a.s$, it is sufficient to show that for any ϵ ,$0 < \epsilon < 1$ $P(X_{n-k+1,n} < (1-\epsilon) \quad i.o) = 0$. We have

$$P(X_{n-k+1,n} < 1-\epsilon) = (1-\epsilon)^n + \binom{n}{1}(1-\epsilon)^{(n-1)}\epsilon + \ldots + \binom{n}{k-1}(1-\epsilon)^{(n-k+1)}\epsilon^{k-1}$$

Hence $\sum_{n=1}^{+\infty} P(X_{n-k+1,n} < 1-\epsilon) < \infty$. By Borel-Contelli lemma

$$P(X_{n-k+1,n} < (1-\epsilon) \quad i.o) = 0 \quad or \quad X_{n-k+1,n} \longrightarrow 1 \quad a.s$$

Remark 1 When k=1 ,$X_{n-k+1} = X_{n,n} = max(X_1, X_2, \ldots, x_n)$.By the above lemma we have $X_{n,n} \to 1 \quad a.s$ which is welknown .

Proof of the Theorem:

k=1 implies $D_{1,n} = max(X_1, X_2, \ldots, X_n) = X_{n,n}$ and from Remark 1 ,$X_{n,n} \to 1 \quad a.s$. For $k \geq 2$,note that $X_{n-k+1,n} \leq D_{k,n} \leq X_{n,n}$, by Lemma 1, the proof is immediate.

Theorem 2 *Let $X_1 \leq X_2 \leq \ldots \leq X_n$ be the order statistics of n observation from U(0,1) distribution function.Then*

$$\limsup \frac{n(1 - D_{k,n})}{\ln \ln n} = 1 \quad a.s \tag{5}$$

Lemma 2 *Under the setup Theorem 2*

$$\limsup \frac{n(1 - X_{n-k+1,n})}{\ln \ln n} = 1 \quad a.s \tag{6}$$

proof of lemma :

When k=1 ,the result follows from ,(see for example Galambos(1978),
pp. 223). With no loss of generality we prove the Lemma for k=2. Equivalantly, for any$\epsilon > 0$,we establish that

$$P(\frac{n(1 - X_{n-1,n})}{\ln \ln n} > (1 + \epsilon) \quad i.o) \;=\; 0 \tag{7}$$

$$P(\frac{n(1 - X_{n-1,n})}{\ln \ln n} > (1 - \epsilon) \quad i.o) \;=\; 1 \tag{8}$$

Define $b_n = (1 + \epsilon)\frac{\ln \ln n}{n} \quad and \quad A_n = \left(X_{n-1,n} < 1 - b_n\right)$ .Then
$P(A_n) = F^n(1 - b_n) + nF^{n-1}(1 - b_n)(1 - F(1 - b_n)) = (1 - b_n)^n + n(1 - b_n)^{n-1}(b_n)$
Note that

$$(1 - b_n)^n = \left(1 - \frac{(1 + \epsilon)\ln \ln n}{n}\right)^{\ln \ln n \frac{n}{\ln \ln n}} \;\simeq\; e^{(-(1+\epsilon)\ln \ln n)} \simeq \frac{1}{(\ln n)^{1+\epsilon}}$$

Hence

$$P(A_n) \simeq \frac{1}{(\ln n)^{1+\epsilon}} + \frac{(1+\epsilon)\ln\ln n}{(\ln n)^{1+\epsilon}} \tag{9}$$

Observe that $P(A_n) \to 0 \quad as \quad n \to \infty$.In order to establish (7) it is enough to show that (See , Barndorff-Nielsen(1961)). $\sum_n P(A_n \cap A_{n+1}^c) < \infty \quad or \quad P(A_n \cap A_{n+1}^c \quad i.o) = 0$ $b_n \downarrow implies$ that $d_n = 1 - b_n \uparrow as \quad n \to \infty$. Then $A_{n+1}^c = (X_{n,n+1} \geq d_{n+1})$. Note that $d_n \leq d_{n+1}$ implies that $(X_{n,n+1} > d_{n+1}) \subseteq (X_{n,n+1} > d_n)$. Hence

$$\begin{aligned}(A_n \cap A_{n+1}^c) &= (X_{n-1,n} < d_n, X_{n,n+1} > d_{n+1}) \\ &\subseteq (X_{n-1,n} < d_n, X_{n,n+1} \geq d_n) \\ &\subseteq (X_{n-1,n} < d_n, X_{n+1} \geq d_n)\end{aligned}$$

From (9) ,one can find a n_0 such that for $n \geq n_0$

$$\begin{aligned}P(A_n \cap A_{n+1}^c) &\leq P(X_{n-1,n} < d_n)P(X_{n+1} \geq d_n) \\ &\leq \frac{1}{(\ln n)^{1+\frac{\epsilon}{2}}} \frac{(1+\epsilon)\ln\ln n}{n} \leq \frac{1}{n(\ln n)^{1+\frac{\epsilon}{3}}}\end{aligned}$$

Now $\sum_n \frac{1}{n(\ln n)^{1+\frac{\epsilon}{3}}} < \infty$ implies that $\sum_n P(A_n \cap A_{n+1}^c) < \infty$ By Borel-Cantelli Lemma one gets

$$P(A_n \cap A_{n+1}^c \quad i.o) = 0$$

From the fact that $P(A_n) \to 0 \quad as \quad n \to \infty$,we conclude that $P(A_n i.o) = 0$

$$i.e. \quad P(X_{n-1,n} < 1 - \frac{(1+\epsilon)\ln\ln n}{n} \quad i.o) = 0$$

Similarly ,one can show that for any $k > 1$,

$$P(X_{n-k+1,n} < 1 - \frac{(1+\epsilon)\ln\ln n}{n} \quad i.o) = 0 \tag{10}$$

From the relation $X_{n-k+1,n} < X_{n,n}$ and $\limsup \frac{n(1-X_{n,n})}{\ln\ln n} = 1 \quad a.s$,(see for example Galambos(1978),pp. 223), one can show that

$$P(X_{n-k+1,n} < 1 - \frac{(1-\epsilon)\ln\ln n}{n} \quad i.o) = 1 \tag{11}$$

(10) and (11) complete the proof of the Lemma.

proof of the theorem:

It is enough to show that for any $\epsilon > 0$,

$$P(\frac{n(1-D_{k,n})}{\ln\ln n} > (1+\epsilon) \quad i.o) = 0 \tag{12}$$

$$P(\frac{n(1-D_{k,n})}{\ln\ln n} > (1-\epsilon) \quad i.o) = 1 \tag{13}$$

Observe that

$$X_{n-k+1,n} \leq D_{k,n} \leq X_{n,n} \Leftrightarrow \frac{1-X_{n,n}}{a_n}$$
$$\leq \frac{1-D_{k,n}}{a_n} \leq \frac{1-X_{n-k+1,n}}{a_n} \quad where \quad a_n = \frac{\ln\ln n}{n}$$

Hence (12) and (13) are immediate consequences of Lemma 2 . Consequently,the Theorem is proved.

3 Boundary crossing problems

Here we study the boundary crossing random variable related to the selection differentials . Define for any small $\epsilon > 0$

$$Z_n = \begin{cases} 1 \quad if \quad D_{k,n} < (1-\epsilon) \\ \\ 0 \quad Otherwise \end{cases}$$

and $N(\epsilon) = \sum_{n=k}^{\infty} Z_n$. Since $P(D_{k,n} < (1-\epsilon) \quad i.o) = 0$, $N(\epsilon)$ will be a proper random variable.

Theorem 3 *Under the setup of Theorem 1,*$E(N^{\lambda}(\epsilon)) < \infty$ *for all* $\lambda > 0$.

Proof of the Theorem:

Note that $X_{n-k+1,n} < (1-\epsilon)$ whenever $D_{k,n} < (1-\epsilon)$ and $D_{k,n} < (1-\epsilon)$ whenever $X_{n,n} < (1-\epsilon)$. Let

$$Z_n^{'} = \begin{cases} 1 \quad if \quad X_{n-k+1,n} < (1-\epsilon) \\ \\ 0 \quad Otherwise \end{cases} \qquad Z_n^{''} = \begin{cases} 1 \quad if \quad X_{n,n} < (1-\epsilon) \\ \\ 0 \quad Otherwise \end{cases}$$

and $N^{'}(\epsilon) = \sum_{n=k}^{\infty} Z_n^{'}$, $N^{''}(\epsilon) = \sum_{n=k}^{\infty} Z_n^{''}$ .Then $N^{''}(\epsilon) \leq N(\epsilon) \leq N^{'}(\epsilon)$ .Hence

$$E(N^{''}(\epsilon))^{\lambda} \leq E(N(\epsilon))^{\lambda} \leq E(N^{'}(\epsilon))^{\lambda} \quad , \lambda > 0 \tag{14}$$

Proceeding as in Slivka and Severo(1970) ,for $\lambda > 1$,one can show that

$$\begin{aligned} E(N^{'}(\epsilon))^{\lambda} &\leq \sum_{n=k}^{\infty} n^{\lambda-1} P(Z_n^{'} = 1) = \sum_{n=k}^{\infty} n^{\lambda-1} \sum_{r=0}^{k-1} \binom{n}{r} (1-\epsilon)^{n-r} \epsilon^r \\ &= \sum_{r=0}^{k-1} \sum_{n=k}^{\infty} n^{\lambda-1} \binom{n}{r} (1-\epsilon)^{n-r} \epsilon^r < \infty \end{aligned}$$

From (14) all moment of N(ϵ) random variable are finite .Also

$$E(N^{'}(\epsilon)) = \sum P(Z_n^{'} = 1) = \sum_{n=k}^{\infty} \sum_{r=0}^{k-1} \binom{n}{r} (1-\epsilon)^{n-r} \epsilon^r < \infty \; or \; E((N(\epsilon)) < \infty$$

It can be show that, $E(N^{''}(\epsilon)) = \sum_{n=1}^{\infty} (1-\epsilon)^n = \frac{1-\epsilon}{\epsilon}$.
For k=2, $E(N^{'}(\epsilon)) = \sum_{n=2}^{\infty} (1-\epsilon)^n + \sum_{n=2}^{\infty} n(1-\epsilon)^{n-1}\epsilon \quad = \frac{2(1-\epsilon)}{\epsilon}$.
Similarly for k=3

$$E(N^{'}(\epsilon)) = \sum_{n=3}^{\infty} (1-\epsilon)^n + \sum_{n=3}^{\infty} n(1-\epsilon)^{n-1}\epsilon + \sum_{n=3}^{\infty} \frac{n(n-1)}{2} (1-\epsilon)^{n-2}\epsilon^2 = \frac{3(1-\epsilon)}{\epsilon}$$

In general for fixed k $E(N^{'}(\epsilon)) = \frac{k(1-\epsilon)}{\epsilon}$. Using $E(N^{''}(\epsilon)) \leq E(N(\epsilon)) \leq E(N^{'}(\epsilon))$,one can give the following bound for $E(N(\epsilon))$

$$i.e. \quad \frac{1-\epsilon}{\epsilon} \leq E(N(\epsilon)) \leq \frac{k(1-\epsilon)}{\epsilon}$$

Theorem 4 *Under the setup of Theorem 2,$E(N^{\lambda}(\epsilon))$ divergent for all $\lambda > 0$.*

Proof of the Theorem:

Under the assumption of Theorem 2 for any small $\epsilon > 0$ and fixed $k < n$, define

$$\Lambda_n = \begin{cases} 1 & if\ D_{k,n} < 1 - \frac{(1+\epsilon)\ln\ln n}{n} \\ 0 & Otherwise \end{cases} \quad \Lambda_n^{'} = \begin{cases} 1 & if\ X_{n,n} \leq 1 - \frac{(1+\epsilon)\ln\ln n}{n} \\ 0 & Otherwise \end{cases}$$

where $N_1(\epsilon) = \sum_{n=k}^{\infty} \Lambda_n$, $N_1^{'}(\epsilon) = \sum_{n=k}^{\infty} \Lambda_n^{'}$ and $D_{k,n} \leq X_{n.n} \Rightarrow N_1(\epsilon) \geq N_1^{'}(\epsilon) \Rightarrow E(N_1(\epsilon))^{\lambda} \geq E(N_1^{'}(\epsilon))^{\lambda}$. Proceeding as in Slivka(1969), one can show that for $0 < \lambda < 1$

$$E(N_1^{'}(\epsilon))^{\lambda} = \sum_{n=k}^{\infty} n^{\lambda-1} P(\Lambda^{'}(n) = 1)$$

We have $P(\Lambda^{'}(n) = 1) = P\Big(X_{n,n} \leq 1 - \frac{(1+\epsilon)\ln\ln n}{n}\Big) = \Big(1 - \frac{(1+\epsilon)\ln\ln n}{n}\Big)^n$, One can find an integer n_0 such that for all $n \geq n_0$

$$\Big(1 - \frac{(1+\epsilon)\ln\ln n}{n}\Big)^n \geq e^{-(1+\frac{3\epsilon}{2})\ln\ln n} = \frac{1}{(\ln n)^{1+\frac{3\epsilon}{2}}}$$

Consequently ,$\sum_{n=k}^{\infty} n^{\lambda-1} P(\Lambda^{'}(n) = 1) = \infty \quad , Since, \quad \sum \frac{1}{n^{1-\lambda}} \frac{1}{(\ln n)^{1+\frac{3\epsilon}{2}}} = \infty$

$E(N_1^{'}(\epsilon))^{\lambda} = \infty$ and in turn ,$E(N_1(\epsilon))^{\lambda} = \infty \quad for \quad 0 < \lambda < 1$. i.e, $E(N_1(\epsilon))^{\lambda} = \infty$ for all $\lambda > 0$.

4 Results for other distributions

Let (X_n) be i.i.d random variables with continuous distribution function F bounded to the right.Let $\alpha > 0$ be such that ,$F(\alpha) = 1$.Define

Y= F(X), then F(X) is U(0,1). Suppose that $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ is the order statistics of $X_1, X_2, \ldots, X_n$.Then $Y_{j,n} = F(X_{j,n}), j = 1, 2, \ldots, n$ implies $Y_{1,n} \leq Y_{2,n} \leq \ldots \leq Y_{n,n}$ is an order statistics for $Y_1, Y_2, \ldots, Y_n$ where $Y_j = F(X_j), j = 1, 2, \ldots, n$. Let,

$$D^*_{k,n} = \frac{X_{n-k+1,n} + X_{n-k+2,n} + \ldots + X_{n,n}}{k},$$
$$D_{k,n} = \frac{Y_{n-k+1,n} + Y_{n-k+2,n} + \ldots + Y_{n,n}}{k}$$

Clearly $X_{n-k+1,n} \leq D^*_{k,n} \leq \alpha, n \geq k$ and from Lemma 1 ,it's known that

$$\begin{aligned} & Y_{n-k+1,n} \longrightarrow 1 \quad a.s \quad or \quad F(X_{n-k+1,n}) \longrightarrow 1 \quad a.s \\ \Longrightarrow \quad & X_{n-k+1,n} \longrightarrow F^{-1}(1) \quad a.s \quad or \quad X_{n-k+1,n} \\ & \longrightarrow \alpha \quad a.s \Rightarrow D^*_{k,n} \longrightarrow \alpha \quad a.s \end{aligned}$$

We now give some examples as application of Theorem2.

Example 1 *Consider $X_1, X_2, \ldots, X_n$ i.i.d Normal(0,1) and $X_{1,n} < X_{2,n} < \ldots < X_{n,n}$ be the corresponding order statistics.Then*

$\liminf D^*_{k,n} e^{\frac{D^{*2}_{k,n}}{2}} \frac{\ln \ln n}{n} = \frac{1}{\sqrt{2\Pi}} \quad a.s$

Define $Y_{i,n} = F(X_{i,n})$,i=1,2,. . . ,n,where $Y_{1,n} < Y_{2,n} < \ldots < Y_{n,n}$ are order statistics from U(o,1). Hence from Lemma 2

$$\limsup \frac{n(1 - Y_{n-k+1,n})}{\ln \ln n} = 1 \quad a.s$$
$$\Rightarrow P\Big(1 - Y_{n-k+1,n} > (1 \pm \epsilon)\frac{\ln \ln n}{n} \quad i.o\Big) = \begin{matrix} 0 \\ 1 \end{matrix}$$

$$1 - Y_{n-k+1,n} > \frac{(1 \pm \epsilon) \ln \ln n}{n} \Longleftrightarrow 1 - F(X_{n-k+1,n}) > \frac{(1 \pm \epsilon) \ln \ln n}{n}$$

Since $X_{n-k+1} \to \infty \quad a.s$, from Feller(1968),pp. 175,one has

$$1 - F(X_{n-k+1,n}) \simeq \frac{1}{\sqrt{2\pi} X_{n-k+1,n}} e^{-\frac{X^2_{n-k+1,n}}{2}} \quad a.s$$

Consequently from (15),

$$P\left(X_{n-k+1,n}e^{\frac{X^2_{n-k+1,n}}{2}}(\frac{\ln\ln n}{n}) < \frac{1}{\sqrt{2\pi}(1+\epsilon)} \quad i.o\right) = 0$$

$$P\left(X_{n-k+1,n}e^{\frac{X^2_{n-k+1,n}}{2}}(\frac{\ln\ln n}{n}) < \frac{1}{\sqrt{2\pi}(1-\epsilon)} \quad i.o\right) = 1$$

Take $\frac{1}{(1\mp\epsilon)} = 1 \pm \delta$. *Then*

$$P\left(X_{n-k+1,n}e^{\frac{X^2_{n-k+1,n}}{2}}(\frac{\ln\ln n}{n}) < \frac{(1-\delta)}{\sqrt{2\pi}} \quad i.o\right) = 0$$

$$P\left(X_{n-k+1,n}e^{\frac{X^2_{n-k+1,n}}{2}}(\frac{\ln\ln n}{n}) < \frac{(1+\delta)}{\sqrt{2\pi}} \quad i.o\right) = 1$$

$$or \qquad \liminf\left(X_{n-k+1,n}e^{\frac{X^2_{n-k+1,n}}{2}}(\frac{\ln\ln n}{n})\right) = \frac{1}{\sqrt{2\pi}} \quad a.s$$

From the relation $X_{n-k+1,n} < D^*_{k,n} < X_{n,n}$ *one gets,* $\liminf\left(D^*_{k,n}e^{\frac{D^{*2}_{k,n}}{2}}(\frac{\ln\ln n}{n})\right) = \frac{1}{\sqrt{2\pi}}$ *a.s. Since F is N(o,1),we have* $V^*_{k,n} = D^*_{k,n}$. *This relation implies that for any* $\delta > 0$, $V^*_{k,n}e^{\frac{V^{*2}_{k,n}}{2}} = \frac{1-\delta}{\sqrt{2\pi}}\frac{n}{\ln\ln n}$ *a.s, or equivalently* $V^{*2}_{k,n} + 2\ln V^*_{k,n} > 2\ln n - 2\ln\ln n - 2\ln\sqrt{2\pi} - \epsilon$ *a.s . Define* $\Theta_n = 2\ln n - 2\ln\ln n - 2\ln\sqrt{2\pi}$ *.From the fact that* $x^2 + 2\ln x$ *is an increasing function in x,the lower bound for* $V^*_{k,n}$ *is obtained from the equation* $V^{*2}_{k,n} + 2\ln V^*_{k,n} - \Theta_n = 0$. *The plot of the lower bound is given in the Figure-2.*

Example 2 $X_1, X_2, \ldots, X_n$ *i.i.d standard exponential with the corresponding order statistics of* $X_{1,n} < X_{2,n} < \ldots < x_{n,n}$. *One can show that for fixed k ,*

$$\liminf\left(D^*_{k,n} - (\ln n - \ln\ln\ln n)\right) = 0 \quad a.s$$

Define $Y_{k,n} = F(X_{k,n})$ *,i=1,2,. . . ,n,where* $Y_{1,n} < Y_{2,n} < \ldots < Y_{n,n}$ *are order statistics from uniform(o,1).Hence from Lemma 2,one has*

$$P\Big(1 - Y_{n-k+1,n} > (1 \pm \epsilon)\frac{\ln\ln n}{n} \quad i.o\Big) = \begin{matrix} 0 \\ 1 \end{matrix} \tag{15}$$

Note that $1 - F(X_{n-k+1,n}) = e^{-X_{n-k+1}}$*, and from (15)*

$$P\Big(e^{-X_{n-k+1,n}} > (1 \pm \epsilon)\frac{\ln\ln n}{n} \quad i.o\Big) = \begin{matrix} 0 \\ 1 \end{matrix} \quad or$$

$$P\Big(X_{n-k+1,n} < \ln - \ln\ln\ln n - \ln(1 \pm \epsilon) \quad i.o\Big) = \begin{matrix} 0 \\ 1 \end{matrix}$$

Taking $\ln(1+\epsilon) = \delta_1$ *and* $\ln(1-\epsilon) = -\delta_2$*, for* $\delta = max(\delta_1, \delta_2)$ *one gets*

$$P\Big(X_{n-k+1,n} - (\ln - \ln\ln\ln n) < \begin{matrix} - & \delta \\ & \delta \end{matrix} \quad i.o\Big) = \begin{matrix} 0 \\ 1 \end{matrix} \Rightarrow$$

$$\liminf\Big(X_{n-k+1,n} - (\ln - \ln\ln\ln n)\Big) = 0 \quad a.s$$

The relation $X_{n-k+1,n} < D^*_{k,n} < X_{n,n}$ *implies,*

$$\liminf\left(D^*_{k,n} - (\ln n - \ln\ln\ln n)\right) = 0 \quad a.s \tag{16}$$

For unit exponential ,we have $\mu = 1$ *and* $\sigma = 1$*.Hence* $V^*_{k,n} = D^*_{k,n} - 1$*.Using (16) ,one gets* $D^*_{k,n} > \ln n - \ln\ln\ln n - \delta$ $a.s$ *. Taking* $\gamma_n = \ln n - \ln\ln\ln n - \delta$*.Then*
$V^*_{k,n} = D^*_{k,n} - 1 > \gamma_n - 1$ $a.s$ *. By choosing n large,one can get a* $V^*_{k,n}$ *adequately large.*

Example 3 $X_1, X_2, \ldots, X_n$ *be i.i.d with pareto distribution function,*

$$F(x) = \begin{cases} 0 & if \quad x < 1 \\ 1 - \frac{1}{x^\alpha} & if \quad x \geq 1 \end{cases} \quad , \alpha > 0$$

One can show that , $\liminf \frac{D^*_{k,n}(\ln\ln n)^{\frac{1}{\alpha}}}{n^{\frac{1}{\alpha}}} = 1 \quad a.s \quad$ *. Define* $Y_{k,n} = F(X_{i,n})$ *,i=1,2,. . . ,n,where* $Y_{1,n} < Y_{2,n} < \ldots < Y_{n,n}$ *are order statistics from uniform(o,1). Hence from Lemma 2*

$$P\Big(1 - Y_{n-k+1,n} > (1 \pm \epsilon)\frac{\ln\ln n}{n} \quad i.o\Big) = \begin{matrix} 0 \\ 1 \end{matrix} \tag{17}$$

We have $1 - F(X_{n-k+1,n}) = \frac{1}{X^{\alpha}_{n-k+1,n}}$ *and hence,*

$$\begin{aligned} 1 - Y_{n-k+1,n} > \frac{(1 \pm \epsilon)\ln\ln n}{n} &\iff 1 - F(X_{n-k+1,n}) > \frac{(1 \pm \epsilon)\ln\ln n}{n} \\ &\iff X^{\alpha}_{n-k+1,n} < \frac{n}{(1 \pm \epsilon)\ln\ln n} \\ &\quad , From(17) \end{aligned}$$

$$P\left(\frac{X_{n-k+1,n}(\ln\ln n)^{\frac{1}{\alpha}}}{n^{\frac{1}{\alpha}}} < \frac{1}{(1\pm\epsilon)} \quad i.o\right) = \begin{matrix} 0 \\ 1 \end{matrix} \Rightarrow P\left(\frac{X_{n-k+1,n}(\ln\ln n)^{\frac{1}{\alpha}}}{n^{\frac{1}{\alpha}}} < (1 \mp \delta) \quad i.o\right) = \begin{matrix} 0 \\ 1 \end{matrix}$$

for $\delta > 0, i.e, \liminf \left(\frac{X_{n-k+1,n}(\ln\ln n)^{\frac{1}{\alpha}}}{n^{\frac{1}{\alpha}}}\right) = 1 \quad a.s,\ X_{n-k+1,n} < D^*_{k,n} < X_{n,n}$ *implies that*

$$\liminf \left(\frac{D^*_{k,n}(\ln\ln n)^{\frac{1}{\alpha}}}{n^{\frac{1}{\alpha}}}\right) = 1 \quad a.s \tag{18}$$

For Pareto Distribution $\mu = \frac{\alpha}{\alpha-1}$ *and* $\sigma = \sqrt{\frac{\alpha}{(\alpha-1)^2(\alpha-2)}}$ *, when ever* $\alpha > 2$*. Then* $V^*_{k,n} = \frac{D^*_{k,n}-\mu}{\sigma}$ *. Using (18),one gets* $D^*_{k,n} > (1-\delta)\frac{n^{\frac{1}{\alpha}}}{(\ln\ln n)^{\frac{1}{\alpha}}}\ a.s$ *. Taking* $\lambda_n = (1-\delta)\frac{n^{\frac{1}{\alpha}}}{(\ln\ln n)^{\frac{1}{\alpha}}}$*, one gets* $\frac{D^*_{k,n}-\mu}{\sigma} < \frac{\lambda_n-\mu}{\sigma}$ *f.o.i.e,* $V^*_{k,n} > \frac{\lambda_n-\mu}{\sigma}$ *a.s By choosing n large ,one can get* $V^*_{k,n}$ *sufficiently large.*

Remark 2 When $\alpha < 2$,we take $D^*_{k,n} - m$ as selection differential , where m is the median which is equal to $2^{\frac{1}{\alpha}}$.

Remark 3 One can show that selection differential for Freshet Distribution behaves similar to Pareto Distribution

We plot the lower bound of selection differentials for various value of n for Normal,Exponential and Pareto distributions in Figures 1 and 2. Figure-1 shows lower bound of selection differential of pareto dis-

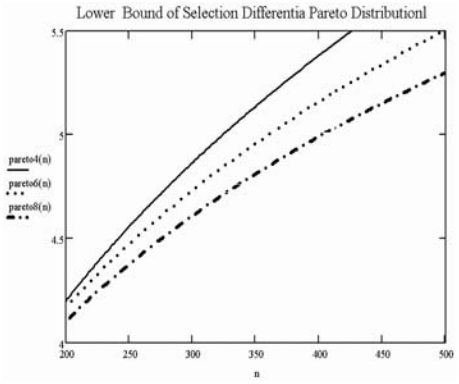


Figure 1: Pareto Distribution

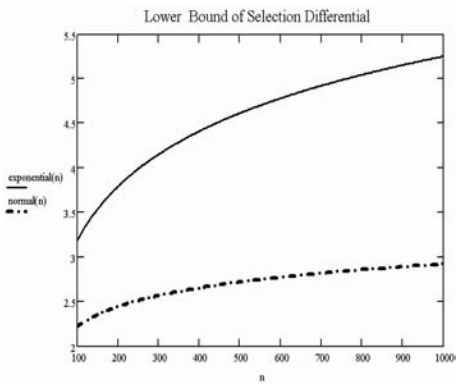


Figure 2: Normal and Exponential Distribution

tribution ,where α taken as 4,6,8 and Figure-2 represent lower bound of selection differential for standard normal distribution and unit exponential distribution .

References

[1] Barndorff-Nielsen, O.(1961).On the Rate of Growth of the Partial Maxima of a Sequence of Independent Identically Distributed Random Variables.*Math. Scand.*,9,383-394.

[2] Feller, W.(1968) .An Introduction to Probability Theory and Its Applications.*John Wiley & Sons*,1st ed.

[3] Galambos, J.(1978).The Asymptotic Theory of Extreme Order Statistics . *Wiley & Sons*,1st ed.

[4] Nagaraja,H.N. (1981). Some Finite Sample Results for The Selection Differential. *Ann. Inst. Statist. Math.*, 33,437-448.

[5] Nagaraja,H.N. (1982). Some Nondegenerate Limit Laws for The Selection Differential. *The Annals of Statistics*,Vol. 10,No. 4,1306-1310.

[6] Slivka,J. (1969).On the Law of the Iterated Logarithm.*Proc. of Nat. Aca. of Sci. of U.S.A* , Vol. 63, No. 2 , 289-291.

[7] Slivka ,J. and Severo,N.C.(1970).On the Strong Law of Large.*Proc. Amer. Math. Soci.*, Vol. 24, No. 4. 1970,729-734.

[8] Suresh,R.P.(1993).Some Asymptotic Results for The Induced Percentile Selection Differential.*The Indian Journal of Statistics(Sankhya)*,Vol. 55 ,A,120-129.

Central Limit Theorem For Linearly Negative Dependent Fuzzy Random Variables

R .Zaman, and B. Sadeghpour Gildeh

Payame Noor University of Ahvaz
Department of Statistics, University of Mazandaran

Abstract: In this paper, central limit theorems for linearly negative dependent fuzzy random variables is discussed by invoking the Hakuhara metric.

Keywords: Central limit theorem; Fuzzy numbers; Fuzzy set-valued random variables; Negatively dependent; Linearly negative dependent.

1 Introduction

The concept of fuzzy random variable was introduced by Kwakernaak[3] and Puri and Ralescu[4]. H-C.Wu[9], provided central limit theorem (CLT) for $\alpha-$cut fuzzyrandom variables. Newman (1984) used liearly negative dependent random variables which requires linear combinations (with positive constant) of negative dependent random variables to be negatively dependent. Roussas[8], assumed negative association (which impies linearly negative dependence) to obtain CLT for arrays of negative dependent and negative associated random variables. Terán[7], provided CLT when measurements are modelled as trapezoidal fuzzy intevals and the t-norm is either the produc or the Łukasiewicz t-norm. In this paper, we will be obtained CLT for linearly negative dependent fuzzy random variables. We suppose that, the fuzzy-random variables have the following form $\widetilde{X} = \widetilde{\mu} \oplus \widetilde{I}_{\{E\}}$, where E a real random variable is distributed as $\widetilde{X}$ but with zero mean[7].Therefore we use the $D_{p,q}-$Distance defined on set of fuzzy numbers, to conculate the variance of fuzzy random variable[6],

$$\begin{aligned} E(\widetilde{X}) &= \widetilde{\mu} \\ Dvar(\widetilde{X}) &= var(E). \end{aligned}$$

In section 2, we recall some basic concepts of fuzzy numbers.

In section 3, the definition linearly negative dependent fuzzy random variables have been presented, and finally, in section 4 we prove central limit theorems by standardizing linearly negative dependent fuzzy random variables $\widetilde{X}$ by invoking the Hakuhara metric.

2 Preliminary

In this section we first recall some notions of fuzzy sets, fuzzy numbers and some operations on fuzzy numbers. And then we present $D_{p,q}$-distance defined on the space of fuzzy numbers [5].

Definition 1 ([6]) Let E be a universal set, then a fuzzy set $\widetilde{A}$ of E is defined by its membership function $\widetilde{A} : E \to [0,1]$. For all $x \in E$, $\widetilde{A}(x)$ is the membership grade of x to $\widetilde{A}$.

Definition 2 ([6]) A_α is called the $\alpha - level(cut)$ set of $\widetilde{A}$, defined by $A_\alpha = \{x \in E : \widetilde{A}(x) \geq \alpha\}$.

Definition 3 ([6]) A fuzzy number is a fuzzy set of $\mathbb{R}$ such that the following condition are satisfied
a) $\widetilde{A}$ is normal, that is, there exist x_0 such that $\widetilde{A}(x_0) = 1$,

b) $\widetilde{A}$ is convex, that is, $\forall x_1, x_2 \in \mathbb{R}$ and $\lambda \in [0,1]$:

$$\widetilde{A}(\lambda x_1 + (1-\lambda)x_2) \geq min(\widetilde{A}(x_1), \widetilde{A}(x_2)),$$

c) $\widetilde{A}$ is upper semicontinuous with compact support.
According to the definition 2.2, $\alpha - level$ set of a fuzzy number is a closed interval, denote by $A_\alpha = [A_\alpha^-, A_\alpha^+]$, i.e.

$$A_\alpha^- = \inf\{x \in \mathbb{R} : \widetilde{A}(x) \geq \alpha\}, \text{ and } A_\alpha^+ = \sup\{x \in \mathbb{R} : \widetilde{A}(x) \geq \alpha\}.$$

The set of all fuzzy number is denoted by $F(\mathbb{R})$. Let $\widetilde{a}$ and $\widetilde{b}$ be two closed fuzzy numbers. If there exists a closed fuzzy number $\widetilde{c}$ such that $\widetilde{c} \oplus \widetilde{b} = \widetilde{a}$ (note that the fuzzy addition is commutative), then $\widetilde{c}$ is unique. In this case, $\widetilde{c}$ is called the Hukuhara difference of $\widetilde{a}$ and $\widetilde{b}$ and is denoted by $\widetilde{a} \ominus_H \widetilde{b}$ *(see Puri and Ralescu[4])*

Definition 4 ([6]) The $D_{p,q}$-distance, indexed by parameters $1 \leq p \leq \infty, 0 \leq q \leq 1$, between two fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ is a nonnegative function on $F(\mathbb{R}) \times F(\mathbb{R})$ given as follows

$$D_{p,q}(\widetilde{A}, \widetilde{B}) = \begin{cases} [(1-q)\int_0^1 |A_\alpha^- - B_\alpha^-|^p \, d\alpha + q\int_0^1 |A_\alpha^+ - B_\alpha^+|^p \, d\alpha]^{\frac{1}{p}}, \\ \qquad\qquad \text{if} \quad p < \infty, \\ (1-q)\text{sup}_{0<\alpha\leq 1}\left(|A_\alpha^- - B_\alpha^-|\right) + q\text{inf}_{0<\alpha\leq 1}(|A_\alpha^+ - B_\alpha^+|) \\ \qquad\qquad \text{if} \quad p = \infty. \end{cases},$$

The analytical properties of $D_{p,q}$ depend on the first parameter p, while the second parameter q of $D_{p,q}$ characterizes the subjective weight attributed to the sides of the fuzzy numbers. if there are no reason to distinguish any side of fuzzy numbers, $D_{p,\frac{1}{2}}$ is recommended. $(F(\mathbb{R}), D_{p,q})$ is a complete metrice space. Let $(\Omega, \mathcal{A}, P)$ be a probability space.

Definition 5 ([5]) A Mapping $\widetilde{X} : \Omega \to F(\mathbb{R})$ is said to be a fuzzy random variable associated with $(\Omega, \mathcal{A})$ if and only if

$$\{(\omega, x) : x \in X_\alpha(\omega)\} \in \mathcal{A} \times \mathcal{B},$$

where $\mathcal{B}$ denote the σ-field of Borel set in $\mathbb{R}$.

Definition 6 Let $\widetilde{X}$ be a fuzzy random variable with fuzzy expectation $\widetilde{\mu}$. We say that $\widetilde{X}$ is an H-fuzzy random variable if the Hukuhara diference $\widetilde{X}(\omega) \ominus_H \widetilde{\mu}$ exists for all $\omega \in \Omega$.

Proposition 2.1 Let $\widetilde{X}$ be an H-fuzzy random variable with fuzzy expectation $\widetilde{\mu}$.Then the fuzzy expectation of $\widetilde{X} \ominus_H \widetilde{\mu}$ is a crisp number with value 0.

Definition 7 Let E be a real normal standard random variable and $\widetilde{\mu} = \widetilde{0}$. We say that $\widetilde{X}$ is a normal standard fuzzy random variable.

Definition 8 ([6]) Let $\widetilde{X}$ be a fuzzy random variable and $\widetilde{A} \in F(\mathbb{R})$. The $D_{2,q}-$*mean square dispersion* of $\widetilde{X}$ about $\widetilde{A} \in F(\mathbb{R})$, is given by

$$DMSD(\widetilde{X}, \widetilde{A}) = E\left([D_{2,q}(\widetilde{X}, \widetilde{A})]^2\right) = \int_{\Omega} [D_{2,q}(\widetilde{X}(\omega), \widetilde{A})]^2 dP(\omega).$$

Definition 9 ([6]) The central $D_{2,q}-$*mean square dispersion* of $\widetilde{X}$ is called $Dvar(\widetilde{X})$ and given by

$$\begin{aligned} Dvar(\widetilde{X}) &= E\left([D_{2,q}(\widetilde{X}, \widetilde{\mu}_{\widetilde{X}})]^2\right) = \int_{\Omega} \left[(1-q)\int_0^1 (X_\alpha^-(\omega) - (\mu_{\widetilde{X}})_\alpha^-)^2 d\alpha \right. \\ &\quad \left. + q\int_0^1 (X_\alpha^+(\omega) - (\mu_{\widetilde{X}})_\alpha^+)^2 d\alpha \right] dP(\omega). \end{aligned} \tag{2}$$

Theorem 1 ([6]) Let $\widetilde{X}$ be a fuzzy random variable and $\widetilde{A} \in F(\mathbb{R})$, then

$$Dvar(\widetilde{X} \oplus \widetilde{A}) = Dvar(\widetilde{X}).$$

Proof:

$$\begin{aligned} (\widetilde{X}(\omega) \oplus \widetilde{A})_\alpha &= [X_\alpha^-(\omega) + A_\alpha^-, X_\alpha^+(\omega) + A_\alpha^+], \\ \left(\widetilde{E}(\widetilde{X} \oplus \widetilde{A})\right)_\alpha &= [(\mu_{\widetilde{X}})_\alpha^- + A_\alpha^-, (\mu_{\widetilde{X}})_\alpha^+ + A_\alpha^+]. \end{aligned}$$

Then, by using the definition of $Dvar(\widetilde{X}), Dvar(\widetilde{X} \oplus \widetilde{A}) = Dvar(\widetilde{X})$.

3 Linearly negative dependent fuzzy random variables

In this section we recall the definition of negatively dependent real-valued random variables, and then we present a new definition of negatively dependent fuzzy random variables based on $\alpha - level$ set .

Definition 10 ([2])Random variables X and Y are negatively dependent (ND) if

$$P\{X \leq x; Y \leq y\} \leq P\{X \leq x\}P\{Y \leq y\}$$

for all $x, y \in R$.

Definition 11 Fuzzy random variables $\widetilde{X}$ and $\widetilde{Y}$ are negatively dependent (ND) if for all $\alpha \in [0,1]$

$$\begin{aligned} &P\left\{X_\alpha^- \leq x, Y_\alpha^- \leq y\right\} \leq P\{X_\alpha^- \leq x\}P\{Y_\alpha^- \leq y\}, \\ &P\left\{X_\alpha^+ \leq x, Y_\alpha^+ \leq y\right\} \leq P\{X_\alpha^+ \leq x\}P\{Y_\alpha^+ \leq y\}, \end{aligned} \tag{1}$$

where $X_\alpha^- = inf\ \{x \in \mathbb{R} : \widetilde{X}_\alpha(\omega) \geq \alpha\}$ *and* $X_\alpha^+ = sup\ \{x \in \mathbb{R} : \widetilde{X}_\alpha(\omega) \geq \alpha\}$. X_α^- *and* X_α^+ *are real-valued random variables.*

Definition 12 A sequence of random variables, $\{X_j\}$ is said to be linearly negative dependent (LIND) if for any disjoint subsets of indices A,B and positive λ_js, $\sum_{k\in A} \lambda_k X_k$ and $\sum_{l\in B} \lambda_l X_l$ are negatively dependent.

Proposition 3.1(Newman, 1980) Suppose $X_1, ..., X_m$ are linearly negatively dependent (LIND). Then

$$\left| E\exp[i\sum_{j=1}^{m} r_j X_j] - \prod_{j=1}^{m} E\exp[r_j X_j] \right| \leq \sum_{\substack{k,l=1 \\ k<l}}^{m} |r_k r_l cov(X_k, X_l)|$$

Definition 13 A sequence of fuzzy random variables,$\{\widetilde{X}_j\}$ is said to be LIND if for any disjoint subsets of indices A,B and positive $\lambda_j^{,}$s

$$\sum_{k\in A}\lambda_k(X_k)_\alpha^+ \ and \ \sum_{l\in B}\lambda_l(X_l)_\alpha^+,$$
$$\sum_{k\in A}\lambda_k(X_k)_\alpha^- \ and \ \sum_{l\in B}\lambda_l(X_l)_\alpha^-$$

for all $\alpha\in[0,1,$ *are negativly dependent.*

4 Central limit theorems

Theorem 2 Let $\{\widetilde{X}_i\}$ be a sequence of H-fuzzy random variables which are independent.Then $D_n = \frac{\bigoplus_{i=1}^n \widetilde{X}_i \ominus_H \bigoplus_{i=1}^n \widetilde{\mu}_i}{\sqrt{Dvar(\bigoplus_{i=1}^n \widetilde{X}_i)}}$ converge in distribution to a N(0,1) f.r.v.

Proof: Based on the definition of $\widetilde{X}$,we have

$$\begin{aligned} D_n &= \frac{\bigoplus_{i=1}^n \widetilde{X}_i \ominus_H \bigoplus_{i=1}^n \widetilde{\mu}_i}{\sqrt{Dvar(\bigoplus_{i=1}^n \widetilde{X}_i)}} = \frac{\left(\bigoplus_{i=1}^n \widetilde{\mu}_i \oplus \widetilde{I}_{\{\sum_{i=1}^n E_i\}}\right) \ominus_H \bigoplus_{i=1}^n \widetilde{\mu}_i}{\sqrt{Dvar(\bigoplus_{i=1}^n \widetilde{X}_i)}} \\ &= \frac{\widetilde{0} \oplus \widetilde{I}_{\{\sum_{i=1}^n E_i\}}}{\sqrt{var(\sum_{i=1}^n \widetilde{E}_i)}} = \widetilde{0} + \frac{\sum_{i=1}^n E_i}{\sqrt{var(\sum_{i=1}^n \widetilde{E}_i)}} \end{aligned}$$

By CLT for real random variables, we can say that $\frac{\sum_{i=1}^n E_i}{\sqrt{var(\sum_{i=1}^n E_i)}}$ converge in distribution to a N(0,1) r.v. therefore, $D_n \xrightarrow{d} N(0,1) f.r.v.$

Theorem 3 Let $\{\widetilde{X}_i\}$ be a sequence of LIND fuzzy random variables, such that

$$1) \quad \sigma_n^{'2} = Dvar(\bigoplus_{i=1}^{n} \widetilde{X}_i) \to \infty$$

$$2) \quad \sigma_n^{'2} \sum\sum_{i<j} cov(E_i, E_j) \to 0 \quad as \quad n \to \infty$$

$$3) \quad \sum_{i=1}^{n} E(E_i^2 I_{\{|E_i|>\epsilon\sigma_n'\}}) = O(\sigma_n^{'2}).$$

Then $\sigma_n^{'-1} \sum_{i=1}^{n} E_i$ *converge in distribution to a N(0,1) r.v. so that, by considering* $\widetilde{\mu}_i = \widetilde{0}$ *we have* $\sigma_n^{'-1} \bigoplus_{i=1}^{n} \widetilde{X}_i \xrightarrow{d} N(0,1) f.r.v.$

Proof:

$$\sigma_n^2 = \sum_{i=1}^{n} Dvar(\widetilde{X}_i) = \sum_{i=1}^{n} var(E_i) \geq Dvar(\bigoplus_{i=1}^{n} \widetilde{X}_i) = var(\sum_{i=1}^{n} E_i) = \sigma_n^{'2}$$

since

$$var(\sum_i E_i) = \sum_i var(E_i) + \sum\sum_{i<j} cov(E_i, E_j)$$

and

$$1 = \lim_{n\to\infty} (\frac{\sigma_n^2}{\sigma_n^{'2}} + \frac{1}{\sigma_n^{'2}} \sum\sum_{i<j} cov(E_i, E_j)) = \lim_{n\to\infty} \frac{\sigma_n^2}{\sigma_n^{'2}}$$

it suffices to show that $\sigma_n^{-1} \sum_{i=1}^{n} E_i \xrightarrow{d} N(0,1)$.

By proposition 3.1 we have,

$$\left| Ee^{it\sigma_n^{-1}\sum_{j=1}^{n} E_j} - \prod_{j=1}^{n} Ee^{it\sigma_n^{-1}E_j} \right| \leq \frac{-t^2}{\sigma_n^2} \sum\sum_{i<j} cov(E_i, E_j)$$

which goes to zero as $n \to \infty$. Let $\{Z_j\}$ be an sequence of independent r.v. where Z_j is distributed as E_i.By construction and Lindberg-Feller CLT,

$$\prod_{j=1}^{n} Ee^{it\sigma_n^{-1}E_j} = \prod_{j=1}^{n} Ee^{itz_n^{-1}E_j} \to e^{\frac{-t^2}{2}}$$

and by theorem 4.1 we have,

$$\sigma_n^{'-1} \bigoplus_{i=1}^{n} \widetilde{X}_i \xrightarrow{d} N(0,1) f.r.v.$$

which completes the proof.

References

[1] Ash, R.B., and Doleans-Dade, C.A.(2000).*Probability Measure Theory, second Edition*

[2] Joe, H.,(1997). *Multivariate models and dependence concepts, Chapman and Hall.*

[3] Kwakernaak, H.,(1978) Fuzzy random variables I: definitions and theorems, *Inform.Sci.,*15 , 1-29.

[4] Puri, M.L., and Ralescu, D.A. (1986). Fuzzy random variabales, *J. Math.Anal.Appl.* 114, 409-422.

[5] Puri, M.L., and Ralescu, D.A. (1983). Differentials of fuzzy functions, *J.Math.Anal.Appl.* 91, 552-558.

[6] Sadeghpour-Gildeh, B., and Gien, D. (2001). La distance-Dp,q et le coefficient de corrélation entre deux variables aléatoires floues, Rencontres francophones sur la logique floue et ses applications LFA' 01, November 26-27, , Mons, Belgium, pp. 97-102.

[7] Terán, P. (2007) Probabilistic foundations for measurement modelling with fuzzy-random variables, *Fuzzy set and Systems*, 158, 973-986.

[8] Taylor, R.L., Patterson, R.F., and Bozorgnia, A.(2001), Limit theorems for negatively dependent random variables.*Nonlinear.Anal.*47, 1288-1295.

[9] Wu, H-C.(1999) The central limit theorems for fuzz random varables,*Inform.Sci.* 15, 239-256.

The Effect of Covariates on the Accuracy of Diagnostic or Screening Tests Using ROC Method

Najaf Zare[1], Mahshid Namdari[1], Faeze Jahanpour [2]

[1] *Department of Biostatistics, Faculty of Medicine, Shiraz University of Medical Sciences*
[2] *Faculty of Nursing and Midwifery, Shiraz University of Medical Sciences*

Abstract: Medical diagnostic tests are designed to discriminate between different states of health or medical conditions. Before diagnostic tests are implemented in practice, it is imperative that their accuracy or ability to discriminate, and factors that affect test's accuracy is studied. The accuracy of a diagnostic test can be summarized in a receiver operating characteristic (ROC) curve, a plot of true positive (TP) versus false positive (FP) rates associated with varying thresholds (c) for the test result (Y). In this approach we use ROC regression to compare ROC curves for continuous test results and for evaluating multiple covariates affecting on the accuracy of medical test. We consider a ROC model for which the ROC curve is a parametric function of covariates but distributions of the diagnostic test results are not specified (parametric distribution free approach). The general ROC regression model is $ROC_{X,X_D}(t) = g\{\sum_{k=1}^{K} \gamma_k h_k(t) + \beta X + \beta_D X_D\}$. The ROC curve is a function of covariates that can be either common to all subjects (X), or specific to those with disease (X_D). We use an estimation procedure based on binary indicators defined by the test result for a diseased subject exceeding various specified quantiles of the distribution of test results from non-diseased subjects with the same covariate values and then estimate parameters by fitting binary generalized linear models to these indicator variables. The methodology is applied to child's body temperature data set, where we use the regression framework to compare the accuracy of three body temperature measuring techniques (Right Tympanic, Left Tympanic, Axillary) in detecting febrile children. Our model shows that there is no difference in the accuracy of these measuring techniques. We also analyze factors such as antipyretic use, environment temperature and fever intensity (Rectal temperature), that supposed to influence the capacity of these techniques for distinguishing febrile children.

Keywords: Accuracy, Diagnostic tests, Discriminate, ROC regression analysis, Sensitivity, Specificity.

1 Introduction

Research into new diagnostic, screening, and prognostic disease markers and tests has exploded in recent years.The development of diagnostic tests for newly identified conditions is an essential first step in disease management.Diagnostic tests with improved accuracy or decreased cost are also being sought for established diseases. Screening biomarkers and tests have the potential to detect disease at an early stage, when it is more treatable. In each of these settings, the primary question is of classification accuracy: How well does the marker distinguish between the two groups of individuals, the "cases" and the "controls" (Janes,Pepe,2006).

Recently it has been recognized that various factors can affect the test performance beyond the disease status(Pepe (1997) and Pepe (2000)). Those factors include different test settings and/or subject's demographic data. One example is that for certain test whose test subjects include both men and women or both younger and older people, tests performance may vary between men and women or between younger and older people. Pepe listed several factors that can affect test performance, such as factors associated with test subject or tester, test settings and severity of disease (Pepe (2003),pp.48-49). It is therefore important to understand such influence to determine the optimal and suboptimal conditions or populations to perform such tests. If we find the test doesn't perform well for certain condition or population, then we may need to modify the test or even develop a new test for those situations. On the other hand, if we find that a factor doesn't influence test performance, we can relax the conditions under which the test is performed (Zhang,Pepe (2005)).

Comparing performance between several different tests is a special case of modelling covariate effects. When a new diagnostic

test is developed, before it can be used in the practice, frequently we need to compare it with an existing test to evaluate whether the new test provides better discrimination between cases and controls. Under certain situations (e.g., cost and invasiveness of the test), a new test is favored as long as it is proven to be non-inferior to its closest competitor (Zhang,Pepe (2005)).

The concept of covariate adjustment has been well studied in epidemiological and clinical research, as well as in statistics more broadly (Janes,Pepe (2006)).

In the field of medical diagnostic testing, the receiver operating characteristics(ROC) curve has long been used as a standard statistical tool to assess the accuracy of tests that yield continuous or ordinal results (Alonzo,Pepe (2006)).

Let D be a binary variable taking the value 1 for diseased subjects and 0 for non diseased subjects. Let the variable Y denote the continuous test result, and use the convention that higher values of Y are considered more indicative of disease .The ROC curve is motivated as follows : if a treshold value c is used to classify subjects as diseaseed or not on the basis of Y, then the true-positive and false-positive rates can be written as:$TPR(c) = P(Y \geq c \mid D = 1)$ and $FPR(c) = P(Y \geq c \mid D = 0)$ (Cai,Pepe (2002)).

The TPR, also called the sensitivity, is the proportion of diseased subjects correctly detected by the test. On the other hand, FPR or (1-specificity) is defined as the proportion of non-diseased subjects erroneously deemed positive by the test (Alonzo,Pepe (2006)).

ROC curve is a plot of true positive (TP) versus false positive (FP) rates associated with varying thresholds c for the test result Y (Pepe (2000)).

ROC regression methodology is used to identify factors that affect the discriminating capacity of non-binary test. In this approach we use ROC regression to compare ROC curves for continuous test results and for evaluating multiple covariates affecting on the accuracy of medical test.

A review of the three major existing approaches to ROC regression is provided in a paper written by Pepe (1998). The first approach which was proposed by Tosteson and Begg uses regression models for the test outcome and infers covariate effects on the corresponding ROC curves. The second approach considers regression models for the area under the ROC curve (AUC), a common summary measure of the ROC curve, was proposed by Thompson and Zucchini. Finally, a parametric distribution-free (PDF) approach that directly models the ROC curve has been proposed by Pepe (1997). In a detailed comparison of the three approaches, Pepe (1998) notes several major advantages to the latter approach, including the facts that it can accommodate multiple test types and continuous covariates and that models can pertain only to restricted portions of the ROC curve that are of interest .

Thus, we consider the direct modelling approach in this paper. We consider a ROC model for which the ROC curve is a parametric function of covariates but distributions of the diagnostic test results are not specified (parametric distribution free approach, PDF). We are interested in determining the effect of a covariate vector X on the accuracy of a continuous diagnostic test Y.

2 Material and Method

2.1 Modelling ROC curve

Let (Y_D) and $(Y_{\bar{D}})$ denote test result random variables from diseased (D) and non-diseased $(\bar{D})$ populations, respectively. We assume that larger values of Y are more indicative of disease and smaller values are less indicative of disease.

For estimating the ROC curve , S_D and $S_{\bar{D}}$ denote the survivor functions for Y in the diseased and non-diseased populations

$$S_D(y) = P(Y \geq c \mid D = 1)) = P(Y_D \geq c)$$
$$S_{\bar{D}}(y) = P(Y \geq c \mid D = 0) = P(Y_{\bar{D}} \geq c)$$

and note that $S_D(c) = TPR(c)$ and $S_{\bar{D}}(y) = FPR(c)$, then the

ROC curve can be written as:

$$ROC(t) = S_D(S_{\bar{D}}^{-1}(t)), t \in T$$

where t is the set of possible FP rates attainable by varying $c \in (-\infty + \infty)$, i.e $T = \{S_{\bar{D}}(c), c \in (-\infty, +\infty)\}$.

This follows simply from the definition of the ROC curve. Let $c = S_{\bar{D}}^{-1}(t)$ .That is,c is the treshold corresponding to the false positive fraction t, so that $P(Y \geq c \mid D = 0) = t$.The corresponding true positive fraction is $P(Y \geq c \mid D = 1) = S_D(c)$. So the TPR that corresponds to the FPR=t is $ROC(t) = S_D(c) = S_D(S_{\bar{D}}^{-1}(t))$ (Pepe (2003),pp.67-70).

That is, $ROC(t)$ is the probability that a diseased individual has test results (Y_D) that are greater than or equal to the tth quantile of the distribution of test results from non-diseased individuals (Alonzo,Pepe (2002)).

Suppose the ROC curve is modelled parametrically with the form; $ROC_\gamma(t) = g\{\sum_{k=1}^{K} \gamma_k h_k(t)\}$, for some specified link function g, basis functions $h_1, ..., h_k$ and unknown parameter γ .The functions g() and $h_k()$are chosen so that the ROC curve is monotone increasing on the unit square. For example ,with the probit link, $g = \Phi$, the cumulative normal distribution function, and basis functions $h_1(t) = 1$ and $h_2(t) = \Phi^{-1}(t)$. These choices yield the binormal model: $ROC_\gamma(t) = \Phi\{\gamma_1 + \gamma_2\Phi^{-1}(t)\}$.

This approach is referred to as PDF, because the approach specifies a parametric model for the ROC curve but does not assume distributions for the diagnostic test results.

Alonzo and Pepe (2002), developed a method for fitting this regression model based on binary indicators defined by a test result for a diseased subject exceeding various specified quantiles of the distribution of test results from non-diseased subjects. By this explanation binary indicator will be of the form $U_{it} = I[Y_{Di} \geq S_{\bar{D}}^{-1}(t)]$ for $t \in (T)$, where t is a FPR between 0 and 1 and denotes a fixed finite set of such values. The key observation is

$$\begin{aligned} E[U_{it}] &= P(Y_{Di} \geq S_{\bar{D}}^{-1}(t)) \\ &= S_D(S_{\bar{D}}^{-1}(t)) \\ &= ROC_\gamma(t) \\ &= g\{\sum_{k=1}^{K} \gamma_k h_k(t)\} \end{aligned}$$

then for estimating model parameters $\{\gamma_k, k = 1, ..., K\}$we can fit binary generalized linear models to these indicator variables:$\{U_{it}, i = 1, ..., n_D; t \in (T)\}$, where n_D and $n_{\bar{D}}$denote the number of observations for diseased and non-diseased test units.

Alonzo and Pepe's (2002) algorithm for estimating the model parameters is as follows: (1) specify a set of FPRs, $t \in (T)$, to consider

(2) estimate $S_{\bar{D}}^{-1}(t)$ for $t \in (T)$ i.e. calculate the tth quantile of the survivor distribution of the non-diseased test results. This can be accomplished using empirical estimates when applicable or regression quantile methods. We refer the reader to Koenker and Basset (1978), and Heagerty and Pepe (1991).

(3) calculate $U_{it} = I[Y_{Di} \geq S_{\bar{D}}^{-1}]$ for $i = 1, ..., n_D$ and $t \in (T)$

(4) fit the model $E[U_{it}] = g\{\sum_{k=1}^{K} \gamma_k h_k(t)\}$ by solving standard estimating equations for fitting a binary generalized linear model to U_{it} with the link function $g^{-1}$and covariates$\{h_k(t); k = 1, ..., K\}$

$$\sum_{i=1}^{n_D} \sum_{t \in T} S_i(\gamma, t) = \sum_{i=1}^{n_D} \sum_{t \in T} h(t)\omega_\gamma(t) \left(U_{it} - g\{\textstyle\sum_{k=1}^{K} \gamma_k h_k(t)\} \right) = 0$$

where $\omega_\gamma(t) = [(\partial/\partial l)g(l)]/g(l)(1 - g(l))$ with $l = \sum_{k=1}^{K} \gamma_k h_k(t)$

2.2 Modelling covariate effects on ROC curve

It is of interest to determine the effect of covariates, denoted by X and X_D , on the ROC curve, where X represents covariates common to diseased and non-diseased subjects and X_D denotes covariates that

are specific to the diseased state. For example, X denotes age and gender and X_D describes a measure of how severe is the disease.

The ROC curve corresponding to (X, X_D) can be written as $ROC_{X,X_D}(t) = S_{D,X,X_D}(S^{-1}_{\bar{D},X}(t))$,where $t \in (0,1)$ is the FPR and $S_{D,X,X_D}(c) = P(Y_D \geq c \mid X, X_D)$ and $S_{\bar{D},X}(c) = P(Y_{\bar{D}} \geq c \mid X)$ are survivor functions at threshold c. That is, $ROC_{X,X_D}(t)$ is the probability that a diseased individual with disease-specific covariates X_D and common covariates X has test results Y_D that are greater than or equal to the tth quantile of the distribution of test results from non-diseased individuals. The general ROC regression model we consider is $ROC_{X,X_D}(t) = g\{\sum_{k=1}^{K} \gamma_k h_k(t) + \beta X + \beta_D X_D\}$. That is, the ROC curve is a function of covariates common to diseased and non-diseased subjects, covariates specific to diseased subjects, and a function h() which defines the location and shape of the curve. The functions g() and $h_k()$are chosen so that the ROC curve is monotone increasing on the unit square. In practice, g() =Φ, the cumulative normal distribution function, $h_1(t) = 1$, and $h_2(t) = \Phi^{-1}(t)$ are often used. These choices yield the binormal model: $ROC_{X,X_D}(t) = \Phi(\gamma_1 + \gamma_2\Phi^{-1}(t) + \beta X + \beta_D X_D)$.

This model specifies that the ROC curves for different values of X and X_D differ by fixed amounts on the probit scale. If $\beta > 0$, then the discrimination between Y_D and $Y_{\bar{D}}$ increases with increasing values of X . Similarly, if $\beta_D > 0$, diseased subjects with larger values of X_D are more distinct from non-diseased subjects than are diseased subjects with smaller values of X_D. A more flexible model could be fit by including an interaction between X or X_D and $\Phi^{-1}(t)$ allowing the effects of (X, X_D) to differ by varying amounts depending on the FPR t.

Alonzo and Pepe (2002) developed a method for fitting this regression model based on binary indicators defined by a test result for a diseased subject exceeding various specified quantiles of the distribution of test results from non-diseased subjects with the same covariate values.By this explanation binary indicator will be of the form $U_{it} = I[Y_{Di} \geq S^{-1}_{\bar{D},X_i}(t)]$ for $t \in (T)$, where t is a FPR between 0 and 1 and denotes a fixed finite set of such values .The key observation is

$$\begin{aligned} E[U_{it}] &= P(Y_{Di} \geq S^{-1}_{\bar{D},X_i}(t) \mid X_i, X_{Di}) \\ &= S_{D,X_i,X_{Di}}(S^{-1}_{\bar{D},X_i}(t)) \\ &= ROC_{X_i,X_{Di}}(t) \\ &= g(\{\sum_{k=1}^{K} \gamma_k h_k(t) + \beta X + \beta_D X_D\}) \end{aligned}$$

then for estimating model parameters $\{\beta, \beta_D, \gamma_k, k = 1, ..., K\}$we can fit binary generalized linear models to these indicator variables:$\{U_{it}, i = 1, ..., n_D; t \in (T)\}$, where n_D and $n_{\bar{D}}$denote the number of observations for diseased and non-diseased test units.

Alonzo's algorithm for estimating the model parameters is as follows:

(1) specify a set of FPRs, $t \in (T)$, to consider

(2) estimate $S^{-1}_{\bar{D},X_i}(t)$ for $t \in (T)$ i.e. calculate the tth covariate-specific quantile of the survivor distribution of the non-diseased test results.

(3) calculate $U_{it} = I[Y_{Di} \geq S^{-1}_{\bar{D},X_i}(t)]$ for $i = 1, ..., n_D$ and $t \in (T)$

(4) fit the model $E[U_{it}] = g\{\sum_{k=1}^{K} \gamma_k h_k(t) + \beta X + \beta_D X_D\}$ by solving standard estimating equations for fitting a binary generalized linear model to U_{it} with the link function $g^{-1}$and covariates$\{h_k(t), X_i, X_{Di}; k = 1, ..., K\}$

$$\sum_{i=1}^{n_D} \sum_{t \in T} S_i(\gamma, \beta, \beta_D, t) = \sum_{i=1}^{n_D} \sum_{t \in T} \begin{pmatrix} h(t) \\ X_i \\ X_{Di} \end{pmatrix} \omega_{\gamma,\beta}(t)$$
$$\times \left(U_{it} - g\{\textstyle\sum_{k=1}^{K} \gamma_k h_k(t) + \beta X + \beta_D X_D\} \right) = 0 \qquad (1)$$

where $\omega_\gamma(t) = [(\partial/\partial l)g(l)]/g(l)(1-g(l))$ with $l = \sum_{k=1}^{K} \gamma_k h_k(t) + \beta X + \beta_D X_D$

3 Application to child's body temperature data

The methodology is applied to child's body temperature data, where we use the regression framework to compare the accuracy of 3 methods of body temperature measuring (Right Tympanic, Left Tympanic, Axillary) in children, and we also analyze factors such as antipyretic use, environment temperature and fever intensity (Rectal temperature), that supposed to influence the capacity of these three methods for distinguishing febrile children. The measurements were obtained on 220 children who were between ages 3 months and 5 years old, who presented to Ali Asghar's hospital in Booshehr.Exclusion criteria included admission for surgical procedures at least on the ear, axilla or rectal possitions.First we take admission from child's parents for including them in the study.Then child's demographic information was recorded .The tools used to measure temperature were calibrated prior to the study. The order of measurement sites was the ears , then the axilla and at last the rectal.Tympanic recordings were obtained using Omron gentle temp 510 ,axilla recordings were obtained using Omron flex temp 0197 and recctal recordings were obtained using Omron pro temp cE 0473. For each body temperature measuring technique, we asked a trained nurse to obtain that temperature and other nurses were not informed about measured temperatures with other techniques. Environment temperature was measured by the thermometer that was instaled in child's room, the environment temperature ranged from $24°C$to $28.5°C$. The temperature recordings for all four sites ranged from $34.4°C$to $39.3°C$,with a mean of $36.4°C$for the right tympanic, $36.4°C$ for the left tympanic, $36.7°C$ for the axilla and $36.6°C$ for the rectal sites.Fever was detected as a rectal temperature of greater than $38°C$ (Gold standard test).

3.1 Modelling ROC curves for three measuring techniques

For modelling the ROC curves for each of the three body temperature measuring techniques (Right Tympanic, Left Tympanic, Axillary) in detecting febrile children, using the notation introduced before, D corresponds to the indicator of febrile child (a rectal temperature of greater than $38°C$) and Y is temperatures measured according to each

technique. Question of interest is to model ROC curves for each technique.

Of the 220 records for each measuring technique,192 and 28 correspond to febrile and non-febrile children , respectively. Thus, $n_{\bar{D}} = 192$ and $n_D = 28$. Using the algorithm that was described before, we first specified the set of FPRs to be used.Based on the simulation results summarized in (Alonzo , pepe (2002)) , 50 equally spaced FPRs, T = $\{1/51, 2/51, \ldots, 50/51\}$, were used. Next, quantiles of the survivor distribution of $Y_{\bar{D}}$, temperatures for non-febrile children in each technique, were estimated for $t \in T$ using regression quantile methods . Then $U_{it} = I[Y_{Di} \geq S^{-1}_{\bar{D},X_i}]$ was calculated for $i = 1, ..., 192$ and $t \in T$. A probit regression model was then fitted to U_{it} with covariate $\Phi^{-1}(t)$. The following ROC regression model was fitted:

$$ROC_{X,X_D}(t) = \Phi(\gamma_1 + \gamma_2\Phi^{-1}(t))$$

The results are summarized in Table 1. Standard errors were es-

Table 1. Estimated parameters for the ROC analysis of three measuring techniques.

Technique	Variable	Coefficient	Standard error	p-value
Right Tympanic	Intercept	2.53	0.15	0.00
	$\Phi^{-1}(t)$	1.05	0.12	0.00
Light Tympanic	Intercept	2.19	0.11	0.00
	$\Phi^{-1}(t)$	0.84	0.10	0.00
Axillary	Intercept	2.06	0.10	0.00
	$\Phi^{-1}(t)$	0.80	0.09	0.00

timated using 500 bootstrap samples. The jackknife-after-bootstrap suggested 500 bootstrap samples was adequate (Alonzo , pepe (2002)).

By comparing estimated parameters in three techniques we find that the accuracy of these three techniques is not too much different.

3.2 Modelling covariate effects on the ROC curve

For comparing the ROC curves for three techniques (Right Tympanic, Left Tympanic, Axillary) in detecting febrile children.Using the notation introduced before, Y is temperatures measured according to three techniques.Each child contributes three observations to the analysis. Let X_1 be a corresponding indicator variable ,equal to one for the Right tympnic technique and zero for others and X_2 be a corresponding indicator variable ,equal to one for the Left tympnic technique and zero for others .Our full model will be

$$ROC_{X,X_D}(t) = \Phi(\gamma_1 + \gamma_2\Phi^{-1}(t) + \beta_1 X_1 + \beta_2 X_1\Phi^{-1}(t) + \beta_3 X_2 + \beta_4 X_2\Phi^{-1}(t))$$

In the above model we use marker type as a stratification covariate . This model is very flexible, in that it allows the effects of(X, X_D) on the ROC curves to differ by varying amounts depending on the FPR t.Standard errors were estimated using 500 bootstrap samples, with the unit for resampling being the cluster of data for the study subject. With backward regression parameter estimation suggested that interactions between covariates and $\Phi^{-1}(t)$ and stratification covariates were not significant and, thus, the following reduced model was more appropriate:

$$ROC_{X,X_D}(t) = \Phi(\gamma_1 + \gamma_2\Phi^{-1}(t))$$

This model implies that ROC curves for different kinds of techniques is same.Our model shows that there is no difference in the accuracy of these measuring techniques. For analyzing the effect of child's covariates on each of the three techniques (Right Tympanic, Left Tympanic, Axillary), Y is temperatures measured according to each technique. Questions of interest are to determine at different

Table 2. Estimated parameters for the ROC analysis to compare three measuring techniques.

Variable	Coefficient	Standard error	p-value
Intercept	2.22	0.33	0.00
$\Phi^{-1}(t)$	0.87	0.2	0.00

states of antipyretic use and environment temperature how well each technique discriminates febrile children from non-febriles. We also want to know how well each technique would distinguish a child who is severely febrile from non-febriles. Therefore, the covariates of interest, X , are the antipyretic use and environment temperature, and X_D is fever intensity.

We first specified the set of FPRs to be used. 50 equally spaced FPRs, T = {1/51, 2/51, . . . , 50/51}, were used. Next, quantiles of the survivor distribution of $Y_{\bar{D}}$, measured temperature for each technique for non-febrile children , were estimated as a function of antipyretic use and environment temperature for $t \in T$ using regression quantile methods. Then $U_{it} = I[Y_{Di} \geq S^{-1}_{\bar{D},X_i}]$ was calculated for i =1,... ,192 and $t \in T$. A probit regression model was then fitted to U_{it} with covariates $\Phi^{-1}(t)$, X = (antipyretic use , environment temperature), and X_D= fever intensity . The results are summarized in Table 3. By the above estimates we find out the amount of effect of each covariate on each technique, for example the positive coefficient for fever intensity in Left Tympanic and Axillary is positive, implies that it is easier to distinguish highly febrile children from non-febriles than to distinguish moderately febriles from non-febriles.

4 Discussion

The main point we make in this paper is that some covariates can affect the accuracy of medical diagnostic test.Standard binary regression methodology applied to binary variables, $U_{it} = I[Y_{Di} \geq S^{-1}_{\bar{D},X_i}]$

Table 3. Estimated parameters for the ROC analysis of three measuring techniques and childs covariates.

Technique	Variable	Coefficient	Standard error	p-value
Right Tympanic	Intercept	-36.33	19.48	0.06
	$\Phi^{-1}(t)$	74.03	18.30	0.00
	Antipyretic-Use	-0.28	0.29	0.34
	Fever Intensity	1.01	0.51	0.05
	$\Phi^{-1}(t)*$Antipyretic-Use	1.55	0.31	0.00
	$\Phi^{-1}(t)*$Fever Intensity	-1.91	0.48	0.00
Left Tympanic	Intercept	-108.40	1.97	0.00
	$\Phi^{-1}(t)$	-16.38	2.49	0.00
	Antipyretic-Use	-1.82	0.17	0.00
	Environment-Temp	0.56	0.10	0.00
	Fever Intensity	2.50	0.29	0.00
	$\Phi^{-1}(t)*$Environment-Temp	0.65	0.10	0.00
Axillary	Intercept	-89.30	11.41	0.00
	$\Phi^{-1}(t)$	-8.18	2.47	0.001
	Antipyretic-Use	-1.61	0.22	0.00
	Environment-Temp	0.61	0.27	0.00
	Fever Intensity	1.97	0.27	0.00
	$\Phi^{-1}(t)*$ Antipyretic-Use	-0.34	0.17	0.03
	$\Phi^{-1}(t)*$Environment-Temp	0.35	0.09	0.00

yield mechanism for estimating, comparing and evaluating covariate effects on ROC curves.

Alonzo and Pepe(2002) investigated the efficiency of the PDF approach and concluded that the PDF and maximum likelihood (ML) approaches have similar efficiency and MSE, and suggested that PDF estimator is reasonably efficient . Pepe (1998), in a detailed comparison of the three major existing approaches to ROC regression notes several major advantages to PDF approach, including the facts that it can accommodate multiple test types and continuous covariates and that models can pertain only to restricted portions of the ROC curve that are of interest .

Weiss (1991), compared the three body temperature measuring techniques, concluding that their accuracy was similar and suggested that differences between these techniques may be as a result of environmental influences .The issue of whether environment influence Tympanic temperature remains controversial (Baily,Rose (2001)).

References

[1] Alonzo, T.A. Pepe, M.S. (2002). Distribution free ROC analysis using binary regression techniques.*Biostatistics,* 3,3, 421-432.

[2] Bailey, J . Rose,p. (2001). Axillary and tympanic membrane temperature recording in the neonate: a comparative study.*J Adv Nurs,* 6, 165-168.

[3] Cai, T. Pepe, M.S.(2002) .Semiparametric Receiver Operating Characteristic Analysis to Evaluate Biomarkers for Disease.*Journal of the American Statistical Association,* 97,460, 1099-1106.

[4] Janes, H. Pepe, M.S. (2006). Adjusting for covariate effects on classification accuracy using the covariate-adjusted ROC curve.*UW Biostatistics working paper series,* paper 283, 1-23.

[5] Heagerty, P. J. Pepe, M. S. (1991) .Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in children. *Applied Statistics,* 48, 533 - 551.

[6] Koenker, R. Basset, G.(1978) .Regression quantiles.*Econometrica,* 46, 33-50.

[7] Pepe, M.S . (1997). A regression modeling framework for receiver operating characteristic curves in medical diagnostic testing. *Biometrika,* 84, 595-608.

[8] Pepe, M.S.(1998) . Three Approaches to Regression Analysis of Receiver Operating Characteristic Curves for Continuous Test Results.*Biometrics,* 54, 124-135.

[9] Pepe, M.S.(2000) . An interpretation for the ROC curves and inference using GLM procedures. *Biometrics,* 56, 352-359

[10] Pepe, M.S.(2003) .The statistical evaluation of medical tests in classification and prediction, 1st edition Oxford University press, New York.

[11] Weiss, M.(1991) Tympanic infrared thermometry for full term and preterm neonates. *clinical pediatrics supplement,* 30, 42-45.

[12] Zhang, Z. and Pepe, M.S. (2005). A linear regression framework for receiver operating characteristic(ROC) curve analysis. *UW Biostatistics working paper series,* paper 253, 1-22.

A study on the residual Rényi entropy of order statistics

Somayeh Zarezadeh

Department of Statistics, University of Isfahan

Abstract: This paper explores some properties of the residual Rényi entropy and residual Rényi information for the order statistics. In the first, we obtain the relation between the residual Rényi entropy of the kth order statistic from an arbitrary distribution with the residual Rényi entropy of the kth order statistic from uniform distribution. Then we give some bounds for the residual Rényi entropy of these ordered random variables. It is also shown that although the Rényi information between the distribution of the kth order statistic and the underlying distribution is distribution free however the corresponding dynamic Rényi information is model dependent. This result is also proved for the consecutive order statistics.

Keywords: Order statistic, Rényi entropy, Residual lifetime, Rényi information, Incomplete beta function.

1 Introduction

Let X be a non-negative continuous random variable with density function f and distribution function F. In addition suppose that $X_1, ..., X_n$ is a random sample from X. The order statistics of this sample is defined by the arrangement of $X_1, ..., X_n$ from smallest to the largest, denoted as $X_{1:n} \leq ... \leq X_{k:n}$. It is well known that the survival function $\bar{F}_{k:n}(x) = 1 - F_{k:n}, k = 1, ..., n$ is given by

$$\bar{F}_{k:n}(x) = \sum_{i=0}^{k-1} \binom{n}{i} F^i(x) \bar{F}^{n-i}(x) = \frac{\bar{B}_{F(x)}(k, n-k+1)}{B(k, n-k+1)} \qquad (1)$$

where

$$\bar{B}_F(k, n-k+1) = B(k, n-k+1) - B_{F(x)}(k, n-k+1) \qquad (2)$$

and $B_{F(x)}(k, n-k+1)$ is known as the incomplete beta function at $F(x)$. The density function corresponding to this equals

$$f_{k:n}(x) = \frac{1}{B(k, n-k+1)}[F_X(x)]^{k-1}[1-F_X(x)]^{n-k} f_X(x). \qquad (3)$$

Order statistics are used in a wide range of problems, including in characterization of probability distributions, analysis of censored samples, reliability analysis, goodness-of-fit tests, quality control and so on.

In information theory to measure the amount of information in a probability distribution the Rènyi's entropy (Rènyi (1961)) plays a central role. Let X be a non-negative continuous random variable with density function f. The Rényi entropy of X, which we denote by $H_\alpha(f)$, is defined as follows:

$$H_\alpha(f) = \frac{1}{1-\alpha} \log \int_0^\infty f^\alpha(x) dx$$

where $\alpha > 0$, $\alpha \neq 1$. It is well known that when α tends to 1, $H_\alpha(f)$ tends to Shannon entropy (Shannon (1948)) which we denote by $H(f)$. That is

$$\lim_{\alpha \to 1} H_\alpha(f) = H(f) = -\int_0^\infty f(x) \log f(x) dx.$$

The Rényi entropy has a wide range of applications in many fields from electrical engineering, physics, chemistry and computer sciences to economics, biology and medicine genetics. Several properties of the Rényi entropy are explored by Rényi (1961), Morales et al. (1997), Song (2001). Recently, Rényi entropy of parametric distributions are tabulated by Nadarjah and Zografos (2003).

Let the random variable X denote a duration such as the lifetime of a system. Usually in reliability theory and survival analysis, when the system is still alive at time t, one is interested in studying the properties of the residual lifetime of the system. The residual lifetime of the system, which we denote by X_t, is $X_t = X - t | X > t$. In this case the Rényi entropy is not an appropriate measure. Motivated by

this, Asadi et al. (2005) introduced a concept of Rényi entropy for the residual lifetime as follows

$$H_\alpha(X;t) = \frac{1}{1-\alpha}\log\int_t^\infty \frac{f^\alpha(x)dx}{\bar{F}^\alpha(t)} \tag{4}$$

where $\alpha > 0$, $\alpha \neq 1$ and $\bar{F}$ is the survival function of X. If $\lambda_F(t) = \frac{f(t)}{\bar{F}(t)}$ denote the failure rate of X then it can be written

$$H_\alpha(X;t) = \frac{1}{1-\alpha}\log E\left[(\lambda_F(X^*))^{\alpha-1}|X^* > t\right] - \frac{1}{1-\alpha}\log\alpha \tag{5}$$

where X^* is a random variable with survival function $\bar{F}^\alpha(t)$. Several properties of this measure has been derived by these authors. This measure is appropriate for comparison of distributions of residual lifetime. Mahmoudi and Asadi (2008) explored some monotone behavior of the residual Rényi entropy.

The wide scope of applications of order statistics gives attention to the study of information properties of them. Ebrahimi et al. (2004) explored some properties of the Shannon entropy for the order statistics. Recently, some results based on Rényi entropy of the order statistics are studied by Baratpour et al. (2008).

The aim of the present paper is to study some properties of residual Rényi entropy and residual Rényi information of order statistics. In Section 2, we obtain some bounds for residual Rényi entropy and give some examples. In Section 3, we show that although the Rényi information between the distribution of the kth order statistic and the underlying distribution is distribution free however the corresponding dynamic Rényi information is model dependent.

2 Results on residual Rényi entropy of order statistics

In this section we concentrate on the residual entropy of order statistics. First we give the following lemma.

Lemma 1 *Let $U_{k:n}$ is* k*th order statistic of $U(0,1)$ distribution. It can be shown that*

$$H_\alpha(U_{k:n};t) = \frac{1}{1-\alpha} log\bar{B}_t(\alpha(k-1)+1, \alpha(n-k)+1) - \frac{\alpha}{1-\alpha} log\bar{B}_t(k, n-k+1)$$

Now the residual Rényi entropy of order statistics from arbitrary distribution can be found by noting that $U_{k:n} = F_X(X_{k:n}),\ k = 1, ..., n$ where $U_{k:n}$ is the kth order statistic from a random sample of size n from $U(0,1)$ distribution. This transformation formula gives the following theorem for the residual Rényi entropy of the order statistics.

Theorem 1 *Residual Rényi entropy of* k*th order statistic from arbitrary distribution can be written in terms of the Residual Rényi entropy of* k*th order statistic from uniform distribution over the unit interval as follows*

$$H_\alpha(X_{k:n};t) = H_\alpha(U_{k:n}; F(t)) + \frac{1}{1-\alpha} log E[f^{\alpha-1}(F^{-1}(Y_k))]$$

where $Y_k \sim \bar{B}_{F(t)}(\alpha(k-1)+1, \alpha(n-k)+1)$.
Note that, in this paper $X \sim \bar{B}_t(a,b)$ denotes distribution with density function

$$f_X(x) = \frac{1}{\bar{B}_t(a,b)} x^{a-1}(1-x)^{b-1} \quad ; t \le x \le 1. \tag{6}$$

Now we give some examples.

Example 1 Suppose that X is a random variable having the exponential distribution with mean $\frac{1}{\theta}$. Then $f(F^{-1}(y)) = \theta(1-y)$ and we have

$$E[f^{\alpha-1}(F^{-1}(Y_1))] = \frac{\theta^{\alpha-1}\bar{F}^{n\alpha}(t)}{n\alpha\bar{B}_{F(t)}(1, \alpha(n-1)+1)}$$

For $k = 1$, theorem 1 gives

$$H_\alpha(X_{1:n};t) = \frac{log\alpha}{\alpha-1} - log(n\theta)$$

On the other hand, we have

$$H_\alpha(X;t) = \frac{log\alpha}{\alpha - 1} - log\theta$$

Since f is strictly decreasing function, this is other reason of this fact that $X_{1:n}$ has exponential distribution with mean $\frac{1}{n\theta}$ (Asadi et al.(2005)).

Example 2 Let X be a random variable with density function as follows

$$f(x) = \frac{\theta\beta^\theta}{x^{\theta+1}} \qquad x \geq \beta > 0, \theta > 0$$

Then $f(F^{-1}(y)) = \frac{\theta}{\beta}(1-y)^{1+\frac{1}{\theta}}$. Therefore for the first order statistic of random sample of size n from this distribution we have

$$E[f^{\alpha-1}(F^{-1}(Y_1))] = \frac{1}{\bar{B}_{F(t)}(1, \alpha(n-1)) + 1}(\frac{\theta}{\beta})^{\alpha-1}\frac{(\bar{F}(t))^{n\alpha+\frac{\alpha-1}{\theta}}}{n\alpha\theta + \alpha - 1}$$

Then for $k = 1$, theorem 1 gives

$$H_\alpha(X_{1:n};t) = logt + \frac{\alpha}{1-\alpha}lognθ - \frac{1}{1-\alpha}log(\alpha(n\theta+1)-1) \quad ; t \geq \beta, \alpha > \frac{1}{n\theta+1}$$

On the other part, we have

$$H_\alpha(X;t) = logt + \frac{\alpha}{1-\alpha}log\theta - \frac{1}{1-\alpha}log[\alpha(\theta+1)-1] \quad ; t \geq \beta, \alpha > \frac{1}{\theta+1}$$

Since $f(x)$ is strictly decreasing over $x \geq \beta$, then for $\alpha > 1$, $H_\alpha(X;t)$ uniquely determines F (Asadi et al.(2005)). Therefore the amount of $H_\alpha(X_{1:n};t)$ is other reason of this fact that $X_{1:n}$ has Pareto distribution with parameters $n\theta$ and β .

In the sequel we obtain some bounds for residual entropy of order α of order statistics. The following theorem provides bounds for the residual Rényi entropy of the order statistics in terms of the residual Rényi entropy of data distribution and mode of distribution.

Theorem 2 *Let X be a nonnegative continuous random variable with density function f and distribution function F. The residual Rényi entropy of order statistics $X_{k:n}, k = 1, ..., n$ is bounded as follows:*

(a) Let residual Rényi entropy of X, $H_\alpha(X;t) < \infty$. Then for $\alpha > 1 (0 < \alpha < 1)$

$$H_\alpha(X_{k:n};t) \geq (\leq) b_k(t) + H_\alpha(X;t) + \frac{\alpha}{1-\alpha} log\bar{F}(t)$$

where

$$b_k(t) = \frac{\alpha}{1-\alpha}[(k-1)log(k-1) + (n-k)log(n-k) - (n-1)log(n-1) \\ -log\bar{B}_{F(t)}(k, n-k+1)]$$

(b) Let $M = f_X(m) < \infty$, where $m = sup\{x : f_X(x) \leq M\}$ is the mode of the distribution. Then for $\alpha > 0, \alpha \neq 1$

$$H_\alpha(X_{k:n};t) \geq H_\alpha(U_{k:n}; F(t)) - logM$$

Proof: (a) In according to theorem (1), it is enough for attaining a bound of $H_\alpha(X_{k:n};t)$ to achieve a bound for $\frac{1}{1-\alpha}E[f^{\alpha-1}F^{-1}(Y_k)]$. We have $m_k = \frac{k-1}{n-1}$ is mode of the distribution of Y_k. Let $M_k = f_{Y_k}(m_k)$ then for $\alpha > 1 (0 < \alpha < 1)$

$$\begin{aligned} \frac{1}{1-\alpha}E[f^{\alpha-1}(F^{-1}(Y_k))] &\geq (\leq) \frac{1}{1-\alpha}logM_k \\ &+ \frac{1}{1-\alpha}log\int_{F(t)}^{1} f^{\alpha-1}(F^{-1}(y))dy \\ &= \frac{1}{1-\alpha}logM_k + \frac{1}{1-\alpha}log\int_{t}^{\infty} f^{\alpha}(u)du \\ &= \frac{1}{1-\alpha}logM_k + H_\alpha(X;t) + \frac{\alpha}{1-\alpha}log\bar{F}(t) \end{aligned}$$

where the first equality is attained by change of variable $u = F^{-1}(y)$.

(b) By attention to theorem 1 and by considering of mode of data distribution it is proved part (b).

Example 3 In the following we compute the bounds for the residual Rényi entropy of the sample minimum for some well known distributions.

(I) For the uniform distribution over the interval (a, b), $\bar{F}(t) = \frac{b-t}{b-a}$ and $H_\alpha(X;t) = log(b-t)$. Thus, for the sample minimum of this distribution and $\alpha > 1 (0 < \alpha < 1)$ by using part (a) we have

$$H_\alpha(X_{1:n};t) \geq (\leq)\frac{\alpha}{1-\alpha}logn + \frac{1-n\alpha}{1-\alpha}log(b-t) + \frac{(n-1)\alpha}{1-\alpha}log(b-a)$$

Also by using part(b), for $\alpha > 0, \alpha \neq 1$ we have

$$H_\alpha(X_{1:n};t) \geq log(b-t) - \frac{1}{1-\alpha}log[\alpha(n-1)+1] + \frac{\alpha}{1-\alpha}logn$$

For $\alpha > 1$ the difference between two lower bound is

$$\frac{\alpha(n-1)}{1-\alpha}log\frac{b-a}{b-t} + \frac{1}{1-\alpha}log[\alpha(n-1)+1] \leq 0$$

That is the achieved bound of part (b) is more correct than the achieved bound of part (a).

Remark 1 Note that for $U(a, b)$ distribution and any $k = 1, ..., n$ the bound of part (b) is equal to exact amount of $H_\alpha(X_{k:n};t)$. Then the lower bound that it is attained in part (b) of theorem 2 is sharp.

(II) For the exponential distribution with mean $\frac{1}{\theta}$, $H_\alpha(X;t) = \frac{log\alpha}{\alpha-1} - log\theta$ and $\bar{F}(t) = e^{-\theta t}$. Then by using part (a) for the sample minimum and $\alpha > 1 (0 < \alpha < 1)$ we have

$$H_\alpha(X_{1:n};t) \geq (\leq)\frac{\alpha}{1-\alpha}\theta t(n-1) + \frac{\alpha}{1-\alpha}logn - \frac{log\alpha}{1-\alpha} - log\theta.$$

For $\alpha > 1$, the difference between $H_\alpha(X_{1:n};t)$ and the lower bound is as follows

$$L.B.(a) - L.B.(b) = \frac{1}{\alpha-1}[logn + \alpha\theta t(n-1)]$$

which is an increasing function of n. Thus the bound is useful when n is not large. In other hand by using part (b), it can be written

$$H_\alpha(X_{1:n};t) \geq -\theta t - \frac{1}{1-\alpha} log[\alpha(n-1)+1] + \frac{\alpha}{1-\alpha} log\alpha - log\theta.$$

For $\alpha > 1$, the difference between achieved lower bound of part (a) and part (b) is as follows:

$$L.B.(a) - L.B.(b) = \frac{1}{1-\alpha}[\theta t(\alpha(n-1)+1) + log(n-1+\frac{1}{\alpha})] \leq 0.$$

Then it is explored that the achieved lower bound in part (b) is better.

(III) The density function of Pareto distribution with parameters θ and β is

$$\begin{aligned} f_X(x) &= \frac{\theta\beta^\theta}{x^{\theta+1}} \; for \; x \geq \beta > 0, \alpha > 0 \\ &= o \quad otherwise. \end{aligned}$$

We have from parts (a) and (b), respectively

$$H_\alpha(X_{1:n};t) \geq \frac{\alpha}{1-\alpha} log(n\theta) + logt - \frac{1}{1-\alpha} log[\alpha(\theta+1)-1] - \frac{\alpha\theta(n-1)}{1-\alpha} log\frac{\beta}{t}$$

$$H_\alpha(X_{1:n};t) \geq \theta log(\frac{\beta}{t}) - \frac{1}{1-\alpha} log[\alpha(n-1)+1] + \frac{\alpha}{1-\alpha} logn - log(\frac{\theta}{\beta})$$

where $\alpha > 1$.

By noting that if F has increasing failure rate (IFR) then $F_{k:n}$ will also be IFR and equation (5), the following theorem can be proved.

Theorem 3 *Let X be IFR, then it is shown that*

$$H_\alpha(X_{k:n};t) \leq -log\lambda_{F_{k:n}}(t) - \frac{log\alpha}{1-\alpha}; \alpha > 0, \alpha \neq 1$$

Where $\lambda_{F_{k:n}(t)}$ *is the failure rate of* $X_{k:n}$ *and can be written as follows*

$$\lambda_{F_{k:n}}(t) = \frac{1}{\bar{B}_{F(t)}(k, n-k+1)} f(t) F^{k-1}(t) \bar{F}^{n-k}(t).$$

Example 4 Let X have Weibull distribution with distribution function

$$F(x) = 1 - e^{-(\lambda x)^{\theta}} \quad for\ x \geq 0,\ where\ \lambda, \theta > 0.$$

It is well known that this distribution is IFR for $\theta \geq 1$, then for this range of θ and the sample minimum we have

$$H_{\alpha}(X_{k:n}; t) \leq -log(n\theta) - (2\theta - 1) log\lambda + \frac{log\alpha}{\alpha - 1}.$$

3 Residual Rényi information of order statistics

Lemma 2 *Let* $U_{k:n}$ *denote* k*th order statistic from* $U \sim U(0,1)$. *The residual Rényi information between* $U_{k:n}$ *and* U *is as follows*

$$K_{\alpha}(U_{k:n}, U; t) = \frac{1}{1-\alpha} log \frac{B(k, n-k+1)}{\bar{B}_t(\alpha(k-1)+1, \alpha(n-k)+1)} - log\bar{B}_t(k, n-k+1) + log(1-t)$$

Theorem 4 *Suppose that* $U \sim U(0,1)$ *and* X *is a nonnegative continuous random variable with density function* f *and distribution function* F. *Then we can show that the residual Rényi information between the* k*th order statistic and the underlying distribution is related to Rényi information between* $U_{k:n}$ *and* U *as follows*

$$K_{\alpha}(X_{k:n}, X; t) = K_{\alpha}(U_{k:n}, U; F(t))$$

Example 5 Let $X_1, ..., X_n$ be independent and identically distributed random variables showing the lifetime of n components connected in a parallel system. Then the lifetime of the system is $X_{n:n} =$

$max(X_1, ..., X_n)$. The residual Rényi information between the distribution of the system $f_{n:n}$ and the parent distribution f, is given by:

$$K_\alpha(X_{n:n}, X; t) = \frac{\alpha}{\alpha-1} log n - \frac{1}{\alpha-1} log[(n-1)\alpha + 1] + \frac{1}{\alpha-1} log[\frac{\bar{F}^{(n-1)\alpha+1}(t)}{\bar{F}^{1-\alpha}(t)(1-F^n(t))^\alpha}]$$

provided that $\alpha < 1$. This shows that although the Rényi information between the distribution of the kth order statistics and the underlying distribution is distribution free however the corresponding dynamic Rényi information is model dependent.

Lemma 3 *Let $U_1, ..., U_n$ be a random sample from uniform distribution over the unit interval. Then residual Rényi information between consecutive order statistics can be written as follows*

$$K_\alpha(X_{k+1;n}, X_{k:n}; t) = -log\frac{\bar{B}_t(k+1, n-k)}{\bar{B}_t(k, n-k+1)} + \frac{1}{1-\alpha} log\frac{B(k+1, n-k)}{\bar{B}_t(k+\alpha, n-\alpha-k+1)}$$

Theorem 5 *Let X be a nonnegative random variable with density function f and distribution function F. If $X_{k:n}$ is* k*th order statistic from this distribution then the residual Rényi information between consecutive order statistics is*

$$K_\alpha(X_{k+1:n}, X_{k:n}; t) = K_\alpha(U_{k+1:n}, U_{k:n}; F(t))$$

References

[1] Asadi, M., Ebrahimi, N. and Soofi, E.S. (2005). Dynamic generalized information measures. Statistics and Probability Letters, 71, 85-98.

[2] Baratpur, S., Ahmadi, J. and Arghami, N. R. (2008). Some characterization based on Rényi entropy of order statistics and record values. Comm. Statist. Theo. Meth., 138, 2544-2551.

[3] Ebrahimi, N., Soofi, E.S. and Zahedi, H. (2004). Information properties of order statistics and spacings. IEEE Transactions on Information Theory, IT50, 177-183.

[4] Mahmoudi, M. and Asadi, M. (2008). More on the residual and past entropies of order α. Submitted for publication.

[5] Morales, D., Pardo, L. and Vajda, I. (1997). Some new statistics for testing hypotheses in parametric models. Journal of Multivariate Analysis, 62, 137-168.

[6] Nadarjah, S. and Zografos, K. (2003). Formulas for Renyi information and related measures for univariate distributions. Inform. Sci., 155, 119-138.

[7] Renyi, A. (1961). On measures of entropy and information. Proceeding of the Fourth Berkeley Symposium, 1, 547-561, Berkeley: UC Press.

[8] Song, K. (2001). Renyi information, loglikelihood and an intrinsic distribution measure. Journal of Statistical Planning and Inference, 93, 51-69.

[9] Shannon, C.E. (1948), A mathematical theory of communication. Bell System Tech. J., 27, 379-423.

مشاوران علمی کنفرانس

کمیته علمی کنفرانس، مقالات را به هفت بخش زیر جهت بررسی و داوری تقسیم نمود:

۱- آمار نظری به مسئولیت آقای دکتر نادر نعمت الهی
۲- آمار کاربردی به مسئولیت آقای دکتر هوشنگ طالبی
۳- احتمال و کاربرد آن به مسئولیت آقای دکتر ناصررضا ارقامی
۴- آمار حیاتی به مسئولیت آقای دکتر بهرام سلیمانی
۵- آمار رسمی به مسئولیت آقای دکتر عباس گرامی
۶- آموزش آمار به مسئولیت آقای دکتر علی رجالی
۷- سایر علوم به مسئولیت آقای دکتر ایرج کاظمی

این زیربخش‌ها به تأیید کمیته علمی کنفرانس از نظرات ارزشمند نامبردگان زیر که به ترتیب الفبا مرتب شده‌اند به عنوان مشاوران علمی کنفرانس در داوری مقالات سود برده اند. بدینوسیله مراتب سپاس و قدردانی برگزارکنندگان کنفرانس ازهمه آنها اعلام می‌گردد.

آقایان: دکتر حسینعلی آذرنوش، دکتر ناصررضا ارقامی، دکتر جعفر احمدی، دکتر مجید اسدی، دکتر فرزاد اسکندری، دکتر محمد امینی، دکتر نصراله ایران‌پناه، غلامرضا ایزدی، دکترابوالقاسم بزرگنیا، دکتر محمد بهرامی، دکتر احمد پارسیان، دکتر سعید پولادساز، دکتر جباری خامنه، دکتر محمد خنجری، دکتر مجید جعفری خالدی، دکتر ابوالفضل خاوری‌نژاد، دکتر خدایی، دکتر منوچهر خردمندنیا، دکتر سلیمان خیری، مهدی دوست پرست، دکتر حسن دوستی، دکتر مصطفی رزمخواه، دکتر فرید روحانی، دکتر حسن زائری، دکتر بهرام سلیمانی، دکتر غلامحسین شاهکار، دکتر محمد صالحی، دکتر هوشنگ طالبی، دکتر محمدحسین علامت‌ساز، دکتر مهدی عمادی، دکتر وحید فکور، دکتر سقراط فقیه‌زاده، دکتر حمید قربانی، دکتر ایرج کاظمی، دکتر عباس گرامی، دکتر مجتبی گنجعلی، دکتر محرابی، دکتر غلامرضا محتشمی، دکتر عادل محمدپور، دکتر نادر نعمت‌الهی، دکتر حمیدرضا نواب‌پور، دکتر حسینعلی نیرومند، دکتر محمدقاسم وحیدی اصل.

و **خانم‌ها:** دکتر سیمیندخت براتپور، دکتر ماه‌بانو تاتا، دکتر زهره شیشه‌بر، دکتر ملیحه عباس نژاد، دکتر طاهره یعقوبی.

اسامی اعضای کمیته های علمی و برگزاری نهمین کنفرانس آمار ایران

اعضای کمیته برگزاری

۱. دکتر مجید اسدی (دبیر کنفرانس)	دانشگاه اصفهان
۲. آقای نصراله ایران‌پناه	دانشگاه اصفهان
۳. دکتر محمد بهرامی (مسئول اجرای مسابقه دانشجویی)	دانشگاه اصفهان
٤. دکتر افشین پرورده (مسئول گارگاه های آموزشی)	دانشگاه اصفهان
٥. دکتر منوچهر خردمند نیا (مسئول انتشارات)	دانشگاه اصفهان
٦. دکتر آرش شاهین	دانشگاه اصفهان
۷. دکتر هوشنگ طالبی (مسئول هماهنگی مدعوین خارجی)	دانشگاه اصفهان
۸. دکتر سید محمود طاهری	دانشگاه صنعتی اصفهان
۹. دکتر محمد حسین علامت ساز	دانشگاه اصفهان
۱۰. دکتر ایرج کاظمی (مسئول دبیرخانه کنفرانس)	دانشگاه اصفهان
۱۱. آقای احمد نورالله (مسئول اسکان)	دانشگاه اصفهان

اعضای کمیته علمی

۱. دکتر ناصر رضا ارقامی	دانشگاه فردوسی مشهد
۲. دکتر مجید اسدی	دانشگاه اصفهان
۳. دکتر بهرام سلیمانی	دانشگاه علوم پزشکی اصفهان
٤. دکتر هوشنگ طالبی	دانشگاه اصفهان
٥. دکتر ایرج کاظمی	دانشگاه اصفهان
٦. دکتر محمد حسین علامت ساز (دبیر علمی کنفرانس)	دانشگاه اصفهان
۷. دکتر عباس گرامی	دانشگاه تهران
۸. دکتر نادر نعمت الهی	دانشگاه علامه طباطبایی

پیشگفتار

مجموعه حاضر حاوی مقالات انگلیسی نهمین کنفرانس آمار ایران است که با همت و تلاش اعضای کمیته‌های علمی و برگزاری، مشاوران علمی کنفرانس، کارکنان و تعدادی از دانشجویان گروه آمار از یکسو و مقالات ارزشمند مدعوین و شرکت‌کنندگان گرامی داخل و خارج از کشور از سویی دیگر تهیه شده و هم‌اکنون در اختیار شما قرار دارد.

جلد دیگری شامل مقالات فارسی نیز به همین صورت تهیه شده و در اختیار شما قرار داده شده است. علاوه بر مقالات مدعوین کنفرانس که بدون داوری پذیرش و چاپ شده‌اند، مقالات این دو جلد پس از طی مراحل داوری از میان ۱۸۵ مقاله که برای سخنرانی گزینش شده بوده‌اند برای چاپ پذیرفته شده‌اند. در هر جلد، به طور مستقل، مقالات بر حسب حروف الفبای نویسنده اول آنها مرتب شده است.

تمام تلاش برگزارکنندگان کنفرانس بر این بوده است که مجموعه مقالات کنفرانس قبل از برگزاری آماده و در کنفرانس توزیع گردد. با توجه به محدودیت زمان و کندی مراحل داوری، اصلاحات پیشنهادی داوران برای مقالات پذیرفته شده به جز در موارد جزیی و ویرایش‌های صوری امکان‌پذیر نبوده است. لذا مسئولیت محتوی مقالات بر عهده نویسندگان آنها می‌باشد.

در اینجا لازم است از مسئولین محترم ذیربط انجمن آمار ایران، پژوهشکده آمار ایران، بانک مرکزی و استانداری‌اصفهان که کنفرانس با مشارکت آنها برگزار شده است قدردانی شود. علاوه برآن لازم می‌دانیم از هیأت رئیسه محترم دانشگاههای اصفهان و علوم پزشکی اصفهان که بدون مساعدت و همکاری آنها برگزاری کنفرانس امکان‌پذیر نبود تشکر و قدردانی نماییم.

همچنین از آقای احسان کریم‌زاده که در آماده‌سازی به موقع این مجموعه از هیچ تلاشی فروگذاری نکرده‌اند قدردانی می‌کنیم.

در پایان از قطب داده های ترتیبی و فضایی دانشگاه فردوسی مشهد، سازمان میراث فرهنگی، صنایع دستی و گردشگری و سازمان صدا و سیمای اصفهان به خاطر حمایت ازکنفرانس قدردانی می شود.

برگزارکنندگان کنفرانس

به نام خدا

مجموعه مقالات

(مقالات انگلیسی)

نهمین کنفرانس آمار ایران

۳۰مرداد الی ۱ شهریور۱۳۸۷
دانشگاه اصفهان
اصفهان، ایران