Revista Colombiana de Estadística

Volumen 36. Número 2 - diciembre - 2013 **ISSN 0120 - 1751**

UNIVERSIDAD
NACIONAL
DE COLOMBIA
SEDE BOGOTÁ
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA

Revista Colombiana de Estadística

http://www.estadistica.unal.edu.co/revista
http://es.wikipedia.org/wiki/Revista_Colombiana_de_Estadistica
http://www.emis.de/journals/RCE/
revcoles_fcbog@unal.edu.co

Indexada en: Ulrichsweb, Scopus, Science Citation Index Expanded (SCIE), Web of Science (WoS), SciELO Colombia, Current Index to Statistics, Mathematical Reviews (MathSci), Zentralblatt Für Mathematik, Redalyc, Latindex, Publindex (A_1)

La *Revista Colombiana de Estadística* es una publicación semestral del Departamento de Estadística de la Universidad Nacional de Colombia, sede Bogotá, orientada a difundir conocimientos, resultados, aplicaciones e historia de la estadística. La Revista contempla también la publicación de trabajos sobre la enseñanza de la estadística.

Se invita a los editores de publicaciones periódicas similares a establecer convenios de canje o intercambio.

Dirección Postal:
Revista Colombiana de Estadística
© Universidad Nacional de Colombia
Facultad de Ciencias
Departamento de Estadística
Carrera 30 No. 45-03
Bogotá – Colombia
Tel: 57-1-3165000 ext. 13231
Fax: 57-1-3165327

Adquisiciones:
Punto de venta, Facultad de Ciencias, Bogotá.
Suscripciones:
revcoles_fcbog@unal.edu.co
Solicitud de artículos:
Se pueden solicitar al Editor por correo físico o electrónico; los más recientes se pueden obtener en formato PDF desde la página Web.

Edición en LaTeX: Patricia Chávez R. E-mail: apchavezr@gmail.com

Impresión: Editorial Universidad Nacional de Colombia, Tel. 57-1-3165000 Ext. 19645, Bogotá.

Revista Colombiana de Estadística	Bogotá	Vol. 36	Nº 1
ISSN 0120 - 1751	COLOMBIA	diciembre-2013	Págs. 193-359

Contenido

Editorial

Leonardo Trujillo[a]

Department of Statistics, Universidad Nacional de Colombia, Bogotá, Colombia

Welcome to the second issue of the 36th volume of Revista Colombiana de Estadistica (Colombian Journal of Statistics). We are glad to announce the new Impact Factor (IF) 2012 of our Journal (0.109) from the Journal Citations Report, which almost doubled the previous one from 2011 (0.059). This IF was launched to the community at the end of June this year just after the publication of the last issue. The impact factor is calculated according to the average number of citations received by its published papers in a time window generally one year. This implies our published papers have now more visibility and hopefully have increased quality in order to be cited by other alternative publications (more information at http://wokinfo.com/essays/impact-factor/).

We have kept, as in recent issues, the characteristic of being a Journal entirely published in English language as part of the requirements of being the winners (second year in a row) of an Internal Grant at the National University of Colombia among other many Journals (see editorial of December 2011). We are also very proud to announce that the Colombian Journal of Statistics have maintained its categorization as an A1 Journal by Publindex (Colciencias) which ranges the journals in the country, being A1 the maximum category. Thanks to all the Editorial and Scientific Committees and Patricia Chavez, our assistant in the Journal, as this is a result of the continuous help obtained from all of them. More information available at http://201.234.78.173:8084/publindex/EnIbnPublindex/resultados.do

The Colombian Journal of Statistics invites researchers to submit scientific papers for the Special Issue in Current Topics in Statistical Graphics to be published in December, 2014. This special issue has the purpose of bringing together current advances and uses of well-known and novel graphical methods from different research areas so that the reader finds potential applications to his/her own research field. The special issue aims at adding an extra value to the special issues published in the RCE by publishing in English language all manuscripts accepted. If you are interested in submitting a paper, please contact the Guest Editor via e-mail (fernando.marmolejoramos@adelaide.edu.au), and visit our website for more information. We welcome applications to problems in engineering, manufacturing process, chemical industry, physical sciences, social sciences, and agricultural industries.

The topics in this current issue range over diverse areas of statistics: two papers in Mixture Distributions by Montoya, Figueroa and Puksic and Salinas, Martinez

[a]Editor in Chief
E-mail: ltrujilloo@unal.edu.co

and Moreno; two papers in Multivariate Analysis by Gupta, Johnson and Nagar and Joarder, Omar and Gupta; one paper in Bayesian Analysis by Moala, Ramos and Achcar; one paper in Fuzzy Data by Pak, Parham, Saraj; one paper in Mixed Models by Ferreira, Ferreira, Nunes and Mexia; one paper in Regression Analysis by Turkan and Toktamis; one paper in Survey Sampling by Hussain, Shah, Shabbir and Riaz and one paper in Time Series Analysis by Garcia-Hiernaux. The International Year of Statistics (Statistics2013) is coming to an end. This initiative was promoted around the globe by the International Statistical Institute (ISI, http://www.statistics2013.org/) in order to celebrate the 300th anniversary of Ars Conjectandi, the book on combinatorics and probability written by Jakob Bernoulli and published in 1713. This work developed essential ideas in statistics such as the Law of Large Numbers and the Bernoulli distribution.

The XIII CLAPEM (Latin American Congress of Probability and Mathematical Statistics) keeps growing in the number of participant institutions for its organization. This event will be held for the first time in Colombia at the city of Cartagena with the help of the Latin American Chapter of the Bernoulli Society. CLAPEM is the largest conference gathering scientists in the particular areas of Probability and Mathematical Statistics in the region and takes place every two/three years. The CLAPEM activities include lectures held by invited researchers, satellite meetings, sessions of oral and poster contributions, short courses, and thematic sessions. Three courses have been confirmed by Allison Etheridge (University of Oxford, UK), Bin Yu (Berkeley University, USA) and Paul Embrechts (ETH Zurich, Switzerland). The plenary talks are in charge of Carenne Ludeña (IVIC, Venezuela), Gerard Biau (Universidad Pierre and Marie Curie, France), Roberto Imbuzeiro Oliveira (IMPA, Brazil), Sourav Chatterjee (New York University, USA), Thomas Mikosch (University of Copenhagen, Denmark) and Victor Rivero (CIMAT, Mexico). If you are interested you can also get more details at www.clapem.unal.edu.co, with Ricardo Fraiman (fraimanricardo@gmail.com) or with Leonardo Trujillo (ltrujilloo@unal.edu.co).

Luis Escobar, member of the Editorial Committee of our Journal, has received the 2013 William G. Hunter Award. Professor Escobar works at the Department of Experimental Statistics at Louisiana State University and he has been a continuous collaborator with the National University of Colombia and CIMAT in Mexico. He is the author of many publications including the classic book of Meeker and Escobar (1998), Statistical Methods for Reliability Data, John Wiley and Sons. The Hunter Award was established in 1987 by the Statistics Division of the American Society for Quality (ASQ) in order to promote, encourage and acknowledge outstanding accomplishments during a career in the broad field of applied statistics. Professor Escobar is Colombian and he has got a PhD from Iowa State University. He was also the Associated Editor of Technometrics during fifteen (15) active years, Vicepresident of the Interamerican Statistical Institute (IASI) from 2006 to 2010 and President from 2010 to 2012. We want to congratulate Professor Escobar for this achievement in his career.

Editorial

Leonardo Trujillo[a]

Departamento de Estadística, Universidad Nacional de Colombia, Bogotá, Colombia

Bienvenidos al segundo número del volumen 36 de la Revista Colombiana de Estadistica. Estamos complacidos en anunciar el nuevo factor de impacto (FI) 2012 de nuestra revista (0.109) de acuerdo al Journal Citations Report, el cual casi ha doblado el factor de impacto anterior en 2011 (0.059). Este FI fue dado a conocer al público a finales de junio de este año justo después de la publicación del primer número. El factor de impacto es calculado de acuerdo al número promedio de citaciones recibidas por los artículos publicados en una ventana de tiempo de generalmente un año. Esto implica que nuestros artículos tienen ahora una mayor visibilidad y esperamos también una mayor calidad para que sean citados en otras publicaciones alternas (más información en http://wokinfo.com/essays/impact-factor/). Hemos mantenido, como en números recientes, la característica de ser una revista publicada completamente en inglés como parte de los requisitos por ser los ganadores (en dos años consecutivos) de una convocatoria interna en la Universidad Nacional de Colombia entre otras muchas revistas (ver editorial de diciembre 2011). También estamos orgullosos de anunciar que la Revista Colombiana de Estadística ha mantenido su categoría A1 de Publindex (Colciencias) la cual reconoce la calidad de las revistas científicas del país, siendo A1 la máxima categoría. Gracias a todos los miembros de los Comités Científico y Editorial y en particular a Patricia Chávez, nuestra asistente, puesto que este es un resultado del trabajo en equipo. Más información se encuentra disponible en http://201.234.78.173:8084/publindex/EnIbnPublindex/resultados.do

La Revista Colombiana de Estadística invita a investigadores que deseen someter artículos para el Número Especial titulado "Current Topics in Statistical Graphics" a ser publicado en diciembre, 2014. Este número especial tiene como propósito el dar a conocer avances recientes de métodos gráficos en diferentes áreas de investigación. El número especial será publicado completamente en inglés. Si Ud. está interesado en someter un artículo, por favor contactar al Editor Invitado vía e-mail (fernando.marmolejoramos@adelaide.edu.au) y visitar nuestra página web con el fin de encontrar más información al respecto. Estaremos atentos a recibir aplicaciones para problemas de ciencias sociales, ingeniería, física, la agricultura, la industria, química, entre otros.

Los temas de este segundo número varían sobre áreas muy diversas de la estadística: dos artículos en Análisis Multivariado de Gupta, Johnson, Nagar y de Joarder,

[a]Editor General
E-mail: ltrujilloo@unal.edu.co

Omar, Gupta; dos artículos en Mezcla de Distribuciones de Montoya, Figueroa, Puksic y de Salinas, Martínez and Moreno; un artículo en Análisis Bayesiano de Moala, Ramos and Achcar; un artículo en Análisis de Regresión de Turkan y Toktamis; un artículo en Datos Difusos de Pak, Parham, Saraj; un artículo en Modelos Mixtos de Ferreira, Ferreira, Nunes, Mexia; un artículo en Muestreo de Hussain, Shah, Shabbir y Riaz y un artículo en Series de Tiempo de García - Hiernaux. El Año Internacional de la Estadística (Statistics2013) está llegando a su fin. Esta iniciativa fue promovida en todo el mundo por el International Statistical Institute (ISI, http://www.statistics2013.org/) con el fin de celebrar el aniversario 300 de la obra Ars Conjectandi, un libro en combinatoria y probabilidad escrito por Jakob Bernoulli y publicado en 1713. Este trabajo fue la base de ideas esenciales en estadística como la Ley de los Grandes Números y la Distribución Bernoulli.

El XIII CLAPEM (Congreso Latinoamericano de Probabilidad y Estadística Matemática) se mantiene en crecimiento en cuanto al número de instituciones participantes en su organización. Este congreso será organizado por primera vez en Colombia en la ciudad de Cartagena con el apoyo de la Sociedad Bernoulli. El CLAPEM es la conferencia más importante de la región que reúne cada dos años a científicos interesados en las áreas de la Probabilidad y la Estadística Matemática. Las actividades del CLAPEM incluyen conferencias invitadas, contribuciones orales, cursos cortos, posters, reuniones satélite y sesiones temáticas. Tres cursos han sido confirmados a cargo de Allison Etheridge (University of Oxford, UK), Bin Yu (Berkeley University, USA) y Paul Embrechts (ETH Zurich, Switzerland). Las conferencias plenarias estarán a cargo de Carenne Ludeña (IVIC, Venezuela), Gerard Biau (Universidad Pierre and Marie Curie, Francia), Roberto Imbuzeiro Oliveira (IMPA, Brasil), Sourav Chatterjee (New York University, USA), Thomas Mikosch (Universidad de Copenhage, Dinamarca) y Victor Rivero (CIMAT, Mexico). Si Ud. está interesado puede obtener más información de este Congreso en www.clapem.unal.edu.co, con Ricardo Fraiman (fraimanricardo@gmail.com) o Leonardo Trujillo (ltrujilloo@unal.edu.co).

Luis Escobar, miembro de nuestro Comité Editorial, ha recibido el Premio William G. Hunter 2013. El Profesor Escobar trabaja en el Departamento de Estadística Experimental en Louisiana State University y ha sido un continuo colaborador con la Universidad Nacional de Colombia y el CIMAT en México. El es el autor de muchas publicaciones incluido el libro clásico de Meeker and Escobar (1998), Statistical Methods for Reliability Data, John Wiley and Sons. El Premio Hunter fue establecido en 1987 por la División de Estadísticas de la American Society for Quality (ASQ) con el fin de promover, motivar y reconocer resultados de gran relevancia en la carrera de un investigador en el área de la estadística aplicada. El Profesor Escobar es de nacionalidad colombiana y terminó sus estudios de Doctorado en Iowa State University. Tambien, fue Editor Asociado de Technometrics durante quince (15) años activos, Vicepresidente del Interamerican Statistical Institute (IASI) de 2006 a 2010 y Presidente de 2010 a 2012. Felicitaciones al profesor Escobar por este logro en su carrera.

Revista Colombiana de Estadística
Diciembre 2013, volumen 36, no. 2, pp. 193 a 208

Profile Likelihood Estimation of the Vulnerability $P(X > v)$ and the Mixing Proportion p Parameters in the Gumbel Mixture Model

Estimación de verosimilitud perfil de los parámetros de vulnerabilidad $P(X > v)$ y proporción de mezcla p en el modelo Gumbel de mezclas

José A. Montoya[1,a], Gudelia Figueroa[1,b], Nuša Pukšič[2,c]

[1]Departamento de Matemáticas, División de Ciencias Exactas y Naturales, Universidad de Sonora, Hermosillo, México

[2]Institute of Metals and Technology, Ljubljana, Slovenia

Abstract

This paper concerns to the problem of making inferences about the vulnerability $\theta = P(X > v)$ and the mixing proportion p parameters, when the random variable X is distributed as a mixture of two Gumbel distributions and v is a known fixed value. A profile likelihood approach is proposed for the estimation of these parameters. This approach is a powerful though simple method for separately estimating a parameter of interest in the presence of unknown nuisance parameters. Inferences about θ, p or (θ, p) are given in terms of profile likelihood regions and can be easily obtained on a computer. This methodology is illustrated through a real problem where the main purpose is to model the size of non-metallic inclusions in steel.

***Key words*:** Invariance principle, Likelihood approach, Likelihood region, Mixture of distributions.

Resumen

En este artículo consideramos el problema de hacer inferencias sobre el parámetro de vulnerabilidad $\theta = P(X > v)$ y la proporción de mezcla p cuando X es una variable aleatoria cuya distribución es una mezcla de dos distribuciones Gumbel y v es un valor fijo y conocido. Se propone el enfoque de verosimilitud perfil para estimar estos parámetros, el cual es un método simple, pero poderoso, para estimar por separado un parámetro de interés en presencia de parámetros de estorbo desconocidos. Las inferencias sobre θ, p o (θ, p) se presentan por medio de regiones de verosimilitud perfil y se

[a]Professor. E-mail: montoya@mat.uson.mx
[b]Professor. E-mail: gfiguero@gauss.mat.uson.mx
[c]Research assistant. E-mail: nusa.puksic@imt.si

pueden obtener fácilmente en una computadora. Esta metodología se ilustra mediante un problema real donde se modela el tamaño de inclusiones no metálicas en el acero.

***Palabras clave*:** enfoque de verosimilitud, mezcla de distribuciones, principio de invarianza, región de verosimilitud.

1. Introduction

Facilities such a electric power, water supplies, communications and transportation are a part of what is named society infrastructure, although in a broad definition this also includes some basic societal functions like education, national defense and financial and health systems. On the other hand, the term critical infrastructure is often used to denote the collection of all large technical systems characterized as public, like electric power, water supply systems, transportation, communications and health systems. All these services are considered a part of a nation critical infrastructure and they are essential for the quality of everyday life. Natural disasters, adverse weather conditions, technical failures, human errors, labor conflicts, sabotage, terrorism and many other situations can disturb the appropriate flow of these services and a severe strain on the society could occur. Hence, national security is directly linked to the vulnerability of critical infrastructure, and problems related with human error or technical failures should be prevented. In particular, it is known that steel inclusions formed during the steel production process degrade the mechanical properties of the steel. Special interest is focused on the control of non-metallic inclusions due to their harmful effect, because their size, amount and chemical composition have a great influence on steel properties and are linked to its vulnerability. Actually, big inclusions can turn out to be dangerous, leading to the failure of the finished steel product. The steel industry fixes some critical limits for these inclusions and those limits depend on the purpose of the steel products. The increasing demand for cleaner steels has led to the continuous improvement of steelmaking practices and modeling the type and distribution of these inclusions has become significant concern in the steel industry.

Murray & Grubesic (2007) define vulnerability of an infrastructure system as the probability that at least one disturbance with negative societal consequence X, could be larger than some (critical) value v, during a given period of time T. Hence, they argued that a simple measure for the vulnerability of an infrastructure system can be formulated as

$$P(X > v) = 1 - F(v)$$

where $F(x)$ denotes the probability distribution function of the random variable X. Skewed distributions such as the exponential, lognormal, log-logistic, and power law distributions have been considered by many authors in a number of different real life situations, like Rosas-Casals, Valverde & Solé (2007) and also by Murray & Grubesic (2007). However, mixture models would be preferable when the random variable X is generated from k distinct random processes. To our knowledge, only

few authors like Zheng (2007) and Barrera-Núñez, Meléndez-Frigola & Herraiz-Jaramillo (2008) have used mixture models to explain vulnerability analysis.

In statistics, a mixture model is a probabilistic model adequate for representing the presence of subpopulations within an overall population, and it is not required for the observed data-set to identify the sub-population to which an individual observation belongs. Formally, given a finite set of probability density functions $f_1(x), \ldots, f_k(x)$ and weights $p_1, \ldots, p_k$ where $p_i \geq 0$ and $\sum p_i = 1$, the density associated with a mixture distribution can be written as $f(x) = \sum p_i f_i(x)$. Mixture distributions arise in a natural way in many areas such as engineering science, medicine, biology, hidrology, geology, as shown in Titterington, Smith & Makov (1985) as well as in Lindsay (1995).

The Gumbel distribution occurs as the limit of maxima of most standard distributions, particularly for the normal distribution. Kotz & Nadarajah (2000) describe in detail this distribution. Actually, the Gumbel distribution has been one of the models used for quantifying the risk associated with extreme rainfalls; it has been also used to model flood levels, the magnitude of earthquakes and even sport records. Some recent applications belong to the engineering area, such as in risk-based engineering, software reliability and structural engineering. Mixture models for Gumbel distributions are of special importance. For example, Chen, Huang & Zhong-Xian (1995) have used a Gumbel mixture model to estimate the seismic risk of the Chinese mainland. Beretta & Murakami (2001) found that a Gumbel mixture model is useful for modeling two types of steel inclusions.

Maximum likelihood estimation for the shape and scale parameters can be found in Evans, Hastings & Peacock (1993) and Johnson, Kotz & Balakrishnan (1994), and parameter estimation for the mixture of two Gumbel distributions is included by Raynal & Guevara (1997), Tartaglia, Caporali, Cavigli & Moro (2005) and Ahmad, Jaheen & Modhesh (2010). However, inferences about the vulnerability parameter $\theta = P(X > v)$ for the Gumbel mixture case has not been carefully studied, despite the actual importance of this kind of analysis. In many applications, inferences about the parameters θ and p, where p is the mixture proportion, can be more relevant than inferences concerning some other model parameters. This will be illustrated with a real data set related to the size of non-metallic inclusions in steel. This data set has two kinds of inclusions, classified as Type 1 and Type 2 inclusions, where p denotes the proportion for the first type of inclusion.

Let the distribution of X be a mixture of two independent Gumbels:

$$f(x; \mu_1, \sigma_1, \mu_2, \sigma_2, p) = p f_1(x; \mu_1, \sigma_1) + (1-p) f_2(x; \mu_2, \sigma_2) \tag{1}$$

where

$$f_i(x; \mu_i, \sigma_i) = \frac{1}{\sigma_i} \exp\left[-\left(\frac{x-\mu_i}{\sigma_i}\right)\right] \exp\left\{-\exp\left[-\left(\frac{x-\mu_i}{\sigma_i}\right)\right]\right\}$$

$-\infty < \mu_i < \infty$, $\sigma_i > 0$, $i = 1, 2$, $-\infty < x < \infty$, and $0 < p < 1$. A fundamental statistical problem is concerned with making inferences on the vulnerability parameter

$$\theta = P(X > v; \mu_1, \sigma_1, \mu_2, \sigma_2, p) = 1 - \int_{-\infty}^{v} f(x; \mu_1, \sigma_1, \mu_2, \sigma_2, p)\, dx \qquad (2)$$

and the mixing proportion p, based on a sample $x_1, \ldots, x_n$ from X. In this paper, we analyze this problem considering $\sigma_1 > 0$, $\sigma_2 > 0$, $0 < p < 1$, $-\infty < \mu_1 < \mu_2 < v$, and v a known fixed value. Our purpose is to estimate these parameters using the profile likelihood function. The profile likelihood approach can be useful in many situations and it is a powerful though simple method for separately estimating a parameter of interest in the presence of unknown nuisance parameters. Inferences about θ, p or (θ, p) can be given in terms of profile likelihood regions which are easily obtained with a computer. This methodology will be illustrated with a real data set concerning the size distribution of non-metallic inclusions in steel.

2. Profile Likelihood Approach

In this section we describe the estimation procedure that will be used to make inferences about the parameters of interest. This approach is based in Sprott (1980), Kalbfleisch (1985), and Sprott (2000). Let $\boldsymbol{x}_o = (x_1, \ldots, x_n)$ be an observed sample from a distribution with likelihood function $L(\boldsymbol{\psi}, \boldsymbol{\lambda}; \boldsymbol{x}_o)$, where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_{d_\psi})$ represents the parameter of interest and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{d_\lambda})$ is a nuisance parameter. The profile likelihood function of $\boldsymbol{\psi}$ is

$$L_P(\boldsymbol{\psi}; \boldsymbol{x}_o) = L\big[\boldsymbol{\psi}, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\psi}); \boldsymbol{x}_o\big] \qquad (3)$$

The quantity $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\psi})$ that maximizes $L(\boldsymbol{\psi}, \boldsymbol{\lambda}; \boldsymbol{x}_o)$ for a specified value of $\boldsymbol{\psi}$, is called the restricted maximum likelihood estimate of the nuisance parameter $\boldsymbol{\lambda}$.

Usually $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\psi})$ exists and is unique for each value of $\boldsymbol{\psi}$, so the definition (3) applies. Formally, $L_P(\boldsymbol{\psi}; \boldsymbol{x}_o)$ can be defined as

$$L_P(\boldsymbol{\psi}; \boldsymbol{x}_o) = \sup_{\boldsymbol{\lambda}} L(\boldsymbol{\psi}, \boldsymbol{\lambda}; \boldsymbol{x}_o) \qquad (4)$$

Since $L(\boldsymbol{\psi}, \boldsymbol{\lambda}; \boldsymbol{x}_o)$ is proportional to the probability of the observed sample as a function of the parameters of the model, then the supremum exists and it is finite. The profile likelihood function can be used to rank parameter values according to their plausibilities. Now, the relative profile likelihood function of $\boldsymbol{\psi}$ is a standardized version of (4), and takes a value of one at the maximum of the profile likelihood function of $\boldsymbol{\psi}$,

$$R_P(\boldsymbol{\psi}; \boldsymbol{x}_o) = \frac{L_P(\boldsymbol{\psi}; \boldsymbol{x}_o)}{\sup_{\boldsymbol{\psi}} L_P(\boldsymbol{\psi}; \boldsymbol{x}_o)}$$

Hence, the relative profile likelihood varies between 0 and 1. Values of $\boldsymbol{\psi}$ that are supported by the observed sample $\boldsymbol{x}_o$ will result in values of $R_P(\boldsymbol{\psi}; \boldsymbol{x}_o)$ close to one. In contrast, values of $\boldsymbol{\psi}$ with $R_P(\boldsymbol{\psi}; \boldsymbol{x}_o)$ close to zero become implausible, given the sample $\boldsymbol{x}_o$. Morever, if the the maximum likelihood estimate (mle) of

$(\boldsymbol{\psi}, \boldsymbol{\lambda})$ exists and is unique, then the relative profile likelihood function of $\boldsymbol{\psi}$ can be defined as

$$R_P\left(\boldsymbol{\psi}; \boldsymbol{x}_o\right) = \frac{L_P(\boldsymbol{\psi}; \boldsymbol{x}_o)}{L(\widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\lambda}}; \boldsymbol{x}_o)} = \frac{L\big[\boldsymbol{\psi}, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\psi}); \boldsymbol{x}_o\big]}{L(\widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\lambda}}; \boldsymbol{x}_o)}$$

where $\widehat{\boldsymbol{\psi}}$ is the mle of $\boldsymbol{\psi}$. Note that $\widehat{\boldsymbol{\lambda}} = \widehat{\boldsymbol{\lambda}}\big(\widehat{\boldsymbol{\psi}}\big)$ is the ordinary mle of $\boldsymbol{\lambda}$. The relative profile likelihood function is the maximum relative likelihood that ψ can attain when $\boldsymbol{\lambda}$ is unknown and free to vary arbitrarily. Thus, $R_P(\boldsymbol{\psi}; x_o)$ ranks all possible values of $\boldsymbol{\psi}$ according to their maximum plausibilities and supported by the observed data.

A level c profile likelihood region for $\boldsymbol{\psi}$ is given by

$$\mathcal{R}_P(c) = \{\boldsymbol{\psi} : R_P(\boldsymbol{\psi}; x_o) \geq c\} \tag{5}$$

where $0 \leq c \leq 1$. When $\boldsymbol{\psi}$ is a scalar this region will be an interval if R_P is unimodal and the union of disjoint intervals when R_P is multimodal. Each specific value of $\boldsymbol{\psi}$ within this region has an associated relative profile likelihood $R_P(\boldsymbol{\psi}; \boldsymbol{x}_o) \geq c$, and values outside this region will have a relative profile likelihood $R_P(\boldsymbol{\psi}; \boldsymbol{x}_o) < c$. At level c, this region separates plausible values of $\boldsymbol{\psi}$ from the implausible ones. When varying c from 0 to 1, a complete set of nested likelihood regions is obtained and these converges to the mle $\widehat{\boldsymbol{\psi}}$ as $c \to 1$. Computer algorithms are usually used to find the mle or the borders of a profile likelihood regions given in (5).

In most of the cases, a profile likelihood region $\mathcal{R}_P(c)$ is an approximate confidence region for $\boldsymbol{\psi}$, so it is called a likelihood-confidence region, or a likelihood-confidence interval when $\boldsymbol{\psi}$ is a scalar. Under the null hypothesis $H_0 : \boldsymbol{\psi} = \boldsymbol{\psi}_0$ the likelihood ratio statistic $-2\ln\left[R_P(\boldsymbol{\psi}_0; x)\right]$ usually converge, in distribution, to a chi-squared distribution with $d_{\boldsymbol{\psi}}$ degrees of freedom (Serfling 1980). When this is true, the set $\mathcal{R}_P(c)$ becomes a $100\,(1-\alpha)\,\%$ confidence region for $\boldsymbol{\psi}$, where $c = \exp(-\chi^2_{d_{\boldsymbol{\psi}},1-\alpha}/2)$ and $\chi^2_{d_{\boldsymbol{\psi}},1-\alpha}$ represents the quantile of probability $1-\alpha$ of a chi-squared distribution with $d_{\boldsymbol{\psi}}$ degrees of freedom. For example, if $d_{\boldsymbol{\psi}} = 1$, $\boldsymbol{\psi}$ is a scalar parameter, then the profile likelihood region at level $c = 0.15$ becomes a confidence region for $\boldsymbol{\psi}$, with an approximate 95% confidence level. In a similar way, if $d_{\boldsymbol{\psi}} = 2$, the level $c = 0.05$ profile likelihood region for $\boldsymbol{\psi}$ will be a confidence region with an approximate 95% confidence level.

Some authors like Montoya, Díaz-Francés & Sprott (2009) and Figueroa (2012) suggest to include the precision of the measuring instrument to avoid unbounded likelihoods, which usually occurs when the continuous approximation to the likelihood function is used and regularity conditions are not satisfied. The unboundedness and also the flatness of a profile likelihood function have been used to propose alternative approaches to estimate nuisance parameters, like in Smith & Naylor (1987) and Green, Roesch, Smith & Strawderman (1994), who criticized the profile likelihood function for being flat and uninformative, overlooking that it can be indicative that a simpler (limiting) model might be a good alternative to explain the data (Cheng & Iles 1990). Although here there is no problem of unbounded likelihoods and to our knowledge, a flat profile likelihood can also be obtained even when including the precision of the measuring instrument, there are some others

circumstances where incorporating this information is reasonable; for example, when the instrument measures with different precision or when it produces many repeated observations, like in the example included in Section 4.

3. Inferences about θ and p Using the Profile Likelihood in the Gumbel Mixture Model

Let X be a random variable from a two-component Gumbel mixture model with density function $f(x;\mu_1,\sigma_1,\mu_2,\sigma_2,p)$ given in (1), where $\sigma_1 > 0$, $\sigma_2 > 0$, $0 < p < 1$, $-\infty < \mu_1 < \mu_2 < v$ and v is a known fixed value. In this case, the parameters of interest are the vulnerability parameter $\theta = P(X > v)$ and the mixing proportion p. Although the parametrization of the Gumbel mixture model involves the five unknown parameters μ_1, σ_1, μ_2, σ_2, and p, the parameter θ has been left out. In order to make profile likelihood inferences about θ and p, it is convenient to use a one to one reparametrization in such a way that θ becomes one of the new parameters and p is included as well. Hence, the vulnerability parameter θ can be written, explicitly, as a function of μ_1, σ_1, μ_2, σ_2 and p,

$$\begin{aligned}\theta &= P(X > v;\mu_1,\sigma_1,\mu_2,\sigma_2,p)\\ &= 1 - P(X \le v;\mu_1,\sigma_1,\mu_2,\sigma_2,p)\\ &= 1 - \left[p\Phi_G\left(\delta_1\right) + (1-p)\,\Phi_G\left(\delta_2\right)\right]\end{aligned}$$

where $\Phi_G(\cdot)$ is the standard Gumbel distribution and $\delta_i = (v-\mu_i)/\sigma_i$, $i = 1, 2$. Here, δ_i is introduced for algebraic and computational simplicity. Note that $\delta_i > 0$ when $\sigma_i > 0$ and $-\infty < \mu_1 < \mu_2 < v$.

3.1. Reparametrizations

Let $\sigma_i = (v-\mu_i)/\delta_i$, $i = 1, 2$. This produces the one to one parametrization $(\mu_1,\sigma_1,\mu_2,\sigma_2,p) \leftrightarrow (\mu_1,\delta_1,\mu_2,\delta_2,p)$ with a Jacobian

$$J_1 = \frac{(v-\mu_1)(v-\mu_2)}{\sigma_1^2\sigma_2^2} > 0$$

The Gumbel mixture model can be reparametrized in terms of $(\mu_1,\delta_1,\mu_2,\delta_2,p)$ when substituting $\sigma_i = (v-\mu_i)/\delta_i$, $i = 1, 2$,

$$f^*(x;\mu_1,\delta_1,\mu_2,\delta_2,p) = pf_1^*(x;\mu_1,\delta_1) + (1-p)f_2^*(x;\mu_2,\delta_2) \tag{6}$$

where

$$f_i^*(x;\mu_i,\delta_i) = \frac{\delta_i}{v-\mu_i}\exp\left[-\delta_i\left(\frac{x-\mu_i}{v-\mu_i}\right)\right]\exp\left\{-\exp\left[-\delta_i\left(\frac{x-\mu_i}{v-\mu_i}\right)\right]\right\}$$

with $-\infty < \mu_1 < \mu_2 < v$, $0 < p < 1$ and $\delta_i > 0$, $i = 1, 2$.

The parameter δ_1 can now be written as

$$\delta_1 = \delta_1(\theta, p, \delta_2) = \Phi_G^{-1}\left[\left(\frac{1-\theta}{p}\right) - \left(\frac{1-p}{p}\right)\Phi_G(\delta_2)\right] \tag{7}$$

where $\Phi_G^{-1}(\cdot)$ denotes the inverse of the standard Gumbel distribution. Again, this produces the one to one parametrization $(\mu_1, \delta_1, \mu_2, \delta_2, p) \leftrightarrow (\mu_1, \theta, \mu_2, \delta_2, p)$ with a Jacobian given by

$$J_2 = p\exp(-\delta_1)\,\Phi_G(\delta_1) > 0$$

Thus, the Gumbel mixture model (6) can be reparametrized in terms of $(\theta, p, \mu_1, \mu_2, \delta_2)$ by substituting in (6) the expression $\delta_1(\theta, p, \delta_2)$ given in (7). Here, θ and p are the parameters of interest and the remaining ones are nuisance parameters.

3.2. Likelihood

In this section a likelihood function for the parameters of the reparametrized mixture model in terms of the parameters of interest θ and p, and the vector of nuisance parameters $\boldsymbol{\lambda} = (\mu_1, \mu_2, \delta_2)$ is presented. This likelihood includes the precision of the measuring instrument because it could provide valuable information that should be included into the analysis. As Lindsey (1999) explains to include the precision of the measuring instrument into the analysis requires no additional computational effort nowadays.

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with density function given in (6) and $\boldsymbol{x}_o = (x_1, \ldots, x_n)$ its observed sample. Since all measuring instruments have finite precision, that is, data can only be recorded to a finite number of decimals, then $\boldsymbol{x}_o$ must *always* be discrete. Thus the observation $X_i = x_i$ can be interpreted as $x_i - h/2 \leq X_i \leq x_i + h/2$, where h is the precision of the measuring instrument, and so is a fixed positive number, as is described in Sprott (2000, p. 10), Montoya, Díaz-Francés, & Sprott (2009) and Figueroa (2012). Therefore, for $\boldsymbol{x}_o = (x_1, \ldots, x_n)$, the resulting likelihood function of $(\theta, p, \boldsymbol{\lambda})$ is proportional to the probability of the observed sample,

$$\begin{aligned} L(\theta, p, \boldsymbol{\lambda}; \boldsymbol{x}_o) &\propto \prod_{i=1}^{n} \int_{x_i - h/2}^{x_i + h/2} f^*\left[x_i; \mu_1, \delta_1(\theta, p, \delta_2), \mu_2, \delta_2, p\right] \\ &= \prod_{i=1}^{n} \left\{ p\left[F_1\left(x_i + \frac{h}{2}; \theta, p, \mu_1, \delta_2\right) - F_1\left(x_i - \frac{h}{2}; \theta, p, \mu_1, \delta_2\right)\right] + \right. \\ &\quad \left. (1-p)\left[F_2\left(x_i + \frac{h}{2}; \mu_2, \delta_2\right) - F_2\left(x_i - \frac{h}{2}; \mu_2, \delta_2\right)\right]\right\} \end{aligned}$$

where

$$\begin{aligned} F_1(z; \theta, p, \mu_1, \delta_2) &= \exp\left\{-\exp\left[-\delta_1(\theta, p, \delta_2)\left(\frac{z-\mu_1}{v-\mu_1}\right)\right]\right\} \\ &= \Phi_G\left\{\Phi_G^{-1}\left[\left(\frac{1-\theta}{p}\right) - \left(\frac{1-p}{p}\right)\Phi_G(\delta_2)\right]\left(\frac{z-\mu_1}{v-\mu_1}\right)\right\} \end{aligned}$$

and

$$F_2(z;\mu_2,\delta_2) \;=\; \exp\left\{-\exp\left[-\delta_2\left(\frac{z-\mu_2}{v-\mu_2}\right)\right]\right\} = \Phi_G\left[\delta_2\left(\frac{z-\mu_2}{v-\mu_2}\right)\right]$$

with $0<\theta<1$, $0<p<1$, $\delta_2>0$, $0\le[(1-\theta)/p]-[(1-p)/p]\,\Phi_G(\delta_2)\le 1$, $-\infty<\mu_1<\mu_2<v$ and v is a known fixed value.

Since the mle $(\widehat{\theta},\widehat{p},\widehat{\boldsymbol{\lambda}})$ cannot be obtained analytically, it must be calculated numerically. For computational convenience, maximization of the log-likelihood function of the original parametrization can be used to obtain $(\widehat{\mu}_1,\widehat{\sigma}_1,\widehat{\mu}_2,\widehat{\sigma}_2,\widehat{p})$ the maximum likelihood estimators of $(\mu_1,\sigma_1,\mu_2,\sigma_2,p)$, and by the invariance property of the likelihood function, the mle of θ is

$$\widehat{\theta} \;=\; 1-\left[\widehat{p}\Phi_G(\widehat{\delta}_1)+(1-\widehat{p})\Phi_G(\widehat{\delta}_2)\right]$$

where $\widehat{\delta}_i=(v-\widehat{\mu}_i)/\widehat{\sigma}_i$.

3.3. Profile likelihood

Profile likelihood inference about the vector (θ,p) in the presence of the vector of nuisance parameters $\boldsymbol{\lambda}=(\mu_1,\mu_2,\delta_2)$ is based on the relative profile likelihood function of (θ,p). The profile likelihood function of (θ,p) is

$$L_P(\theta,p;\boldsymbol{x}_o)=\sup_{\lambda} L(\theta,p,\boldsymbol{\lambda};x_o)$$

and its associated relative profile likelihood function is

$$R_P(\theta,p;\boldsymbol{x}_o)=\frac{L_P(\theta,p;\boldsymbol{x}_o)}{\sup\limits_{\theta,p} L_P(\theta,p;\boldsymbol{x}_o)}=\frac{L_P(\theta,p;\boldsymbol{x}_o)}{L(\widehat{\theta},\widehat{p},\widehat{\lambda};\boldsymbol{x}_o)} \tag{8}$$

where $(\widehat{\theta},\widehat{p},\widehat{\boldsymbol{\lambda}})$ is the mle of (θ,p,λ). Note that $R_P(\theta,p;\boldsymbol{x}_o)$ will be a surface sitting above the parameter space $(0,1)\times(0,1)$ and its maximum value 1 occurs at $(\theta,p)=(\widehat{\theta},\widehat{p})$. A convenient way to visualize $R_P(\theta,p;\boldsymbol{x}_o)$ in two dimensions is by plotting contours of level c, obtained by solving $R_P(\theta,p;\boldsymbol{x}_o)=c$. The regions obtained from this contour plot are likelihood-based confidence regions for (θ,p). For instance, if $c=0.05$ the region delimited by the contour plot is a likelihood-based confidence region for (θ,p), with an approximate 95% confidence level.

On the other hand, profile likelihood inferences about the scalar parameter θ are based on the relative profile likelihood function of θ. In this case the profile likelihood and relative profile likelihood functions are

$$\begin{aligned} L_P(\theta;\boldsymbol{x}_o) &= \sup_{p,\lambda} L(\theta,p,\boldsymbol{\lambda};\boldsymbol{x}_o)=\sup_{p} L_P(\theta,p;\boldsymbol{x}_o) \\ R_P(\theta;\boldsymbol{x}_o) &= \frac{L_P(\theta;\boldsymbol{x}_o)}{\sup\limits_{\theta} L_P(\theta;\boldsymbol{x}_o)}=\frac{\sup\limits_{p} L_P(\theta,p;\boldsymbol{x}_o)}{\sup\limits_{\theta,p} L_P(\theta,p;\boldsymbol{x}_o)} \end{aligned} \tag{9}$$

Similarly, the profile likelihood and the relative profile likelihood for parameter p are given by

$$\begin{aligned} L_P(p;\boldsymbol{x}_o) &= \sup_{\theta,\lambda} L(\theta,p,\lambda;\boldsymbol{x}_o) = \sup_{\theta} L(\theta,p;\boldsymbol{x}_o) \\ R_P(p;\boldsymbol{x}_o) &= \frac{L_P(p;\boldsymbol{x}_o)}{\sup\limits_{p} L_P(p;\boldsymbol{x}_o)} = \frac{\sup\limits_{\theta} L(\theta,p;\boldsymbol{x}_o)}{\sup\limits_{\theta,p} L_P(\theta,p;\boldsymbol{x}_o)} \end{aligned} \tag{10}$$

3.4. Computational implementation

Relative profile likelihoods given in (8), (9) and (10) can be obtained through the computation of $L_P(\theta,p;\boldsymbol{x}_o)$, which can be easily implemented using the R 'stats' package function constrOptim, for each specified value of (θ,p). A feasible region is defined for $AB - C \geq 0$, where

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{bmatrix},\ B = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \delta_2 \end{bmatrix},\ \text{and } C = \begin{bmatrix} 0 \\ -v \\ 0 \\ \Phi_G^{-1}\left(\dfrac{1-\theta-p}{1-p}\right) \\ -K \end{bmatrix}$$

K is an upper bound for user-defined δ_2, serving as a tuning value. Based on that, the following algorithm is used to display contour lines and profiles of $L_P(\theta,p;x_o)$.

1. Create a full grid from two monotonically increasing grid vectors: $\theta \in S_1 \subset (0,1)$ and $p \in S_2 \subset (0,1)$.
2. Map these points to the function $L_P(\theta,p;\boldsymbol{x}_o)$ and store the values in a matrix $M[\theta,p]$.
3. Calculate the matrix $R[\theta,p] = M[\theta,p]/L^*$, the vector $R[\theta] = \underset{p}{max}\ M[\theta,p]/L^*$ and the vector $R[p] = \underset{\theta}{max}\ M[\theta,p]/L^*$, where $L^* = \underset{\theta,p}{max}\ M[\theta,p]$.
4. Contour plots can be created from θ, p, $R[\theta,p]$ coordinates using the contour function included in the R 'graphics' package. The profile plot can be obtained by plotting θ versus $R[\theta]$ (or p versus $R[p]$).

Note that $R[\theta]$ and $R[p]$ can be used to obtain profile likelihood intervals for θ and p, respectively.

4. Illustrative Example

Quality of steel is strongly influenced by the presence of non-metallic inclusions. Although inferences about the size of large inclusions have largely been

based on the assumption that inclusions are all of a single type and methods of classic extreme value statistics are appropriate, as shown in Murakami (1994), the need to analyze for the presence of multiple inclusions has been noticed by Lund, Johansson & Olund (1998) through experiments in the bearing industry. In particular, Beretta & Murakami (2001) have found that a Gumbel mixture model can be appropriate when studying only two types of inclusions.

The methodology proposed in this paper is implemented in a set of data obtained from an experiment conducted at the Institute of Metals and Technology, Ljubljana, Slovenia and concerns modeling the size distribution of non-metallic inclusions in steel, where mainly two types of inclusions are investigated: (a) Type 1, these are soft elongated inclusions composed mainly of manganese sulfide and (b) Type 2, hard round inclusions comprising mainly aluminum oxides. These two types of inclusions can be distinguished by their shape when seen under a microscope, and their composition has been confirmed by spectroscopic analyses. Round inclusions are much more harmful and sometimes can lead to premature failure of the steel piece.

As a part of this experiment, a standard inspection area S_0 of 0.27 mm^2 is defined. The area of the maximum inclusion in S_0, defined as $A_{\max}$, is measured in $n = 544$ sample areas, from a single steel slab. All the inclusions were measured using automatic image analysis and only those with a cross-section area larger than 3 μm^2 were considered real inclusions in this analysis. Cross-section areas smaller than 3 μm^2 are not clearly visible by light microscopy at 100x magnification; it is not only the matter of sufficient resolution, but it is difficult to verify what may be inclusions and what could be artifacts from image contrast adjustments. It is worthwhile to mention that none of the $A_{\max}$ measurements was smaller than 3 μm^2.

The square root of the measured area $x = \sqrt{A_{\max}}$ is taken; this is called the inclusion size and these are the measurements used in the statistical analysis. The minimum observation is $x_{(1)} = 1.9339$, the maximum $x_{(544)} = 20.8835$, the sample mean $\bar{x} = 7.3184$, the sample standard deviation $s_x = 2.2158$, and quantiles 25, 50 and 75 are $Q_{25} = 5.8583$, $Q_{50} = 7.1091$, $Q_{0.75} = 8.5621$, respectively. The data set formed with these 544 inclusion sizes has many repeated values. This suggests that the precision of the measuring instrument should be included into the analysis. Actually, this set has only 185 different values, some of them repeated even eight times. Now, since the size of each inclusion is not directly measured by the instrument, the precision associated with each of these values must be computed through error propagation techniques. This implies that each measurement has an associated precision h_i, that should be considered in the analysis.

A Gumbel mixture model is proposed to study the size of non-metallic inclusions in steel and the adequacy of this model can be seen in Figures 1 and 2. Although an observation in Figure 2 is apparently an outlier, it turns out to be a possible observation, when the cloud formed by the Q-Q plots of many simulated samples with the estimated model includes the points of the Q-Q plot of the observed sample; this cloud can be seen in Figure 3. The parameter estimates

for the Gumbel mixture model given in (1) were obtained by maximum likelihood estimation and are shown in Table 1.

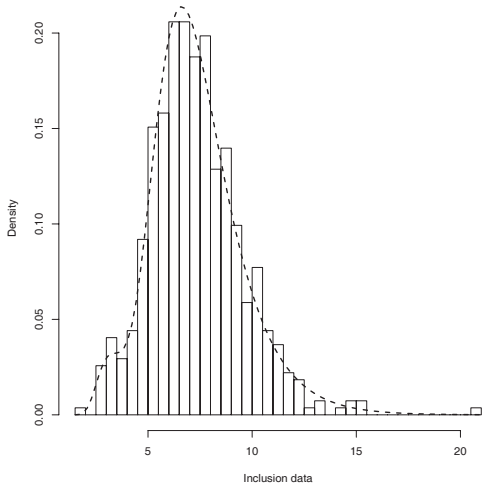


FIGURE 1: Histogram and estimated density for non-metallic inclusion sizes.

$\widehat{\mu}_1 = 3.1943$	$\widehat{\sigma}_1 = 0.7195$	$\widehat{\mu}_2 = 6.6251$	$\widehat{\sigma}_2 = 1.6229$	$\widehat{p} = 0.0597$

TABLE 1: Parameter estimates

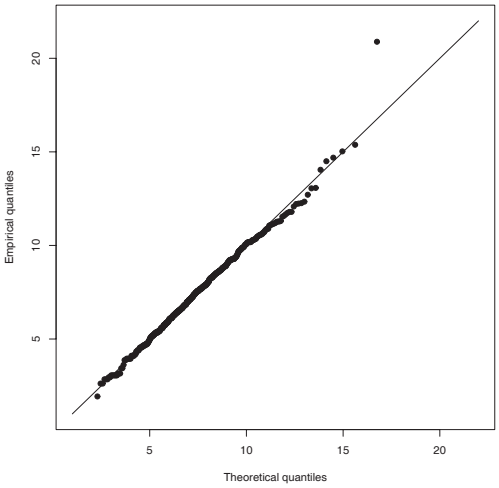


FIGURE 2: Q-Q plot: Theoretical versus Empirical quantiles.

The likelihood approach described in Section 3.2 was used to estimate the probability that the maximum size of a non-metallic inclusion could be greater than a maximum allowed inclusion size v. The problem about fixing a critical

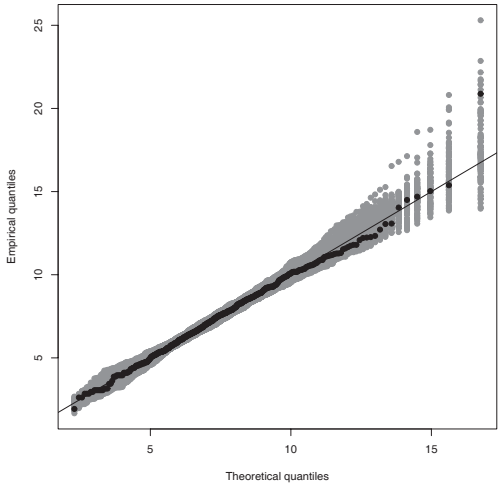


FIGURE 3: Q-Q plot simulation.

size v is that it may differ for different kind of steels, depending on their use or purpose. Here, we fixed $v = 15$ μm, just to show how to model vulnerability.

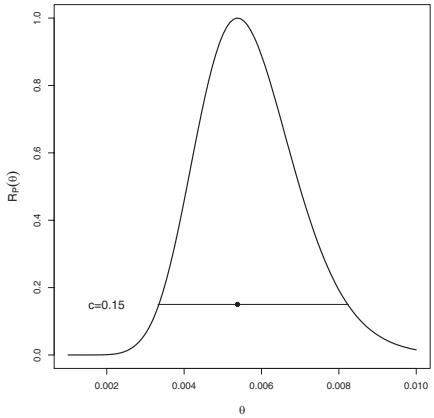


FIGURE 4: Relative profile likelihood of parameter θ.

The profile likelihood function of parameter $\theta = P(X > 15\ \mu\text{m})$ is plotted in Figure 4, and is almost symmetric around its mle $\widehat{\theta} = 0.005380$ which is marked with a dark circle; the 15% profile likelihood interval for θ is $(0.003364, 0.008182)$. It is worthwhile to mention that this information can be used for quality control, because better steel grades can be obtained by reducing its inclusion content. For example, for a plausible value of θ, like $\widehat{\theta} = 0.005380$, a non-metallic inclusion larger than 15, a value merely illustrative, could be associated with a return period of approximately 200 sample areas. This could be low or high depending of the steel and its use. Actually, the information contained in the profile likelihood function of parameter θ is very useful when comparing candidates for the improved steel

grade, since the amount and size of non-metallic inclusions in steel are directly linked with many of its properties.

Besides the importance of having information about parameter θ, it is very informative to characterize parameter p. We are considering two types of non-metallic inclusions and knowledge about the proportion of each of these types is crucial. Here, p denotes the proportion of Type 1 inclusions within the mixture model. Figure 5(a) shows the relative profile likelihood of parameter p, where $\widehat{p} = 0.0597$ is marked with a dark circle and the 15% profile likelihood interval for p results $(0.0294, 0.3422)$. This interval is wide with respect to 1, the length of its parameter space and it does not contain the value $p = 0$. The proportion of Type 1 inclusions can be considered small or large depending on the application of the steel. By the invariance property of the likelihood function, point and interval estimates for Type 2 inclusions can be easily obtained and these are $1-\widehat{p} = 0.9403$ and $(0.6578, 0.9706)$, respectively. The relative profile likelihood for parameter $1 - p$ is shown in Figure 5(b). Plots in Figure 5 show that these likelihoods are strongly asymmetric around their maximum. It is important to mention that Type 2 inclusions are much more harmful since they do not work when the steel is deformed, so they serve as a crack nucleation sites and can lead to premature failure of the steel piece.

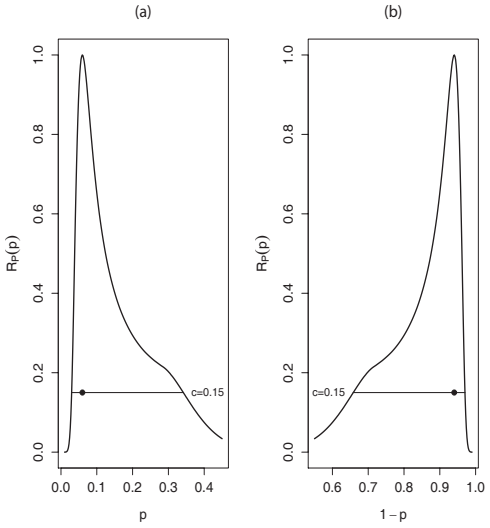


FIGURE 5: Relative profile likelihood of parameter (a) p and (b) $1 - p$.

Inferences about parameters θ and p can be obtained by constructing a likelihood contour plot for these parameters. Figure 6(a) shows the likelihood confidence regions for parameters θ and p at different levels of c. Using the invariance property of the likelihood function, a contour plot for θ and $1-p$ is easily obtained; this is shown in Figure 6(b). This kind of plot allows us to make simultaneous inferences about the proportion of Type 1 or Type 2 inclusions and the probability of exceeding the maximum allowed inclusion size v. These plots can play an important role in the improvement of the steelmaking practices, for example, they

can be used to compare candidates for the improved steel grade through the analysis of parameters θ and p. We consider that the approach used in this statistical analysis adds another dimension to the overall characterization of the steel grade.

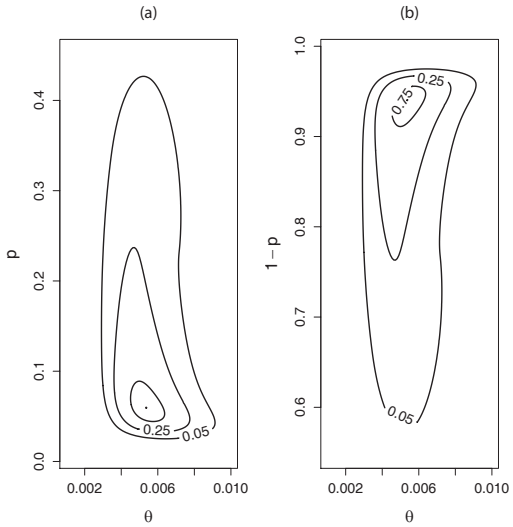


FIGURE 6: Contour plot for parameters (a) θ and p, (b) θ and $1 - p$.

All the parameter inferences were obtained through computational techniques using the R software and the R source code for this example is available upon request.

5. Conclusions

The precision of a measuring instrument turns out to be an important aspect in some applications, like the one included here, where the lack of precision of the measuring instrument caused many repeated observations. The likelihood function allows to include, in a natural way, the precision associated with each of the observations. The profile likelihood function is a simple method devised to handle the estimation of parameters of interest in the presence of unknown nuisance parameters, and it inherits all the information contained in the likelihood function. Estimation statements about parameters $\theta = P(X > v)$, p or (θ, p) in the Gumbel mixture case can be given in terms of profile likelihood confidence regions that were easily obtained through computational techniques. This approach was particularly useful when analyzing the vulnerability of steel.

[Recibido: junio de 2011 — Aceptado: junio de 2013]

References

Ahmad, K. E., Jaheen, Z. F. & Modhesh, A. A. (2010), 'Estimation of a discriminant function based on small sample size from a mixture of two Gumbel distributions', *Communications in Statistics-Simulation and Computation* **39**, 713–725.

Barrera-Núñez, V., Meléndez-Frigola, J. & Herraiz-Jaramillo, S. (2008), A survey on voltage sag events in power systems, *in* 'Transmission and Distribution Conference and Exposition: Latin America, 2008 IEEE/PES', pp. 1–3.

Beretta, S. & Murakami, Y. (2001), 'Largest-extreme-value distribution analysis of multiple inclusion types in determining steel cleanliness', *Metallurgical and Materials Transactions B* **32**, 517–523.

Chen, H. & Huang, Z.-X. (1995), 'Application of Gumbel mixture extreme theory and maximum likelihood to estimate the seismic risk of the Chinese mainland', *Acta Seismologica Sinica* **8**, 325–331.

Cheng, R. C. H. & Iles, T. C. (1990), 'Embedded models in three-parameter distributions and their estimation', *Journal of the Royal Statistical Society. Series B.* **52**(1), 135–149.

Evans, M., Hastings, N. & Peacock, B. (1993), *Statistical Distributions*, John Wiley & Sons, New York.

Figueroa, P. G. (2012), Las funciones de verosimilitud discretizada y restringida perfil en la inferencia científica, Ph.D. Thesis, Universidad de Sonora, Hermosillo, Sonora, México.

Green, E. J., Roesch, F. A. J., Smith, A. F. M. & Strawderman, W. E. (1994), 'Bayesian estimation for the three-parameter Weibull distribution with tree diameter data', *Biometrics* **50**(1), 254–269.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994), *Continuous Univariate Distributions*, Vol. II, John Wiley & Sons, New York.

Kalbfleisch, J. G. (1985), *Probability and Statistical Inference*, Vol. II, Springer-Verlag, New York.

Kotz, S. & Nadarajah, S. (2000), *Extreme Value Distributions. Theory and Applications*, Imperial College Press, London.

Lindsay, B. (1995), *Mixture Models: Theory, Geometry and Applications*, Institute for Mathematical Statistics, Hayward.

Lindsey, J. K. (1999), 'Some statistical heresies', *The Statistician* **48**(1), 1–40.

Lund, T., Johansson, S. & Olund, L. (1998), Nucleation of fatigue in very low oxygen bearing steels, *in* 'Bearing Steels: Into the 21st Century, STP 1327', American Society for Testing and Materials, West Conshohocken, PA, pp. 124–130.

Montoya, J., Díaz-Francés, E. & Sprott, D. A. (2009), 'On a criticism of the profile likelihood function', *Statistical Papers* **50**, 195–202.

Murakami, Y. (1994), 'Inclusion rating by statistics of extreme values and its application to fatigue strength prediction and quality control of materials', *Journal of Research of the National Institute of Standards and Technology* **99**, 345–351.

Murray, A. T. & Grubesic, T. H. (2007), *Critical Infrastructure: Reliability and Vulnerability*, Springer-Verlag, Berlin.

Raynal, J. & Guevara, J. (1997), 'Maximum likelihood estimators for the two populations Gumbel distribution', *Hydrological Science and Technology Journal* **13**, 47–56.

Rosas-Casals, M., Valverde, S. & Solé, R. V. (2007), 'Topological vulnerability of the European power grid under errors and attacks', *International Journal of Bifurcation and Chaos* **17**(7), 2465–2475.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York.

Smith, R. L. & Naylor, J. C. (1987), 'Statistics of the three-parameter Weibull distribution', *Annals of Operations Research* **9**, 577–587.

Sprott, D. A. (1980), 'Maximum likelihood and small samples: Estimation in the presence of nuisance parameters', *Biometrika* **67**, 515–523.

Sprott, D. A. (2000), *Statistical Inference in Science*, Springer-Verlag, New York.

Tartaglia, V., Caporali, E., Cavigli, E. & Moro, A. (2005), 'L-moments based assessment of a mixture model for frequency analysis of rainfall extremes', *Advances in Geosciences* **2**, 331–334.

Titterington, D., Smith, A. & Makov, U. (1985), *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Sons, New York.

Zheng, H. (2007), Investigation of power system blackouts and reliability improvement for power distribution systems, Ph.D. Thesis, The University of Texas, Arlington.

Revista Colombiana de Estadística
Diciembre 2013, volumen 36, no. 2, pp. 211 a 221

The Distribution of a Linear Combination of Two Correlated Chi-Square Variables

Distribución de una Combinación Lineal de Dos Variables Chi-Cuadrado Correlacionadas

Anwar H. Joarder[1,a], M. Hafidz Omar[1,b], Arjun K. Gupta[2,c]

[1]Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

[2]Department of Mathematics and Statistics, Bowling Green State University, Ohio, United States of America

Abstract

The distribution of the linear combination of two chi-square variables is known if the variables are independent. In this paper, we derive the distribution of positive linear combination of two chi-square variables when they are correlated through a bivariate chi-square distribution. Some properties of the distribution, namely, the characteristic function, cumulative distribution function, raw moments, mean centered moments, coefficients of skewness and kurtosis are derived. Results match with the independent case when the variables are uncorrelated. The graph of the density function is presented.

***Key words*:** Bivariate Chi-square Distribution, Correlated Chi-square Variables, Linear Combination, Characteristic Function, Cumulative Distribution, Moments.

Resumen

La distribución de una combinación lineal de dos variables chi cuadrado es conocida si las variables son independientes. En este artículo, se deriva la distribución de una combinación lineal positiva de dos variables chi cuadrado cuando éstas están correlacionadas a través de una distribución chi cuadrado bivariada. Algunas propiedades de esta distribución como la función característica, la función de distribución acumulada, sus momentos, momentos centrados alrededor de la media, los coeficientes de sesgo y curtosis son derivados. Los resultados coinciden con el caso independiente cuando las variables son no correlacionadas. La gráfica de la función de densidad es también presentada.

***Palabras clave*:** combinación lineal, distribución acumulada, distribución chi cuadrado bivariada, función característica, momentos, variables chi cuadrado correlacionadas.

[a]Professor. E-mail: anwarj@kfupm.edu.sa
[b]Associate professor. E-mail: omarmh@kfupm.edu.sa
[c]Professor. E-mail: gupta@bgsu.edu

1. Introduction

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \boldsymbol{X}_n$ be $(N > 2)$ two-dimensional independent normal random vectors from $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with mean vector $\overline{\boldsymbol{X}} = (\overline{X}_1, \overline{X}_2)'$, where $N\overline{X}_i = \sum_{j=1}^{N} X_{ij}$, $(i = 1, 2)$ so that sums of squares and cross product matrix is given by $\sum_{j=1}^{N}(\boldsymbol{X}_j - \overline{\boldsymbol{X}})(\boldsymbol{X}_j - \overline{\boldsymbol{X}})' = \boldsymbol{A}$. Let the matrix $\boldsymbol{A}$ be denoted by $\boldsymbol{A} = (a_{ik}), i = 1, 2;\ k = 1, 2$ where $a_{ii} = mS_i^2, (i = 1, 2), m = N - 1$ and $a_{12} = mRS_1S_2$. That is, S_1 and S_2 are the sample standard deviations based on the bivariate sample, and R is the related product moment correlation coefficient. Also let $\Sigma = (\sigma_{ik})$, $i = 1, 2;$ $k = 1, 2$ where $\sigma_{11} = \sigma_1^2, \sigma_{22} = \sigma_2^2, \sigma_{12} = \rho\sigma_1\sigma_2$ with $\sigma_1 > 0, \sigma_2 > 0$. The quantity $\rho(-1 < \rho < 1)$ is the product moment correlation coefficient between X_{1j} and $X_{2j}(j = 1, 2, \ldots, N)$.

The joint density function $U = mS_1^2/\sigma_1^2$ and $V = mS_2^2/\sigma_2^2$, called the bivariate chi-square distribution, was derived by Joarder (2009) in the spirit of Krishnaiah, Hagis & Steinberg (1963) who studied the bivariate chi-distribution.

The distribution of linear function of random variables is useful in the theory of process capability indices and the study of two or more control variables. See, for example, Glynn & Inglehart (1989) and Chen & Hsu (1995). It also occurs in statistical hypothesis testing and high energy physics (See Bausch 2012).

The density function of positive linear combination of independent chi-square random variables was derived by Gunst & Webster (1973). Algorithms were written by Davies (1980) and Farebrother (1984) for the distribution of the linear combination of independent chi-square variables. The exact density function of a general linear combination of independent chi-square variables is a special case of a paper by Provost (1988) for a more general case of Gamma random variables. Interested readers may go through Johnson, Kotz & Balakrishnan (1994) for a detailed historical account.

By application of the inversion formula to the characteristic function of the sum of correlated chi-squares, Gordon & Ramig (1983) derived an integral form of the cumulative distribution function (CDF) of the sum and the used trapezoidal rule to evaluate it. Since this integral form of the CDF involves integration of complex variables, the percentage points depends on the type of numerical technique you employ. Recently Bausch (2012) has developed an efficient algorithm for numerically computing the linear combination of independent chi-square random variables. He has shown its application in string theory.

In Section 2, some mathematical preliminaries are provided. In Section 3, we derive the density function and the Cumulative Distribution Function of the positive linear combination of two correlated chi-square variables when they are governed through a bivariate chi-square density function given by (6). In Section 4, we derive the characteristic function of the distribution. In Section 5, we also derive some properties of the distribution, namely, raw moments, mean centered moments, coefficient of skewness and kurtosis. The results match with the independent case when the variables are uncorrelated. The results also match with the special case of the sum of two correlated chi-square variables considered by

Joarder & Omar (2013). The graph of the density function of the sum is presented at the end of the paper.

2. Mathematical Preliminaries

Let $f_{X,Y}(x,y)$ be the joint density function of X and Y. Then the following lemma is well known.

Lemma 1. *Let X and Y be two random variables with common probability density function $f_{X,Y}(x,y)$. Further let $Z = X + Y$. Then the density function of Z at z is given by*

$$h_Z(z) = \int_0^\infty f_{X,Y}(z-y,y)dy \tag{1}$$

The duplication of the Gamma function is given below:

$$\Gamma(2z)\sqrt{\pi} = 2^{2z-1}\Gamma(z)\Gamma\left(z+\frac{1}{2}\right) \tag{2}$$

The incomplete Gamma is defined by

$$\gamma(\alpha,x) = \int_0^x t^{\alpha-1}e^{-t}dt \tag{3}$$

where $Re(\alpha) > 0$ (Gradshteyn & Ryzhik 1994, Equation 8.350, p. 949).

The hypergeometric function ${}_pF_q(a_1,a_2,\ldots,a_p;b_1,b_2,\ldots,b_q;z)$ is defined by

$${}_pF_q(a_1,a_2,\ldots,a_p;b_1,b_2,\ldots,b_q;z) = \sum_{k=0}^{\infty} \frac{(a_1)_{\{k\}}(a_2)_{\{k\}}\ldots(a_p)_{\{k\}}}{(b_1)_{\{k\}}(b_2)_{\{k\}}\ldots(b_q)_{\{k\}}}\frac{z^k}{k!} \tag{4}$$

where $a_{\{k\}} = a(a+1)\ldots(a+k-1)$

The following integral will be used:

$$\int_0^\infty x^{a-1}e^{-bx}\gamma(c,dx)dx = \frac{d^c\Gamma(a+c)}{c(b+d)^{a+c}}\,{}_2F_1\left(1,a+c;c+1;\frac{d}{b+d}\right) \tag{5}$$

with $Re(a+b) > 0, b > 0, (a+c) > 0$, (Gradshteyn & Ryzhik 1994).

The following theorem is due to Joarder (2009), although it can be followed from Krishnaiah et al. (1963).

Theorem 1. *The random variables U and V are said to have a correlated bivariate chi-square distribution each with $m(>2)$ degrees of freedom, if its density function is given by*

$$f_{U,V}(u,v) = \frac{(uv)^{(m/2)-1}}{2^m\Gamma^2(m/2)(1-\rho^2)^{m/2}}\exp\left(-\frac{u+v}{2-2\rho^2}\right){}_0F_1\left(\frac{m}{2};\frac{\rho^2uv}{(2-2\rho^2)^2}\right) \tag{6}$$

where ${}_0F_1(;b;z)$ is defined in 4 and $-1 < \rho < 1$.

In case $\rho = 0$, the density function of the joint probability distribution in Theorem 1, would be $f_{U,V}(u,v) = f_U(u)f_v(v)$ where $U \sim X^2_m$ and $V \sim X^2_m$. The product moment correlation coefficient between U and V can be calculated to be ρ^2. For the estimation of correlation coefficient ρ by modern techniques, we refer to Ahmed (1992).

3. The Density Function and the Cumulative Distribution Function

Let c_1 and c_2 be positive numbers so that $T_1 = c_1U + c_2V$. Equivalently, let $T_1 = c_1T$ where $T = U + cV, c = c_2/c_1$ defines a general linear combination of the variables U and V.

Theorem 2. *Let U and V be two chi-square variables each having m degrees of freedom with density function given in Theorem 1. Then for any positive constant c, the density function of $T = U + cV$ is given by*

$$f_T(t) = \frac{\Gamma((m+1)/2)t^{m-1}}{2^m\Gamma(m)[c(1-\rho^2)]^{m/2}} \exp\left(-\frac{t}{2-2\rho^2}\right)$$
$$\times \sum_{k=0}^{\infty} \frac{1}{\Gamma(k+(m+1)/2)} \frac{(t\rho)^{2k}}{(4-4\rho^2)^{2k}c^k k!} {}_1F_1\left(k+\frac{m}{2}; 2k+m; \frac{(c-1)t}{(2-2\rho^2)c}\right) \quad (7)$$

where $m > 2, -1 < \rho < 1$ and $0 \le t < \infty$.

Proof. It follows from (6) that the joint density function of $X = U$ and $Y = cV$ is given by

$$f_{X,Y}(x,y) = \frac{(1-\rho^2)^{-m/2}}{2^m\Gamma^2(m/2)} \left(\frac{xy}{c}\right)^{(m/2)-1}$$
$$\exp\left(-\frac{1}{2-2\rho^2}\left(x+\frac{y}{c}\right)\right) {}_0F_1\left(\frac{m}{2}; \frac{\rho^2}{(2-2\rho^2)^2}\frac{xy}{c}\right)\frac{1}{c}$$

so that, by Lemma 1, the density function of $T = X + Y$ is given by

$$f_T(t) = \frac{(c(1-\rho^2))^{-m/2}}{2^m\Gamma^2(m/2)} \exp\left(-\frac{t}{2-2\rho^2}\right) I(t; m, \rho, c) \quad (8)$$

where

$$I(t; m, \rho, c) = \Gamma(m/2) \sum_{k=0}^{\infty} \frac{1}{\Gamma[k+(m/2)]} \frac{\rho^{2k}}{(2-2\rho^2)^{2k}c^k k!} J(t; m, \rho, c) \quad (9)$$

with $J(t; m, \rho, c) = \int_0^t (t-y)^{k-1+(m/2)} y^{k-1+(m/2)} \exp\left(\frac{(c-1)y}{c(2-2\rho^2)}\right) dy$ □

Substituting $y = st$ we have

$$J(t; m, \rho, c) = \Gamma(k + (m/2))t^{2k+m-1} \sum_{j=0}^{\infty} \frac{\Gamma[(k + j + (m/2))]}{\Gamma(2k + j + m)} \frac{(c-1)^j t^j}{(2 - 2\rho^2)^j c^j j!} \quad (10)$$

which, by (4), can be expressed as

$$J(t; m, \rho, c) = t^{2k+m-1} \frac{\Gamma^2(k + (m/2))}{\Gamma(2k + m)} \, {}_1F_1\left(k + \frac{m}{2}; 2k + m; \frac{(c-1)t}{(2 - 2\rho^2)c}\right)$$

Plugging this in (9) and simplifying, we have

$$I(t; m, \rho, c) = \frac{\Gamma(m/2)\sqrt{\pi}}{2^{m-1}} \, t^{m-1} \sum_{k=0}^{\infty} \frac{1}{\Gamma(k + (m+1)/2)} \frac{(t\rho)^{2k}}{(4 - 4\rho^2)^{2k} c^k k!} \times {}_1F_1\left(k + \frac{m}{2}; 2k + m; \frac{(c-1)t}{(2 - 2\rho^2)c}\right)$$

Substituting this in (8) and simplifying, we have (7).

Figure 1 provides a graphical display of this density function for $m = 5$ and various values of c and ρ.

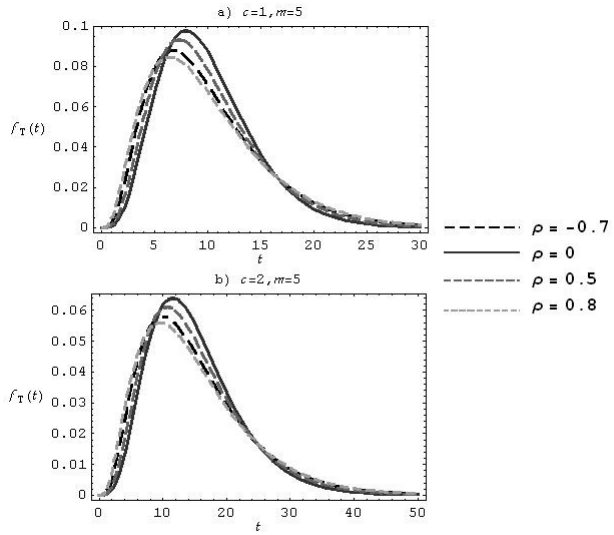


FIGURE 1: Linear combination of chi-square variables for $m = 5$ and various values of ρ.

Theorem 3. *Let T have a density function given by* (7)*. Then the Cumulative Distribution Function of T is given by*

$$F_T(t) = \frac{\Gamma((m+1)/2)}{2^m \Gamma(m)(c(1-\rho^2))^{m/2}} \times \sum_{k=0}^{\infty} \frac{1}{\Gamma(k + (m+1)/2)} \frac{\rho^{2k}}{(4 - 4\rho^2)^{2k} c^k k!} I(k; m, \rho) \quad (11)$$

where $I(k;m,\rho)=\int_{y=0}^{t} y^{m+2k-1}\exp\left(-\frac{y}{(2-2\rho^2)}\right){}_1F_1\left(k+\frac{m}{2};2k+m;\frac{(c-1)y}{(2-2\rho^2)c}\right)dy$, $0\le t<\infty,-1<\rho<1,m>2$ *and* c *is any positive constant.*

Proof. It is immediate from Theorem 2 □

The CDF in (11) is not in closed form. However, if $\rho=0$, a closed form expression is presented in Theorem 5.

Theorem 4. *Let* U *and* V *be two independent chi-square variables each having* $m(>2)$ *degrees of freedom. Then for any positive constant* c*, the density function of* $T=U+cV$ *is given by*

$$f_T(t)=\frac{t^{m-1}e^{-t/2}}{2^m c^{m/2}\Gamma(m)}\,{}_1F_1\left(\frac{m}{2};m;\frac{(c-1)t}{2c}\right),\ 0\le t<\infty \tag{12}$$

Proof. Putting $\rho=0$ in Theorem 2, we have (12). □

If $c=1$, then (12) simplifies to the density function of X^2_{2m} as expected. The equation (10) is a special case of Provost (1988)

Theorem 5. *Let* U *and* V *be two independent chi-square variables each having* $m(>2)$ *degrees of freedom. Then the Cumulative Density Function of* $T=U+cV$ *is given by*

$$F(t)=\frac{1}{c^{m/2}}\sum_{k=0}^{\infty}\frac{(m/2)_{\{k\}}}{\Gamma(k+m)}\frac{(c-1)^k}{c^k k!}\gamma(k+m,t/2) \tag{13}$$

where $m>2$ *and* $\gamma(\alpha,x)$ *is defined in* (3)*.*

Proof. By substituting $\rho=0$ in (12), we have

$$F(t)=\frac{1}{2^m\Gamma(m)c^{m/2}}\int_0^t y^{m-1}\exp\left(-y/2\right){}_1F_1\left(\frac{m}{2};\frac{(c-1)y}{2c}\right)dy$$

which simplifies to (13). □

By substituting $c=1$ in (13), we have $F(t)=\gamma(m,t/2)/\Gamma(m)$ which is the Cumulative Distribution Function X^2_{2m}. Bausch (2012) developed and efficient algorithm for computing linear combination of independent chi-square variables.

4. The Characteristic Function

The quantity i in this section is defined by the imaginary number $i=\sqrt{-1}$.

Theorem 6. *Let* U *and* V *be two chi-square variables each having* $m(>2)$ *degrees of freedom* $-1<\rho<1$ *with density function given in Theorem 1. Then the characteristic function* $\phi_{U,V}(w_1,w_2)=E(e^{iw_1U+iw_2V})$ *of* U *and* V *at* w_1 *and* w_2 *is given by*

$$\phi_{U,V}(w_1,w_2)=[(1-2iw_1)(1-2iw_2)+4w_1w_2\rho^2]^{-m/2} \tag{14}$$

where $m>2$ *and* $-1<\rho<1$.

Proof. See Omar & Joarder (2010). □

The characteristic function of the linear combination of two correlated chi-square variables is derived below.

Theorem 7. *Let U and V be two chi-square variables each having m degrees of freedom. Then for any known constant c, the characteristic function of $T = U + cV$ at w is given by the following:*

$$\phi_T(w) = [(1 - 2iw)(1 - 2icw) + 4w^2c\rho^2]^{-m/2} \tag{15}$$

where $m > 2$ and $-1 < \rho < 1$.

Proof. By definition,the characteristic function of $T = U + cV$ is given by $\phi_T(w) = E(e^{iwT}) = E[e^{iw(U+cV)}] = E[e^{i(wU+cwV}]$. □

By (14), $E[e^{i(wU+cwV)}] = \phi_{U,V}(w, cw)$ and can be written as $\phi_{U,V}(w, cw) = [(1 - 2iw)(1 - 2iw) + 4wcw\rho^2]^{-m/2}$, which is (15).

The corollary below follows from Theorem 7.

Corollary 1. *Let U and V be two independent chi-square variables each having the same degrees of freedom m. Then for any positive constant c, the characteristic function of $T = U + cV$ at w is given by the following:*

$$\phi_T(w) = [(1 - 2iw)(1 - 2iwc)]^{-m/2},\ m > 2 \tag{16}$$

Since the above can be expressed as $\phi_T(w) = \phi_U(w)\phi_{cV}(w)$, clearly the random variable T is the linear combination of two independent random variables U and V. In case $c = 1$, the equation (16) will be specialized to the characteristic function of a chi-square variable with $2m$ degrees of freedom.

The following results are for any general bivariate distribution.

Theorem 8. *Let X and Y have a bivariate distribution with density function $f_{X,Y}(x, y)$ and characteristic function $\varphi_{X,Y}(w_1, w_2) = E(e^{iw_1X+iw_2Y})$. Then for any constant c, the characteristic function of $T = X + cY$ at w is given by the following:*

$$\phi_T(w) = \phi_{X,Y}(w, cw) \tag{17}$$

Proof. By definition, the characteristic function of $T = X + cY$ is given by $\phi_T(w) = E(e^{iwT}) = E[e^{iw(X+cY)}] = E[e^{i(wX+cwY)}] = \phi_{X,Y}(w, cw)$. □

Corollary 2. *Let X and Y have a bivariate distribution with density function $f_{X,Y}(x, y)$ and characteristic function $\varphi_{X,Y}(w_1, w_2) = E(e^{iw_1X+iw_2Y})$. Then, the characteristic function of $T = X + Y$ at w is given by the following:*

$$\phi_T(w) = \phi_{X,Y}(w, w) \tag{18}$$

5. Moments, Coefficient of Skewness and Kurtosis

The following theorem is due to Joarder, Laradji, & Omar (2012).

Theorem 9. *Let U and V have the bivariate chi-square distribution with density function with common degrees of freedom m and density function in Theorem 1. Then for $a > -m/2, b > -m/2$ and $-1 < \rho < 1$, the (a, b)-th product moment of U and V, denoted by $\mu'_{a,b;\rho}(U, V) = E(U^a V^b)$, is given by*

$$\mu'_{a,b;\rho}(U,V) = 2^{a+b}(1-\rho^2)^{a+b+(m/2)}\frac{\Gamma(a+(m/2))\Gamma(b+(m/2))}{\Gamma^2(m/2)} \times {}_2F_1\left(a+\frac{m}{2}, b+\frac{m}{2}; \frac{m}{2}; \rho^2\right) \quad (19)$$

where $m > 2, -1 < \rho < 1$ and ${}_2F_1(a_1, a_2; b; z)$ is defined in (4).

Theorem 10. *Let T have a density function given by (7). Then the first four moments of T are respectively given by*

$$E(T) = (c+1)m \quad (20)$$

$$E(T^2) = (c^2+1)m(m+2) + 2cm(m+2\rho^2) \quad (21)$$

$$E(T^3) = (c^3+1)m(m+2)(m+4) + 3c(c+1)(m(m+2)(m+4\rho^2)) \quad (22)$$

$$\begin{aligned} E(T^4) &= (c^4+1)[m(m+2)(m+4)(m+6)] \\ &\quad + 4c(c^2+1)[m(m+2)(m+4)(m+6\rho^2)] \\ &\quad + 6c^2 m(m+2)[m(m+2) + 8(m+2)\rho^2 + 8\rho^4] \end{aligned} \quad (23)$$

where $c > 0, m > 2$ and $-1 < \rho < 1$.

Proof. The moment expressions between (20) and (23) inclusively follow from Theorem 9 by tedious algebraic simplification. □

Let T have a density function given by (7). Then the a-th moment of T denoted by $E(T^a) = E(U + cV)^a$, where c is any non-negative constant, is given by

$$\mu'_a(T) = \sum_{j=0}^{a} \binom{a}{j} c^{a-j} \mu'_{j,a-j;\rho}(U,V) \quad (24)$$

where $\mu'_{j,a-j;\rho}(U, V) = E(U^j V^{a-j})$ is given by Theorem 9.

The centered moments of T of order a is given by $\mu_a = E(T - E(T))^a, a = 1, 2, \ldots$ That is the second, third and fourth order mean corrected moments are respectively given by

$$\mu_2 = E(T^2) - \mu^2 \quad (25)$$

$$\mu_3 = E(T^3) - 3E(T^2)\mu + 2\mu^3 \quad (26)$$

$$\mu_4 = E(T^4) - 4E(T^3)\mu + 6E(T^2)\mu^2 - 3\mu^4 \quad (27)$$

Where $\mu = E(T)$. The explicit forms for the centered moments of the linear combination of bivariate chi-square random variables are given in the following theorem.

Theorem 11. *Let T have a density function given by* (7)*. The second to fourth centered moments of T are given by the following:*

$$\mu_2 = 2m(1 + c^2 + 2c\rho^2) \tag{28}$$

$$\mu_3 = 8(c+1)m(c^2 - c + 1 + 3c\rho^2) \tag{29}$$

$$\begin{aligned}\mu_4 = 12m\big[&2c^2m + (c^4+1)(m+4) \\ &+ 4c(4c^2 + 4c + 4 + c^2m + m)\rho^2 + 4c^2(m+2)\rho^4\big]\end{aligned} \tag{30}$$

where $m > 2, c$ is any positive constant and $-1 < \rho < 1$.

Proof. The moments between (28) to (30) inclusively follow from (25),(26) and (27) with tedious algebraic simplifications. □

In case $\rho = 0$,the moments match with that of $T = U + cV$ where U and V have independent chi-square distributions each with degrees of freedom $m(> 2)$.

The skewness and kurtosis of a random variable T are given by the moment ratios $\alpha_i(T) = \mu_i \mu_2^{-i/2}, i = 3, 4$. The theorem below follows from Theorem 11.

Theorem 12. *Let T have a density function given by* (7)*. The coefficient of skewness and kurtosis of T where c is any non-negative constant, are given by*

$$\alpha_3(T) = \frac{2\sqrt{2}(c+1)(3c\rho^2 + c^2 - c + 1)}{\sqrt{m}(2c\rho^2 + c^2 + 1)^{3/2}} \tag{31}$$

and

$$\alpha_4(T) = 3 + \frac{12}{m(2\rho^2 c + c^2 + 1)^2}(2c^2\rho^4 + 4c(c^2 + c + 1) + c^4 + 1) \tag{32}$$

respectively, where $m > 2, c$ is any positive constant and $-1 < \rho < 1$.

In case $\rho = 0$, the above coefficient of skewness and kurtosis simplifies to, as expected, that for $T = U + cV$ where U and V are independent chi-square with the same degrees of freedom $m(> 2)$. In case $c = 1, \rho$ decreases to 0 and the degrees of freedom m increases indefinitely, then the coefficient of skewness and that of kurtosis converges to 0 and 3 as expected.

6. Conclusion

We have developed the distributional characteristics of linear combination of correlated chi-square variables. Based on the results in the paper, efficient computational algorithms can be developed along the line of Bausch (2012) who developed an efficient algorithm for computing linear combination of independent chi-square variables.

Acknowledgments

The authors are grateful to two anonymous referees and the Editor for many constructive suggestions that led to the current version of the paper. The first two authors gratefully acknowledge the excellent research facility provided by King Fahd University of Petroleum and Minerals, Saudi Arabia especially through the project IN111019.

[Recibido: febrero de 2013 — Aceptado: julio de 2013]

References

Ahmed, S. (1992), 'Large sample pooling procedure for correlation', *The Statistician* **41**, 415–428.

Bausch, J. (2012), On the efficient calculation of a linear combination of chi-square variables with an application in counting string vacua. arXiv:1208.2691.

Chen, S. & Hsu, N. (1995), 'The asymptotic distribution of the process capability index c_{pmk}', *Communications in Statistics - Theory and Methods* **24**, 1279–1291.

Davies, R. (1980), 'Algorithm as 155. The distribution of a linear combination of χ^2 random variables', *Applied Statistics* **29**, 332–339.

Farebrother, R. (1984), 'Algorithm AS 204. The distribution of a positive linear combination of χ^2 random variables', *Applied Statistics* **33**, 332–339.

Glynn, P. & Inglehart, D. (1989), 'The optimal linear combination of control variates in the presence of asymptotically negligible bias', *Naval Research Logistics Quarterly* **36**, 683–692.

Gordon, N. & Ramig, P. (1983), 'Cumulative distribution function of the sum of correlated chi-squared random variables', *Journal of Statistical Computation and Simulation* **17**(1), 1–9.

Gradshteyn, I. & Ryzhik, I. (1994), *Table of Integrals, Series and Products*, Academic Press.

Gunst, R. & Webster, J. (1973), 'Density functions of the bivariate chi-square distribution', *Journal of Statistical Computation and Simulation* **2**, 275–288.

Joarder, A. (2009), 'Moments of the product and ratio of two correlated chi-square random variables', *Statistical Papers* **50**(3), 581–592.

Joarder, A., Laradji, A., & Omar, M. (2012), 'On some characteristics of bivariate chi-square distribution', *Statistics* **46**(5), 577–586.

Joarder, A. & Omar, M. (2013), 'Exact distribution of the sum of two correlated chi-square variables and its application', *Kuwait Journal of Science and Engineering* **40**(2), 60–81.

Johnson, N., Kotz, S. & Balakrishnan, N. (1994), *Continuous Univariate Distributions*, Vol. 1, Wiley.

Krishnaiah, P., Hagis, P. & Steinberg, L. (1963), 'A note on the bivariate chi distribution', *SIAM Review* **5**, 140–144.

Omar, M. & Joarder, A. (2010), Some properties of bivariate chi-square distribution and their application, Technical Report 414, Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Saudi Arabia.

Provost, S. (1988), 'The exact density of a general linear combination of gamma variables', *Metron* **46**, 61–69.

Revista Colombiana de Estadística
Diciembre 2013, volumen 36, no. 2, pp. 223 a 238

Generalized Portmanteau Tests Based on Subspace Methods

Tests de Portmanteau generalizados basados en métodos de subespacios

Alfredo García-Hiernaux[a]

Quantitative Economics Department, Universidad Complutense de Madrid, Spain

Abstract

The problem of diagnostic checking is tackled from the perspective of the subspace methods. Two statistics are presented and their asymptotic distributions are derived under the null hypothesis. The procedures are devised to deal with univariate and multivariate processes, are flexible and able to separately check regular and seasonal correlations. The performance in finite samples of the proposals is illustrated via Monte Carlo simulations and two examples with real data.

***Key words*:** Diagnostic checking, Portmanteau test, Residual autocorrelation, Residuals.

Resumen

Este artículo trata el problema de la diagnosis residual desde la perspectiva de los métodos de subespacios. Se presentan dos estadísticos y sus distribuciones asintóticas bajo la hipótesis nula. Ambos estadísticos pueden usarse con procesos univariantes o multivariantes, son flexibles y permiten contrastar separadamente las correlaciones regulares y estacionales. El comportamiento en muestras finitas de las dos propuestas se ilustra mediante simulaciones de Monte Carlo y dos ejemplos con datos reales.

***Palabras clave*:** autocorrelación residual, diagnosis de residuos, test de Portmanteau, residuos.

1. Introduction

Since the seminal work by Box & Pierce (1970), or the enhanced version by Ljung & Box (1978), many studies have focused in the ability of the statistical

[a]Professor. E-mail: agarciah@ucm.es

tests to determine the adequacy of a model. The procedures suggested in this paper cope with this problem from a novel perspective.

We use a subspace methods-based approach to derive two tests and their asymptotic distributions under the null of zero correlations up to order k. As subspace methods, the procedures are devised to deal with univariate and multivariate processes that leads to a generalization of Ljung & Box (1978) and Hosking (1980) -which is the Ljung-Box multivariate version- statistics, hereafter Q_{LB} and P_H, respectively.

The flexibility of the tests allows use to obtain gains in terms of statistical power and robustness against non-robust competitors as Q_{LB} and P_H. We propose that these gains can improve by tuning a specific matrix that may be modified by the user. Although this is not investigated in this paper, the question is briefly addressed in the conclusion. However, no comparison against robust statistics is performed as ours do not belong to this type of test. Our proposals are also able to separately test seasonal correlations. When applied to seasonal data, our tests present a gain in terms of degrees of freedom with respect to alternatives devised to cope with seasonality, as McLeod (1978) or Ursu & Duchesne (2009), and in terms of statistical power when compared to Q_{LB}. A Monte Carlo study shows that the finite sample properties of one of our tests outperform those of Q_{LB} in terms of nominal size, when the number of lags chosen grows, and in statistical power.

Finally, results in Aoki (1990), Casals, Sotoca & Jerez (1999) and Casals, García-Hiernaux & Jerez (2012) imply that Multiple-Source Error (MSE) state space, Single-Source Error (SSE) state space and VARMAX models are equally general and freely interchangeable. This means that our derivation of the distribution for the residuals of a VARMA model permits to test the adequacy of its equivalent MSE or SSE state space model. Consequently, our procedures can be sequentially used to determine the system order in a state space model (since the null hypothesis can always be written as residuals with system order equal to zero) which is a critical decision in the subspace methods literature and applied data modeling.

The plan of the paper is as follows. Section 2 presents previous results in subspace methods that will be used later. Some distributional results and the two tests proposed are derived in Sections 3 and 4, respectively. Lastly, Section 5 compares the performance of our proposals with Ljung-Box and Hoskings' tests using Monte Carlo experiments and two applications to real data.

To express the results precisely, we introduce the following notation which will be use throughout the paper: $\xrightarrow{d}$ means *converges in distribution to*, $\xrightarrow{a.s.}$ means *converges almost surely to* and *plim* means *convergence in probability*. These three concepts are defined, e.g., in White (2001). Furthermore, $\boldsymbol{I}_n$ will be an n-dimensional identity matrix and $\boldsymbol{A}_m$ a square $m-$by$-m$ matrix, unless defined otherwise. The proofs of the propositions are given in the Appendix.

2. Previous Results in Subspace Methods

Consider a linear fixed-coefficients system that can be described by the following state space model:

$$\boldsymbol{x}_{t+1} = \boldsymbol{\Phi}\boldsymbol{x}_t + \boldsymbol{E}\boldsymbol{\psi}_t \tag{1a}$$

$$\boldsymbol{z}_t = \boldsymbol{H}\boldsymbol{x}_t + \boldsymbol{\psi}_t \tag{1b}$$

where $\boldsymbol{x}_t$ is a state n-vector, n being the true order of the system. In addition, $\boldsymbol{z}_t$ is an observable output m-vector, which is assumed to be zero-mean, $\boldsymbol{\psi}_t$ is an unobservable input m-vector, and $\boldsymbol{\Phi}$, $\boldsymbol{E}$ and $\boldsymbol{H}$ are parameter matrices with dimensions $(n \times n)$, $(m \times m)$ and $(n \times m)$, respectively. We suppose that the following assumptions hold in (1a-1b).

Assumptions. A.1: $\boldsymbol{\psi}_t$ is a sequence of zero-mean uncorrelated variables with $\mathrm{E}(\boldsymbol{\psi}_t\boldsymbol{\psi}_t') = \boldsymbol{\Gamma}$, $\boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma}$ is a positive definite matrix. A.2: The system is stable and strictly minimum-phase, *i.e.*, all the eigenvalues of $\boldsymbol{\Phi}$ and $(\boldsymbol{\Phi} - \boldsymbol{E}\boldsymbol{H})$ lie inside the unit circle.

We use the SSE, or also called innovations, form (1a-1b) since it is general and simpler than other representations. Its generality is discussed by Casals et al. (2012), who show that SSE, MSE and VARMAX models are equally general and freely interchangeable.

Additionally, throughout the paper we will also use $\bar{\boldsymbol{z}}_t$, a standardized version of $\boldsymbol{z}_t$, defined as $\bar{\boldsymbol{z}}_t = \hat{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{z}_t$, where $\hat{\boldsymbol{\Sigma}} = T^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t\boldsymbol{z}_t'$ and T is the sample size.

García-Hiernaux, Jerez & Casals (2010) show that model (1a-1b) can be transformed into a single equation in matrix form as $\boldsymbol{Z}_f = \boldsymbol{O}\boldsymbol{X}_f + \boldsymbol{V}\boldsymbol{\Psi}_f$, where: a) $\boldsymbol{Z}_f$ is a block Hankel matrix whose columns can be generally defined as $[\boldsymbol{z}_t', \ldots, \boldsymbol{z}_{t+f-1}']'$ and each column is specified by a different value of t such that: $t = p+1, \ldots, T-f+1$;[1] b) p and f are two integers chosen by the user, where $p > n$; and, c) $\boldsymbol{X}_f$ and $\boldsymbol{\Psi}_f$ are as $\boldsymbol{Z}_f$ but with $\boldsymbol{x}_t$ or $\boldsymbol{\psi}_t$, respectively, instead of $\boldsymbol{z}_t$. For simplicity, we assume $p = f$, denoting this integer by i. In this case, $\boldsymbol{Z}_f$ and $\boldsymbol{\Psi}_f$ are $im \times (T - 2i + 1)$ matrices. To simplify the notation, we denote the number of columns of both matrices by $T_* = T - 2i + 1$. Last, as it is detailed in García-Hiernaux et al. (2010), Section 2, matrices $\boldsymbol{O}$ and $\boldsymbol{V}$ are known functions of the original parameter matrices, $\boldsymbol{\Phi}$, $\boldsymbol{E}$ and $\boldsymbol{H}$:

$$\boldsymbol{O} := \begin{pmatrix} \boldsymbol{H}' & (\boldsymbol{H}\boldsymbol{\Phi})' & (\boldsymbol{H}\boldsymbol{\Phi}^2)' & \ldots & (\boldsymbol{H}\boldsymbol{\Phi}^{i-1})' \end{pmatrix}'_{im \times n} \tag{2}$$

$$\boldsymbol{V} := \begin{pmatrix} \boldsymbol{I}_m & \boldsymbol{0} & \boldsymbol{0} & \ldots & \boldsymbol{0} \\ \boldsymbol{H}\boldsymbol{E} & \boldsymbol{I}_m & \boldsymbol{0} & \ldots & \boldsymbol{0} \\ \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{E} & \boldsymbol{H}\boldsymbol{E} & \boldsymbol{I}_m & \ldots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{H}\boldsymbol{\Phi}^{i-2}\boldsymbol{E} & \boldsymbol{H}\boldsymbol{\Phi}^{i-3}\boldsymbol{E} & \boldsymbol{H}\boldsymbol{\Phi}^{i-4}\boldsymbol{E} & \ldots & \boldsymbol{I}_m \end{pmatrix}_{im} \tag{3}$$

[1] From now on all the block Hankel matrices will be defined in a similar way.

Given A.2 and for large values of i and T, $\boldsymbol{X}_f$ is to a close approximation representable as a linear combination of the past of the output, $\boldsymbol{M}\boldsymbol{Z}_p$, where $\boldsymbol{Z}_p := [\boldsymbol{z}'_{t-p}, \ldots, \boldsymbol{z}'_{t-1}]'$ with $t = p+1, \ldots, T-f+1$. Then, the relation between the past and the future of the output can be expressed by:

$$\boldsymbol{Z}_f \simeq \boldsymbol{\beta}\boldsymbol{Z}_p + \boldsymbol{V}\boldsymbol{\Psi}_f \tag{4}$$

where $\boldsymbol{\beta} = \boldsymbol{O}\boldsymbol{M}$. For a given system order n, subspace methods first solve a reduced-rank (as $\boldsymbol{\beta}$ is an im square matrix with rank $n < im$) weighted least squares problem by estimating $\boldsymbol{\beta}$ as:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{Z}_f\boldsymbol{Z}'_p(\boldsymbol{Z}_p\boldsymbol{Z}'_p)^{-1} \tag{5}$$

and splitting it to estimate $\boldsymbol{O}$ and $\boldsymbol{M}$, and then $\boldsymbol{V}$. Finally, the parameter matrices in (1a-1b) can be obtained from the estimates $\hat{\boldsymbol{O}}$, $\hat{\boldsymbol{M}}$ and $\hat{\boldsymbol{V}}$, see, e.g., Katayama (2005).

3. Some Distributional Results

We begin by establishing the null hypothesis that $\boldsymbol{z}_t$ has no correlations different from zero up to lag order k, i.e., H_0: $\rho_j = 0$, $j = 1, 2, \ldots, k$, where ρ_j is the correlation coefficient of order j. It is common in the literature that the user just chooses k to conduct the hypothesis testing. Accordingly, we define i as a function of k, such that i is the integer rounded toward infinity of $(k+1)/2$. However, the tests could be directly adapted to any other value of i, or even different values of p and f.

The first proposal uses a generalized least squares approach. Using the previously defined standardized version of the output and input, we have $\bar{\boldsymbol{Z}}_f = \bar{\boldsymbol{\beta}}\bar{\boldsymbol{Z}}_p + \bar{\boldsymbol{V}}\bar{\boldsymbol{\Psi}}_f$, where $\bar{\boldsymbol{Z}}_p, \bar{\boldsymbol{\Psi}}_p$ are as $\boldsymbol{Z}_p, \boldsymbol{\Psi}_p$ but with $\bar{\boldsymbol{z}}_t, \bar{\boldsymbol{\psi}}_t$ instead of the original $\boldsymbol{z}_t, \boldsymbol{\psi}_t$. Matrix $\bar{\boldsymbol{\beta}}$ can be estimated as (5), but with the standardized matrices $\bar{\boldsymbol{Z}}_p$ and $\bar{\boldsymbol{Z}}_f$ instead of $\boldsymbol{Z}_p$ and $\boldsymbol{Z}_f$. Notice that an immediate consequence of the null hypothesis is that $\bar{\boldsymbol{\beta}} = \boldsymbol{0}_{im}$. By applying the *vec* operator, which stacks the columns of a matrix into a long vector, on $\hat{\bar{\boldsymbol{\beta}}}$ we state the following proposition:

Proposition 1. *Given A.1-A.2, under* H_0, $\sqrt{T_*}vec(\hat{\bar{\boldsymbol{\beta}}}|\bar{\boldsymbol{Z}}_p) \xrightarrow{d} N\big(\boldsymbol{0}, \bar{\boldsymbol{\Pi}}\big)$, *where* $\bar{\boldsymbol{\Pi}}$ *is derived in the Appendix.*

The second test comes from a canonical correlation approach. This one is based on the information held in $\boldsymbol{O}$, which affects $\boldsymbol{Z}_f$ through $\boldsymbol{\beta}$, see (4). The canonical correlation analysis (CCA) between $\boldsymbol{Z}_f$ and $\boldsymbol{Z}_p$ is usually performed by analyzing the product $(\boldsymbol{Z}_f\boldsymbol{Z}'_f)^{-\frac{1}{2}}\boldsymbol{Z}_f\boldsymbol{Z}'_p(\boldsymbol{Z}_p\boldsymbol{Z}'_p)^{-\frac{1}{2}}$, see Katayama (2005) for a detailed description on CCA. From equation (5), one could get the product above from $(\boldsymbol{Z}_f\boldsymbol{Z}'_f)^{-\frac{1}{2}}\hat{\boldsymbol{O}}$, estimating $\boldsymbol{O}$ as $\boldsymbol{Z}_f\boldsymbol{Z}'_p(\boldsymbol{Z}_p\boldsymbol{Z}'_p)^{-\frac{1}{2}}$ and then $\boldsymbol{M}$ as $(\boldsymbol{Z}_p\boldsymbol{Z}'_p)^{-\frac{1}{2}}$, so that the equality $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{O}}\hat{\boldsymbol{M}}$ holds. This second alternative leads to Proposition 2:

Proposition 2. *Given A.1-A.2, under* H_0, $\sqrt{T_*}vec\big((\boldsymbol{Z}_f\boldsymbol{Z}'_f)^{-\frac{1}{2}}\hat{\boldsymbol{O}}|\boldsymbol{Z}_p\big) \xrightarrow{d} N\big(\boldsymbol{0}, \bar{\boldsymbol{\Pi}}\big)$.

4. The Test Statistics

The covariance matrix $\bar{\mathbf{\Pi}}$ is not, generally, the identity matrix. In fact, it is only so when $i = 1$. For $i > 1$ some elements in $\hat{\boldsymbol{\beta}}$ and $(\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\hat{\boldsymbol{O}}$ are perfectly correlated by construction, see equation (8) in the Appendix. However, as the structure of $\bar{\mathbf{\Pi}}$ is known, the following proposition applies.

Proposition 3. *For any random matrix* $\boldsymbol{A}$ *such that* $\sqrt{T_*}vec\boldsymbol{A} \xrightarrow{d} N(\mathbf{0}, \bar{\mathbf{\Pi}})$*, there is an idempotent matrix* $\boldsymbol{P}_{(im)^2}$ *of rank* m^2k*, such that* $\mathcal{S}_A = T_*vec(\boldsymbol{A})'\boldsymbol{P}vec(\boldsymbol{A}) \xrightarrow{d} \chi^2_{m^2k}$.

Corollary 1. Consequently, by combining Propositions 1, 2 and 3, we get that both, $\mathcal{S}_{\boldsymbol{\beta}} = T_*vec(\hat{\tilde{\boldsymbol{\beta}}})'\boldsymbol{P}vec(\hat{\tilde{\boldsymbol{\beta}}})$ and $\mathcal{S}_{\boldsymbol{O}} = T_*vec\big((\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\hat{\boldsymbol{O}}\big)'\boldsymbol{P}vec\big((\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\hat{\boldsymbol{O}}\big)$ converge to a chi-square distribution with m^2k degrees of freedom.

Matrix $\boldsymbol{P}$ is the product of two weighting matrices that average the perfectly correlated elements of $vec(\boldsymbol{A})$ in a vector of m^2k uncorrelated elements. This point deserves further discussion, as it makes the procedure flexible by tuning matrix $\boldsymbol{P}$ according on the specific case. For instance, some $\boldsymbol{P}$ could be chosen with the aim of reducing the effects of outliers or increasing the statistical power of the tests.

We have seen that, when $i > 1$ some elements in $\hat{\tilde{\boldsymbol{\beta}}}$ and $(\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\hat{\boldsymbol{O}}$ are perfectly correlated. Matrix $\boldsymbol{P}$, as it is proposed in the proof of Proposition 3 averages the perfectly correlated elements to obtain a vector of uncorrelated components. The procedure computes each k-order correlation for different non-disjoint subsamples and averages them to obtain a single one. In this way, the effect of an outlier will be mitigated, provided that it only affects a small proportion of the weighted correlations. This will be more likely the more subsamples we use, *i.e.*, the higher i is. Obviously, our statistics do not use robust estimation methods, as M-estimators or MM-estimators, and therefore they are not robust statistics and will perform worse than those methods in the presence of outliers. However, we expect that they present a better performance than non-robust statistics as Q_{LB} in such cases; specifically, innovational outliers, additive outliers or level changes (see, for definitions, Tsay 1988). An example illustrates this feature in the next section.

An interesting point that deserves more attention is that one could easily tune the matrix $\boldsymbol{P}$ according to the data. If we are suspicious about the presence of outliers then, instead of calculating the mean of several k-order correlation (which is the proposal here), the median or the minimum could be used. In these cases, the distribution of the statistics should be derived but the statistics are likely to be more robust.

On the other hand, often in practice, only the low-order correlations are of interest to analysts. Consequently, the possibility of modifying $\boldsymbol{P}$ by increasing the weights of low lags (either *ad-hoc* or using a more sophisticated mechanism) should increase the power of the tests.

In any case, a standard use of the Portmanteau tests is to check the residuals obtained from fitting Vector Autoregressive Moving Average, VARMA, models.

Here we adopt the usual definition of a stationary m-variate ARMA(p,q) process (see, *e.g.*, Liu 2006, p. 14.2). Nevertheless, when $\boldsymbol{z}_t$ are the residuals from a VARMA model, the asymptotic distribution of $\mathcal{S}_\beta$ and $\mathcal{S}_O$ is not as it has been shown. The reason is that A.1 does not hold, as residuals, contrary to innovations, present some linear constraints inherit from the VARMA estimation (see, e.g., Mauricio 2007). In these circumstances, the following proposition establishes the asymptotic distribution of both statistics.

Proposition 4. *When $\boldsymbol{z}_t$ in (1b) are the residuals from a fitted m-vector ARMA(p,q) model, then, under H_0, $\mathcal{S}_\beta$ and $\mathcal{S}_O$ converge in distribution to a $\chi^2_{m^2(k-p-q)}$.*

At this point, notice that testing H_0 in any m-variate process requires (if the Ljung-Box test is used) a Q-matrix that leads to m^2 different statistics. As Hosking (1980) test, ours offer a more natural scalar statistic instead. Further, it is straightforward to see that for $p = 1$ and $f = k + 1$ both, $\mathcal{S}_\beta$ and $\mathcal{S}_O$, are equivalent to: (i) Ljung-Box statistic when $m = 1$ and (ii) Hoskings' statistic when $m \geq 1$ (see, Hosking 1980, p. 605). In short, our procedures generalize Ljung-Box and Hosking's procedures, allowing for different values of p and f.

Furthermore, these results are extended to multiplicative seasonal VARMA(p,q) $\times(P,Q)_s$ models, where s is the seasonal period and (P,Q) are the seasonal autoregressive and moving average orders, respectively (see, Liu 2006, p. 14.36). Regarding this, McLeod (1978), for the univariate case ($m = 1$), and Ursu & Duchesne (2009), for multivariate processes, proved that an adjusted version of the Q-statistic follows a $\chi^2_{m^2(k-p-q-P-Q)}$. With our proposals, if one only identifies and estimates the seasonal parameters (P,Q), $\mathcal{S}_\beta$ or $\mathcal{S}_O$ and Proposition 4 could easily be used to check whether there is seasonal correlation in the residuals, testing H_0: $\rho_j = 0$, $j = s, 2s, \ldots, ks$. The statistics should be computed by replacing $\boldsymbol{Z}_p$ and $\boldsymbol{Z}_f$ by their seasonal counterparts $\boldsymbol{Z}^s_p := [\boldsymbol{z}'_{t-si}, \boldsymbol{z}'_{t-s(i-1)}, \ldots, \boldsymbol{z}'_{t-s}]'$ and $\boldsymbol{Z}^s_f := [\boldsymbol{z}'_t, \boldsymbol{z}'_{t+s}, \ldots, \boldsymbol{z}'_{t+s(i-1)}]'$, where $t = si+1, s(i+1)+1, \ldots, T-s(i-1)$. In those cases $\mathcal{S}_\beta$ and $\mathcal{S}_O$ follow a $\chi^2_{m^2(k-P-Q)}$. Hence, the adequacy of a VARMA$(p,q)\times(P,Q)_s$ model can be checked by sequentially identifying, estimating and applying the tests using the seasonal matrices, $\boldsymbol{Z}^s_p$ and $\boldsymbol{Z}^s_f$, and then the regular ones, $\boldsymbol{Z}_p$ and $\boldsymbol{Z}_f$. The sequential procedure implies a gain in terms of degrees of freedom with respect to Ursu & Duchesne (2009) when testing for seasonal correlation, as we only consider the seasonal part and not the complete model. This may be a great advantage in very short samples.

5. Numerical Examples

In this section we investigate the finite sample properties of the proposed tests. Its performance is compared with that of Ljung-Box (Q_{LB}) and Hosking (P_H) statistics, as they are the most common and cited diagnostic tests in the literature for the univariate and the multivariate case, respectively. As said previously, no comparison against robust methods is made as ours do not fulfill those characteristics. However, in order to analyze its behavior in different situations, we split the

study into some Monte Carlo simulations of univariate processes without outliers contamination and two applications to real data in which, at least the first one, contains documented additive outliers.

5.1. Monte Carlo Simulations

Firstly, we will study how the autocorrelation structure affects the empirical size and power of the tests.

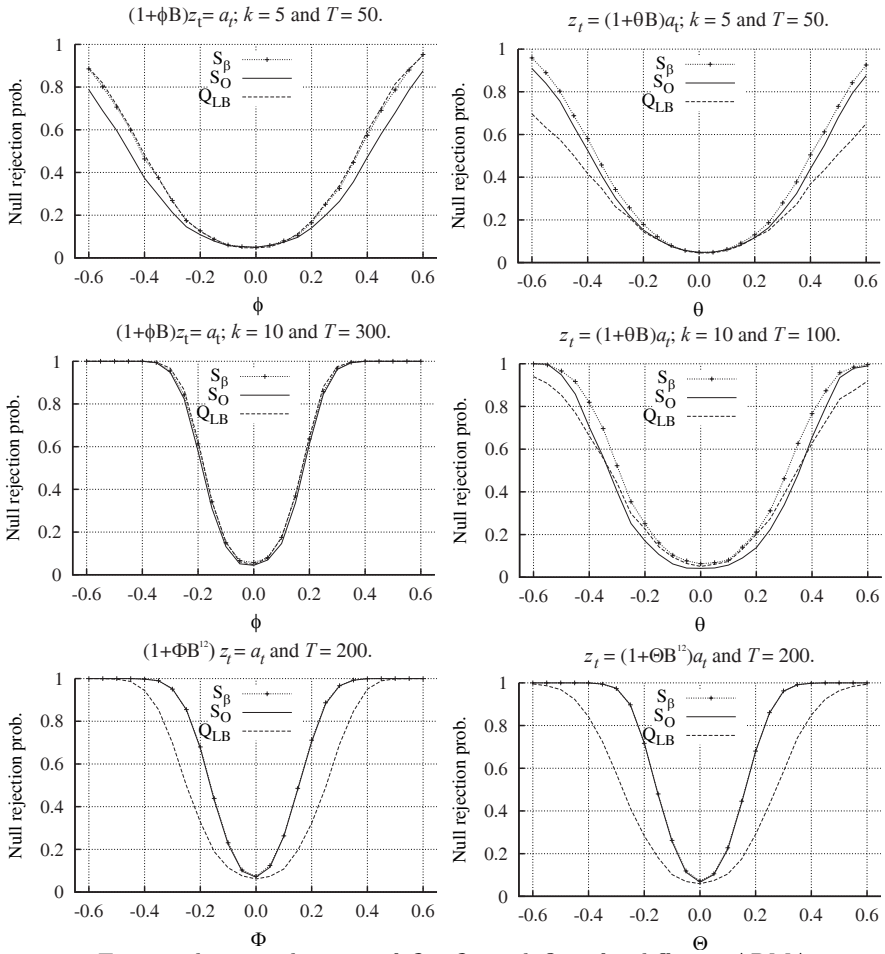


FIGURE 1: Empirical size and power of $\mathcal{S}_\beta$, $\mathcal{S}_O$ and Q_{LB} for different ARMA processes (computed with a χ^2_k at 5% and 5000 replications). The graphs at the bottom depict the size and power for two seasonal processes. In these cases, Q_{LB} is computed with $k = 24$ to be able to capture the seasonal structure, while $\mathcal{S}_\beta$ and $\mathcal{S}_O$ are computed with the seasonal matrices $\boldsymbol{Z}_p^s$ and $\boldsymbol{Z}_f^s$ and $k^s = 2$.

Figure 1 presents the empirical size and power of $\mathcal{S}_O$, $\mathcal{S}_\beta$ and Q_{LB} for alternative AR(1) and MA(1) processes, with different k (lags) and T (sample size).

Hosking's test is omitted as it coincides with Q_{LB} in univariate processes.[2] The most noticeable features of this exercise are:

1. In processes without seasonality and short samples ($T = 50$):
   a) Q_{LB} and $\mathcal{S}_\beta$ perform very similarly with autoregressive structures, both being slightly more powerful than $\mathcal{S}_O$.
   b) The empirical power of Q_{LB} is clearly outperformed by our two proposals when MA structures. This result partially coincides with Monti (1994) who proposes a test using the residual partial autocorrelations whose behavior is better than that of Q_{LB} if the order of the MA is understated. However, in that case it was shown that Q_{LB} was more powerful if the order of the AR part was understated. In contrast, we did not find any evidence of this when applying $\mathcal{S}_\beta$.
2. The asymptotically equivalence of the three tests is observed when T grows. For $T = 300$ and a AR(1) process the performance of the three tests is almost identical. When $T = 200$ and a MA(1) process our tests still outperform Q_{LB}, although less evidently than when $T = 50$.
3. In seasonal processes, $\mathcal{S}_O$ and $\mathcal{S}_\beta$ clearly outperform Q_{LB} in terms of statistical power. Not surprisingly, this enhancement is even bigger with seasonal MA(1) processes. The explanation comes from the fact that $\mathcal{S}_O$ and $\mathcal{S}_\beta$ are computed with the seasonal matrices $\boldsymbol{Z}_p^s$ and $\boldsymbol{Z}_f^s$ defined in Section 4 and the test is then computed with $k^s = 2$. However, Q_{LB} is computed with $k = 24$ to be able to capture the seasonal correlation.

Secondly, we analyze the empirical distribution of the statistics under H_0 for white noise samples and increasing values of k. Notice that in those cases the null distribution follows a χ^2_k. In this context, Figure 2 shows that $\mathcal{S}_\beta$ better fits the theoretical distribution than Q_{LB} and $\mathcal{S}_O$, when $k = 15$ and $T = 50$. Interestingly enough, the simulations evidence that Q_{LB} and $\mathcal{S}_O$ empirical distributions get further away from the theoretical one when k increases for a given T. Nevertheless, the distribution of $\mathcal{S}_\beta$ correctly fits its theoretical counterpart regardless of the value of k.[3]

5.2. Two examples with real data

The first example with real data considers the Residence Telephone Extensions Inward Movement known as RESEX series (y_t). The left plot of Figure 3 shows the original monthly series that goes from January 1966 to May 1973, where observations $t = 83, 84$ are larger than the rest. These two outliers have a known cause, namely a bargain month, in which residence extensions could be requested free of

[2]Simulations with higher lags in pure autoregressive, pure moving average or ARMA models show similar or mixed results that do not suggest additional conclusions and, consequently, are not presented here. However, they are available from the author upon request.

[3]Additional simulations not shown here are available from the author upon request.

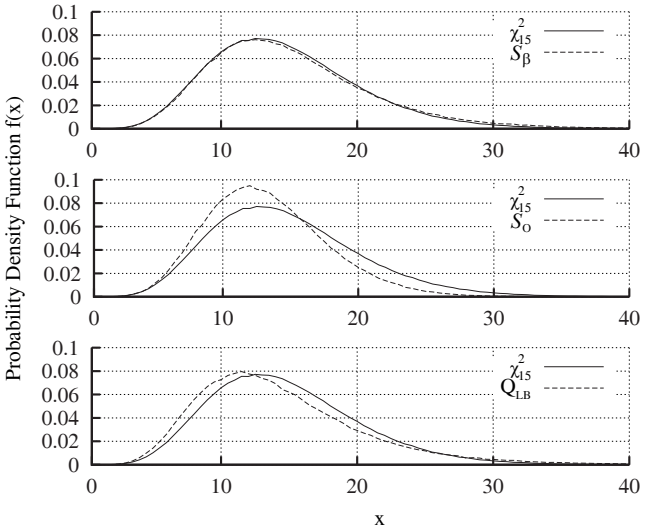


FIGURE 2: Empirical distribution for $\mathcal{S}_\beta$, $\mathcal{S}_O$ and Q_{LB} compared to a theoretical χ^2_{15}; 250,000 replications for $T = 50$ and $k = 15$.

charge. Robust methods identify an AR(1) in the regularly and seasonally differenced transformation ($\nabla\nabla_{12} y_t$), see, e.g., Rousseeuw & Leroy (1987) or Li (2004). On the contrary, standard methods usually do not capture the autocorrelation structure due to the effect of the outliers.

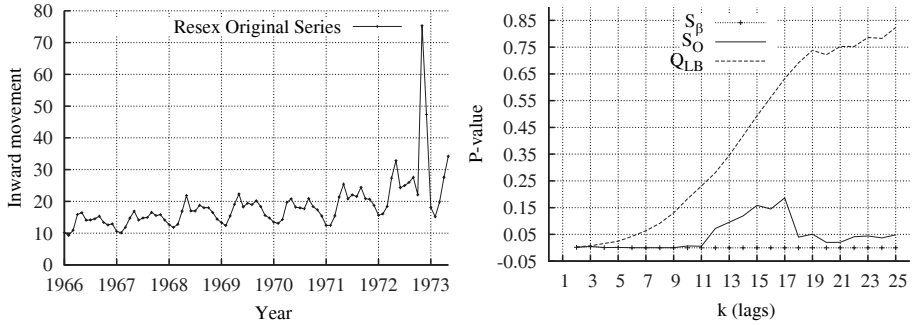


FIGURE 3: Top plot: Original RESEX series (y_t). Bottom plot: P-values of $\mathcal{S}_\beta$, $\mathcal{S}_O$ and Q_{LB} for lags (k) from 1 to 25 obtained by applying the statistics to the transformed series $\nabla\nabla_{12} \log y_t$.

When we apply $\mathcal{S}_\beta$, $\mathcal{S}_O$ and Q_{LB} to the transformed series $\nabla\nabla_{12} \log y_t$, we find that Q_{LB} does not reject the null from $k = 7$ at 5% of significance and from $k = 8$ at 10%. However, $\mathcal{S}_O$ rejects the null at a 5% for all k except when $k = 12 - 17$, where the p-values always remain below 16%. Finally, $\mathcal{S}_\beta$ behaves much better than Q_{LB} and $\mathcal{S}_O$ with this data, rejecting the null at 1% of significance for all k studied. This example is relevant as most empirical works only show the Q_{LB} values for high lags (usually $10, 15$ or 20) without paying attention to the loss of

power when k increases, that can grow dramatically in the presence of outliers. $\mathcal{S}_\beta$ behavior explanation lies in the fact that i has been defined as a positive function of k (see Section 2), so when k grows, i increases. As i is the number of subsamples to compute the autocorrelations of the same order, when i increases, the weight of the contaminated subsamples diminishes.

The second example deals with the logarithms of indices of monthly flour prices in the cities of Buffalo, Minneapolis and Kansas City, over the period from August 1972 to November 1980, which give us 100 observations at each site. The aim of modeling these data is to illustrate the performance of the proposed statistics, as specification tools, and compare it with Q_{LB} and P_H.

Since all series appear non-stationary, we use the log-difference transformation $\boldsymbol{z}_t = \nabla \log(\boldsymbol{y}_t)$, where $\boldsymbol{y}_t$ are the original series. Table 1 shows the results of applying the statistics to $\boldsymbol{z}_t$ with different lags. The first conclusion is that even if all the tests suggest that there are significant correlations, at least up to order one, Q_{LB} presents very low power when a (not-so) large lag is chosen. It seems that the significant correlations at lag one are diluted by insignificant correlations at other lags, and this effect is much more important in Q_{LB} than in $\mathcal{S}_\beta$, $\mathcal{S}_O$ or P_H. In this context, notice that $\mathcal{S}_\beta$ is the only statistic that keeps its p-value under 5% for $k = 5, 10$. Additionally, Q_{LB} only reveals 5 out of 9 correlations statistically significant at 5%, when $k = 1$.

TABLE 1: P-value of the statistics. H_0: There are no correlations up to lag k in $\boldsymbol{z}_t$.

k (lag)	$\mathcal{S}_O$	$\mathcal{S}_\beta$	P_H	Q_{LB}
1	.000*	.000*	.000*	$\begin{pmatrix} .172 & .026^* & .047^* \\ .103 & .027^* & .056 \\ .045^* & .018^* & .066 \end{pmatrix}$
5	.241	.035*	.072	$\begin{pmatrix} .822 & .416 & .506 \\ .716 & .421 & .493 \\ .470 & .309 & .549 \end{pmatrix}$
10	.155	.003*	.082	$\begin{pmatrix} .954 & .744 & .632 \\ .918 & .734 & .545 \\ .779 & .682 & .573 \end{pmatrix}$

* rejects at 5%.

Following the results obtained with Q_{LB} at 5% in Table 1 when $k = 1$, a restricted VAR(1) model $(\boldsymbol{I} - \boldsymbol{\Phi}_1 B)\boldsymbol{z}_t = \boldsymbol{a}_t$ is tentatively specified. Parameter estimates result in:

$$\hat{\boldsymbol{\Phi}}_1 = \begin{pmatrix} 0 & -.188^* & -.035 \\ 0 & -.289^* & 0 \\ -.401^* & .117 & 0 \end{pmatrix}, \quad \hat{\boldsymbol{\Gamma}}_{\boldsymbol{a}} = \begin{pmatrix} 2.263 & 2.296 & 2.202 \\ & 2.496 & 2.364 \\ & & 2.770 \end{pmatrix} \times 10^{-3}, \quad (6)$$

where '0' denotes an entry constrained to be zero and '*' means the parameter is significant at 5%. Table 2 presents the p-value of the diagnostic tests on the residuals of model (6).

TABLE 2: P-value of the statistics. H_0: There are no correlations up to lag k in model (6) residuals.

Statistic	k (lags)			
	2	5	10	15
$\mathcal{S}_O$	.003*	.200	.110	.202
$\mathcal{S}_\beta$	.000*	.003*	.006*	.007*
P_H	.000*	.037*	.052	.256
$Q_{LB}^{\dagger}$	.429	.869	.792	.884

$Q_{LB}^{\dagger}$ is to the lowest p-value among all the elements of the Q_{LB} matrix.
* rejects at 5%.

Q_{LB} suggests that the correlations are zero for $k = 2, 5, 10, 15$ at 10% level of significance, implying that model (6) is appropriate. However, $\mathcal{S}_O$, P_H and $\mathcal{S}_\beta$ reject H_0 for $k = 2$, $k = 2, 5, 10$ and $k = 2, 5, 10, 15$, respectively, at 5% level. Hence, $\mathcal{S}_O$, P_H and particularly $\mathcal{S}_\beta$ strongly evidence that Q_{LB} leads to an inappropriate specification. Instead, if we specify an unrestricted VAR(1), the estimation returns:

$$\hat{\mathbf{\Phi}}_1 = \begin{pmatrix} 1.226^* & -1.355^* & .005 \\ .830^* & -1.027^* & .035 \\ .463 & -.813^* & .142 \end{pmatrix}, \quad \hat{\mathbf{\Gamma}}_{\boldsymbol{a}} = \begin{pmatrix} 2.033 & 2.140 & 2.039 \\ & 2.390 & 2.253 \\ & & 2.647 \end{pmatrix} \times 10^{-3} \quad (7)$$

To check if the residual correlations of model (7) are zero, the four procedures are again employed. Table 3 shows these results. None of the tests rejects H_0 for any value of k. Surprisingly, Q_{LB} presents the smallest evidence in favor of the null out of the four alternative for $k = 2, 5$. Model (7) was proposed by Lütkepohl & Poskitt (1996) and, as it was shown in Grubb (1992), is better than many other alternatives, in particular model (6).

TABLE 3: P-value of the statistics. H_0: There are no correlations up to lag k in model (7) residuals.

Statistic	k (lags)			
	2	5	10	15
$\mathcal{S}_O$	.953	.952	.480	.454
$\mathcal{S}_\beta$	.937	.952	.445	.506
P_H	.945	.951	.601	.838
$Q_{LB}^{\dagger}$	.455	.756	.736	.858

$Q_{LB}^{\dagger}$ is to the lowest p-value among all the elements of the Q_{LB} matrix.
* rejects at 5%.

From this exercise with multiple series we conclude that: (i) multivariate Portmanteau statistics, $\mathcal{S}_{\boldsymbol{\beta}}, \mathcal{S}_{\boldsymbol{O}}$ and P_H, perform better than the multiple Q_{LB}, and (ii) $\mathcal{S}_\beta$ seems to be more powerful than $\mathcal{S}_O$ and P_H when k grows.

6. Concluding Remarks

This work tackles the problem of diagnostic checking from an original viewpoint. Two statistics based on subspace methods are presented and their asymptotic distributions are derived under the null. They generalize the Box-Pierce statistic for single series, the Hoskings' statistic in the multivariate case and are able to separately test seasonal and regular correlations. Monte Carlo simulations and two examples with real data show that our proposals perform better than the common Ljung-Box Q-statistic in many different situations. The procedures can sequentially be used to determine the system order, as the null hypothesis can always be written as $n = 0$, which is a critical decision in the subspace methods literature and the applied data modeling.

Moreover, the subspace structure and the possibility of tuning a weight matrix make the tests more flexible and robust against outliers than non-robust alternatives. In this paper we just propose a particular form for this matrix $\boldsymbol{P}$ (see proof of Proposition 3), but others are possible and could be fitted to the characteristics of the data. A deeper analysis of this point with the suggestion of different matrices $\boldsymbol{P}$ could be the core of a next research.

Finally, the procedures used in the numerical examples and described in the paper are implemented in a MATLAB toolbox for time series modeling called `E4` that can be downloaded from the webpage `www.ucm.es/info/icae/e4`. The source code for all the functions in the toolbox is freely provided under the terms of the GNU General Public License. This site also includes a complete user manual and other materials.

Acknowledgment

Manuel Domínguez, Miguel Jerez and two anonymous referees made useful comments and suggestions to previous versions of this work. The author gratefully acknowledges financial support from Ministerio de Educación y Ciencia, ref. ECO2011-23972 and the Ramón Areces Foundation.

[Recibido: noviembre de 2012 — Aceptado: mayo de 2013]

References

Aoki, M. (1990), *State Space Modelling of Time Series*, Springer Verlag, New York.

Box, G. E. P. & Pierce, D. A. (1970), 'Distribution of residuals autocorrelations in autoregressive-integrated moving average time series models', *Journal of the American Statistical Association* **65**(332), 1509–1526.

Casals, J., García-Hiernaux, A. & Jerez, M. (2012), 'From general state-space to VARMAX models', *Mathematics and Computers in Simulation* **80**(5), 924–936.

Casals, J., Sotoca, S. & Jerez, M. (1999), 'A fast and stable method to compute the likelihood of time invariant state space models', *Economics Letters* **65**(3), 329–337.

García-Hiernaux, A., Jerez, M. & Casals, J. (2010), 'Unit roots and cointegration modeling through a family of flexible information criteria', *Journal of Statistical Computation and Simulation* **80**(2), 173–189.

Grubb, H. (1992), 'A multivariate time series analysis of some flour price data', *Applied Statistics* **41**, 95–107.

Hosking, J. R. M. (1980), 'The multivariate Pormanteau statistic', *Journal of the American Statistical Association* **75**(371), 602–608.

Katayama, T. (2005), *Subspace Methods for System Identification*, Springer Verlag, London.

Li, W. K. (2004), *Diagnostic Checks in Time Series*, Chapman and Hall/CRC, Florida.

Liu, L. M. (2006), *Time Series Analysis and Forecasting*, 2 edn, Scientific Computing Associates Corporation, Illinois.

Ljung, G. M. & Box, G. E. P. (1978), 'On a measure of lack of fit in time series models', *Biometrika* **65**, 297–303.

Lütkepohl, H. & Poskitt, D. S. (1996), 'Specification of echelon form VARMA models', *Journal of Business and Economic Statistics* **14**(1), 69–79.

Mauricio, J. A. (2007), 'Computing and using residuals in time series models', *Computational Statistics and Data Analysis* **52**(3), 1746–1763.

McLeod, A. I. (1978), 'On the distribution of residual autocorrelations in Box-Jenkins model', *Journal of the Royal Statistics Society B* **40**, 296–302.

Monti, A. C. (1994), 'A proposal for residual autocorrelation test in linear models', *Biometrika* **81**, 776–780.

Rousseeuw, P. J. & Leroy, A. M. (1987), *Robust Regression and Outlier Detection*, John Wiley, New York.

Tsay, R. S. (1988), 'Outliers, level shifts, and variance changes in time series', *Journal of Forecasting* **7**, 1–20.

Ursu, E. & Duchesne, P. (2009), 'On multiplicative seasonal modelling for vector time series', *Statistics and Probability Letters* **79**(19), 2045–2052.

White, H. (2001), *Asymptotic Theory for Econometricians*, Academic Press.

Appendix

Proof of Proposition 1. Equation (4) can be written as an equality by including a term that tends to zero at an exponential rate as a result of the minimum-phase assumption. For the lack of simplicity, we neglect this term during the proof and treat equation (4) as an equality. By applying the *vec* operator to the standardized version of equation (4), we have $vec\bar{\boldsymbol{Z}}_f = (\bar{\boldsymbol{Z}}_p' \otimes \boldsymbol{I}_{im})vec\bar{\boldsymbol{\beta}} + vec\bar{\boldsymbol{\Psi}}_f$, where we use that, under H_0, $\bar{\boldsymbol{V}} = \boldsymbol{I}_{im}$. From this, $vec\hat{\bar{\boldsymbol{\beta}}} = [(\bar{\boldsymbol{Z}}_p' \otimes \boldsymbol{I}_{im})'(\bar{\boldsymbol{Z}}_p' \otimes \boldsymbol{I}_{im})]^{-1}(\bar{\boldsymbol{Z}}_p' \otimes \boldsymbol{I}_{im})'vec\bar{\boldsymbol{Z}}_f$, and hence we get $vec(\hat{\bar{\boldsymbol{\beta}}} - \bar{\boldsymbol{\beta}}) = \bar{\boldsymbol{H}}^{-1}\bar{\boldsymbol{A}}'vec\bar{\boldsymbol{\Psi}}_f$, where $\bar{\boldsymbol{H}} = \bar{\boldsymbol{A}}'\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{A}} = \bar{\boldsymbol{Z}}_p' \otimes \boldsymbol{I}_{im}$. Therefore, the covariance of $vec\hat{\bar{\boldsymbol{\beta}}}$ conditional to $\bar{\boldsymbol{Z}}_p$ is $cov[vec\hat{\bar{\boldsymbol{\beta}}}|\bar{\boldsymbol{Z}}_{\boldsymbol{p}}] = \bar{\boldsymbol{H}}^{-1}\bar{\boldsymbol{A}}'(\boldsymbol{\Omega} \otimes \boldsymbol{I}_m)\bar{\boldsymbol{A}}\bar{\boldsymbol{H}}^{-1}$, where $(\boldsymbol{\Omega} \otimes \boldsymbol{I}_m)$ denotes de covariance of $vec\bar{\boldsymbol{\Psi}}$ and we use that, under H_0, $\mathrm{E}(\bar{\boldsymbol{z}}_t\bar{\boldsymbol{z}}_t') = \mathrm{E}(\bar{\boldsymbol{\psi}}_t\bar{\boldsymbol{\psi}}_t') = \boldsymbol{I}_m$. Asymptotically, the Ergodic Theorem (see, Theorem 3.34, White 2001) and H_0 ensure that $T_*^{-1}\bar{\boldsymbol{A}}'(\boldsymbol{\Omega} \otimes \boldsymbol{I}_m)\bar{\boldsymbol{A}} \overset{a.s.}{\to} \bar{\boldsymbol{\Pi}}$ and $T_*\bar{\boldsymbol{H}}^{-1} \overset{a.s.}{\to} \boldsymbol{I}_{(im)^2}$, where $\bar{\boldsymbol{\Pi}}$ has the following structure:

$$\bar{\boldsymbol{\Pi}} = \begin{pmatrix} \boldsymbol{I}_{im^2} & \boldsymbol{\Pi}_{i-1} & \boldsymbol{\Pi}_{i-2} & \dots & \boldsymbol{\Pi}_1 \\ \boldsymbol{\Pi}_{i-1}' & \boldsymbol{I}_{im^2} & \boldsymbol{\Pi}_{i-1} & \dots & \boldsymbol{\Pi}_2 \\ \boldsymbol{\Pi}_{i-2}' & \boldsymbol{\Pi}_{i-1}' & \boldsymbol{I}_{im^2} & \dots & \boldsymbol{\Pi}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Pi}_1' & \boldsymbol{\Pi}_2' & \boldsymbol{\Pi}_3' & \dots & \boldsymbol{I}_{im^2} \end{pmatrix}_{(im)^2} \tag{8}$$

where $\boldsymbol{\Pi}_{i-j}$ is a diagonal im^2 matrix with $\boldsymbol{\omega}_{i-j}$ in the main diagonal,

$$\boldsymbol{\omega}_{i-j} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{I}_{m(i-j)} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}_{im} \text{ and } j = 1, 2, \dots, T_* - 1 \tag{9}$$

Moreover, when $j \geq i$, $\boldsymbol{\omega}_{i-j}$ is an im zero-matrix. This particular composition of $\bar{\boldsymbol{\Pi}}$ is inherited from the structure of $\boldsymbol{\Psi}_f$. Consequently, $\sqrt{T_*}vec(\hat{\bar{\boldsymbol{\beta}}}|\bar{\boldsymbol{Z}}_{\boldsymbol{p}}) \overset{d}{\to} N\big(\boldsymbol{0}, \bar{\boldsymbol{\Pi}}\big)$. ■

Proof of Proposition 2. Let $(\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\hat{\boldsymbol{O}} = (\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\boldsymbol{Z}_f\boldsymbol{Z}_p'(\boldsymbol{Z}_p\boldsymbol{Z}_p')^{-\frac{1}{2}}$, which becomes $(\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\hat{\boldsymbol{O}} = (\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}(\boldsymbol{O}\boldsymbol{M}\boldsymbol{Z}_p + \boldsymbol{\Psi}_f)\boldsymbol{Z}_p'(\boldsymbol{Z}_p\boldsymbol{Z}_p')^{-\frac{1}{2}}$ under the null. Substituting $\boldsymbol{M} = (\boldsymbol{Z}_p\boldsymbol{Z}_p')^{-\frac{1}{2}}$ and vectorizing, we get $vec[(\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}(\hat{\boldsymbol{O}} - \boldsymbol{O})] = \big[\big((\boldsymbol{Z}_p\boldsymbol{Z}_p)^{-\frac{1}{2}}\boldsymbol{Z}_p'\big) \otimes (\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\big]vec\boldsymbol{\Psi}_f$.

The covariance matrix of $vec[(\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}(\hat{\boldsymbol{O}} - \boldsymbol{O})]$ conditional to $\boldsymbol{Z}_p$ is written $\mathrm{E}\Big[\big[\big((\boldsymbol{Z}_p\boldsymbol{Z}_p')^{-\frac{1}{2}}\boldsymbol{Z}_p\big) \otimes (\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\big]vec\boldsymbol{\Psi}_f vec\boldsymbol{\Psi}_f'\big[\big(\boldsymbol{Z}_p'(\boldsymbol{Z}_p\boldsymbol{Z}_p')^{-\frac{1}{2}}\big) \otimes (\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\big]|\boldsymbol{Z}_p\Big]$. By replacing $(\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}} = (\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}$ and using that, under H_0, $\boldsymbol{Z}_f|\boldsymbol{Z}_p = \boldsymbol{Z}_f$, the covariance becomes $\big[\big((\boldsymbol{Z}_p\boldsymbol{Z}_p')^{-\frac{1}{2}}\boldsymbol{Z}_p\big) \otimes (\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\big](\boldsymbol{\Omega} \otimes \boldsymbol{Q})\big[\big(\boldsymbol{Z}_p'(\boldsymbol{Z}_p\boldsymbol{Z}_p')^{-\frac{1}{2}}\big) \otimes (\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\big]$. Again under the null hypothesis, $\sqrt{T_*}(\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}} \overset{a.s.}{\to} \boldsymbol{I}_i \otimes \boldsymbol{\Gamma}^{-\frac{1}{2}}$ and $\sqrt{T_*}(\boldsymbol{Z}_p\boldsymbol{Z}_p')^{-\frac{1}{2}} \overset{a.s.}{\to} \boldsymbol{I}_i \otimes \boldsymbol{\Gamma}^{-\frac{1}{2}}$ hold. Using the properties of the Kronecker

product, we can finally write $cov[vec((\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\hat{\boldsymbol{O}})] \overset{a.s.}{\to} T_*^{-2}\Big[\big[\big((\boldsymbol{I}_i\otimes\boldsymbol{\Gamma}^{-\frac{1}{2}})\boldsymbol{Z}_p\big)\otimes \boldsymbol{I}_i\big]\boldsymbol{\Omega}\big[\big(\boldsymbol{Z}_p'(\boldsymbol{I}_i\otimes\boldsymbol{\Gamma}^{-\frac{1}{2}})\big)\otimes\boldsymbol{I}_i\big]\Big]\otimes\boldsymbol{I}_m$.

On the other hand, the covariance of $vec(\hat{\boldsymbol{\beta}}|\boldsymbol{Z}_p)$ is $\bar{\boldsymbol{H}}^{-1}(\bar{\boldsymbol{Z}}_p\otimes\boldsymbol{I}_{im})(\boldsymbol{\Omega}\otimes\boldsymbol{I}_m)(\bar{\boldsymbol{Z}}_p'\otimes \boldsymbol{I}_{im})\bar{\boldsymbol{H}}'^{-1} \overset{a.s.}{\to} T_*^{-1}\bar{\boldsymbol{\Pi}}$. Finally, as $\lim_{T\to\infty}|\bar{\boldsymbol{Z}}_{\boldsymbol{p}} - (\boldsymbol{I}_i\otimes\boldsymbol{\Gamma}^{-\frac{1}{2}})\boldsymbol{Z}_p| = \boldsymbol{0}$, then both, $vec(\hat{\bar{\boldsymbol{\beta}}}|\boldsymbol{Z}_p)$ and $cov[vec((\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\hat{\boldsymbol{O}})]$, tend asymptotically to $T_*^{-1}\bar{\boldsymbol{\Pi}}$. ■

Proof of Proposition 3. As matrix $\bar{\boldsymbol{\Pi}}$ is known, it is straightforward to see that not all the elements in $\boldsymbol{A}$ are independent, except when $i=1$, that implies $\bar{\boldsymbol{\Pi}} = \boldsymbol{I}_{m^2}$. Given the structure of $\bar{\boldsymbol{\Pi}}$ and using the submatrix Matlab notation: (i) The first im elements of $vec\boldsymbol{A}$, which are $\boldsymbol{A}_{1:im,1:m}$, are uncorrelated as the square submatrix $\bar{\boldsymbol{\Pi}}_{1:im} = \boldsymbol{I}_{im^2}$, and (ii) as the first m rows of $\bar{\boldsymbol{\Pi}}_{i-1}'$ are zeros, then the elements of the submatrix $\boldsymbol{A}_{1:m,m+1:m+2}$ are also uncorrelated with those of $\boldsymbol{A}_{1:im,1:m}$. This occurs for every element in the submatrix $\boldsymbol{A}_{1:m,m+1:im}$ due to the structure of zeros in $\bar{\boldsymbol{\Pi}}_{i-k}'$, $k=1,2,\ldots,i-1$. Then the elements in $\boldsymbol{A}_{1:m,m+1:im}$ are uncorrelated with those of $\boldsymbol{A}_{1:im,1:m}$ and, therefore, $\bar{\boldsymbol{\Pi}}$ is of rank $m^2(2i-1)$. In order to extract m^2k independent elements from $\boldsymbol{A}$, we use the singular value decomposition (SVD) of $\bar{\boldsymbol{\Pi}}$, yielding a matrix $\boldsymbol{B}_{(im)^2\times m^2k}$ such that $\bar{\boldsymbol{\Pi}} \overset{\text{svd}}{=} \boldsymbol{U}\boldsymbol{S}^{\frac{1}{2}}\boldsymbol{S}^{\frac{1}{2}}\boldsymbol{V}' = \boldsymbol{B}\boldsymbol{B}'$. Consequently, we have $\boldsymbol{B}^{\dagger}\bar{\boldsymbol{\Pi}}\boldsymbol{B}'^{\dagger} = \boldsymbol{I}_{m^2k}$, where '†' denotes the Moore-Penrose pseudo inverse, and $\boldsymbol{B}^{\dagger}vec(\boldsymbol{A}) \overset{d}{\to} N\big(\boldsymbol{0}, T_*^{-1}\boldsymbol{I}_{m^2k}\big)$ which leads to $\mathcal{S}_{\boldsymbol{A}} = T_*vec(\boldsymbol{A})'\boldsymbol{P}vec(\boldsymbol{A}) \overset{d}{\to} \chi^2_{m^2k}$, $\boldsymbol{P} = \boldsymbol{B}'^{\dagger}\boldsymbol{B}^{\dagger}$ being a symmetric idempotent matrix of rank m^2k. ■

Proof of Proposition 4. Let the rth autocovariance matrix of the innovations be $\boldsymbol{C}_r = T^{-1}\boldsymbol{\psi}_t\boldsymbol{\psi}_{t-r}'$ and the rth residual autocovariance matrix be $\hat{\boldsymbol{C}}_r = T^{-1}\hat{\boldsymbol{\psi}}_t\hat{\boldsymbol{\psi}}_{t-r}'$. Further, define $\boldsymbol{C} = (\boldsymbol{C}_1\,\boldsymbol{C}_2\,\ldots\,\boldsymbol{C}_k)$ and similary $\hat{\boldsymbol{C}}$. (Hosking 1980) proved that $vec(\hat{\boldsymbol{C}}) = \boldsymbol{D}vec(\boldsymbol{C})$ where $\boldsymbol{D}$ is idempotent of rank $m^2(k-p-q)$. Let $\hat{\bar{\boldsymbol{\beta}}}_{\boldsymbol{*}}$ be as in (5) but using $\bar{\boldsymbol{z}}_t$ instead of $\boldsymbol{z}_t$ and assuming that $\bar{\boldsymbol{z}}_t$ are the standardized residuals from a VARMA(p,q) model. In such a case, $\hat{\bar{\boldsymbol{\beta}}}_{\boldsymbol{*}} \overset{a.s.}{\to} \hat{\mathbb{C}}(\boldsymbol{I}_i\otimes\boldsymbol{I}_m)^{-1} = \hat{\mathbb{C}}$ where:

$$\hat{\mathbb{C}} = \begin{pmatrix} \hat{\boldsymbol{C}}_{\bar{k}-i+1} & \hat{\boldsymbol{C}}_{\bar{k}-i} & \cdots & \hat{\boldsymbol{C}}_1 \\ \hat{\boldsymbol{C}}_{\bar{k}-i+2} & \hat{\boldsymbol{C}}_{\bar{k}-i+1} & \cdots & \hat{\boldsymbol{C}}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\boldsymbol{C}}_{\bar{k}} & \hat{\boldsymbol{C}}_{\bar{k}-1} & \cdots & \hat{\boldsymbol{C}}_{\bar{k}-i+1} \end{pmatrix}_{im} \quad \text{with } \bar{k} \equiv \begin{cases} k \text{ if } k \text{ is odd} \\ k+1 \text{ if } k \text{ is even.} \end{cases} \tag{10}$$

Then, we can write $\boldsymbol{B}^{\dagger}vec(\hat{\bar{\boldsymbol{\beta}}}_*) = \bar{\boldsymbol{D}}\boldsymbol{B}^{\dagger}vec(\hat{\bar{\boldsymbol{\beta}}})$ as it was done by (Hosking 1980), since $\boldsymbol{B}^{\dagger}vec(\hat{\bar{\boldsymbol{\beta}}}_*)$ and $\boldsymbol{B}^{\dagger}vec(\hat{\bar{\boldsymbol{\beta}}})$ have, asymptotically, the same elements as $vec(\hat{\boldsymbol{C}})$ and $vec(\boldsymbol{C})$, respectively, but sorted in different order. Likewise, $\bar{\boldsymbol{D}}$ has the same rows as $\boldsymbol{D}$, but ordered differently, that yields $rank(\bar{\boldsymbol{D}}) = rank(\boldsymbol{D}) = m^2(k-p-q)$. Finally, we previously showed that $\boldsymbol{B}^{\dagger}vec(\hat{\bar{\boldsymbol{\beta}}}|\boldsymbol{Z}_p) \overset{d}{\to} N\big(\boldsymbol{0}, T_*^{-1}\boldsymbol{I}_{m^2k}\big)$ and, con-

sequently, $\boldsymbol{B}^{\dagger} vec(\hat{\bar{\boldsymbol{\beta}}}_{*}|\boldsymbol{Z}_p) \xrightarrow{d} N\big(\boldsymbol{0}, T_{*}^{-1}\bar{\boldsymbol{D}}\big)$, which leads to $T_{*} vec(\hat{\bar{\boldsymbol{\beta}}}_{*})' \boldsymbol{P} vec(\hat{\bar{\boldsymbol{\beta}}}_{*}) \xrightarrow{d} \chi^2_{m^2(k-p-q)}$. ■

Revista Colombiana de Estadística
Diciembre 2013, volumen 36, no. 2, pp. 239 a 260

Testing Equality of Several Correlation Matrices

Prueba de Igualdad de Varias Matrices de Correlación

Arjun K. Gupta[1,a], Bruce E. Johnson[2,b], Daya K. Nagar[3,c]

[1]Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, Ohio 43403-0221, USA

[2]Experient Research Group, 471 Severnside Drive, Severna Park, MD 21146, USA

[3]Instituto de Matemáticas, Facultad de Ciencias Exactas y Naturales, Universidad de Antioquia, Medellín, Colombia

Abstract

In this article we show that the Kullback's statistic for testing equality of several correlation matrices may be considered a modified likelihood ratio statistic when sampling from multivariate normal populations. We derive the asymptotic null distribution of L^* in series involving independent chi-square variables by expanding L^* in terms of other random variables and then inverting the expansion term by term. An example is also given to exhibit the procedure to be used when testing the equality of correlation matrices using the statistic L^*.

***Key words*:** Asymptotic null distribution, Correlation matrix, Covariance matrix, Cumulants, Likelihood ratio test.

Resumen

En este artículo se muestra que el estadístico L^* de Kullback , para probar la igualdad de varias matrices de correlación, puede ser considerado como un estadístico modificado del test de razón de verosimilitud cuando se muestrean poblaciones normales multivariadas. Derivamos la distribución asintótica nula de L^* en series que involucran variables independientes chi-cuadrado, mediante la expansión de L^* en términos de otras variables aleatorias y luego invertir la expansión término a término. Se da también un ejemplo para mostrar el procedimiento a ser usado cuando se prueba igualdad de matrices de correlación mediante el estadístico L^*.

***Palabras clave*:** distribución asintótica nula, matriz de correlación, matriz de covarianza, razón de verosimilitud.

[a]Professor. E-mail: gupta@bgsu.edu
[b]Researcher. E-mail: bruce.johnson@experientresearch.com
[c]Professor. E-mail: dayaknagar@yahoo.com

1. Introduction

The correlation matrix is one of the foundations of factor analysis and has found its way into such diverse areas as economics, medicine, physical science and political science. There is a fair amount of literature on testing properties of correlation matrices. Tests for certain structures in a correlation matrix have been proposed and studied by several authors, e.g, see Aitkin, Nelson, and Reinfurt (1968), Gleser (1968), Aitkin (1969), Modarres (1993), Kullback (1997) and Schott (2007). In a series of papers, Konishi (1978, 1979*a*, 1979*b*) has developed asymptotic expansions of correlation matrix and applied them to various problems of multivariate analysis. The exact distribution of the correlation matrix, when sampling from a multivariate Gaussian population, is derived in Ali, Fraser and Lee (1970) and Gupta and Nagar (2000).

If the covariance matrix of α-th population is given by Σ_α and Δ_α is a diagonal matrix of standard deviations for the population α, then $P_\alpha = \Delta_\alpha^{-1}\Sigma_\alpha\Delta_\alpha^{-1}$ is the correlation matrix for the population α. The null hypothesis that all k populations have the same correlation matrices may be stated as $H : P_1 = \cdots = P_k$.

Let the vectors $\boldsymbol{x}_{\alpha 1}, \boldsymbol{x}_{\alpha 2}, \ldots, \boldsymbol{x}_{\alpha N_\alpha}$ be a random sample of size $N_\alpha = n_\alpha + 1$ for $\alpha = 1, 2, \ldots, k$ from k multivariate populations of dimensionality p. Further, we assume the independence of these k samples. Let $\overline{\boldsymbol{x}}_\alpha = \sum_{i=1}^{N_\alpha} \boldsymbol{x}_{\alpha i}/N_\alpha$, $A_\alpha = \sum_{i=1}^{N_\alpha}(\boldsymbol{x}_{\alpha i} - \overline{\boldsymbol{x}}_\alpha)(\boldsymbol{x}_{\alpha i} - \overline{\boldsymbol{x}}_\alpha)'$ and $S_\alpha = A_\alpha/N_\alpha$. Further, let D_α be a diagonal matrix of the square roots of the diagonal elements of S_α. The sample correlation matrix R_α is then defined by $R_\alpha = D_\alpha^{-1} S_\alpha D_\alpha^{-1}$. Let $n = \sum_{\alpha=1}^{k} n_\alpha$ and $\overline{R} = \sum_{\alpha=1}^{k} n_\alpha R_\alpha$.

Kullback (1967) derived the statistic $L^* = \sum_{\alpha=1}^{k} n_\alpha \ln\{\det(\overline{R})/\det(R_\alpha)\}$ for testing the equality of k correlation matrices based on samples from multivariate populations. This statistic was later examined by Jennrich (1970) who observed that the statistic proposed by Kullback failed to have chi-square distribution ascribed to it. For further results on this topic the reader is referred to Browne (1978) and Modarres and Jernigan (1992).

Although the Kullback's statistic L^* is not equal to the modified likelihood ratio criterion, we here show that it may be considered an approximation of the modified likelihood ratio statistic when sampling from multivariate normal populations.

In Section 2, we show that Kullback's statistic can be viewed as an approximation of the modified likelihood ratio statistic based on samples from multivariate normal populations. Section 3 deals with some preliminary results and definitions which are used in subsequent sections. In sections 4 and 5, we obtain asymptotic null distribution of L^* by expanding L^* in terms of other random variables and then inverting the expansion term by term. Finally, in Section 6, an example is given to demonstrate the procedure to be used when testing the equality of correlation matrices using the statistic L^*. Some results on matrix algebra and distribution theory are given in the Appendix.

2. The Test Statistic

In this section, we give an approximation of the likelihood ratio test statistic λ for testing equality of correlation matrices of several multivariate Gaussian populations. The test statistic λ was derived and studied by Cole (1968*a*, 1968*b*) in two unpublished technical reports (see Browne 1978, Modarres and Jernigan 1992, 1993). However, these reports are scarcely available, and therefore the sake of completeness and for a better understanding it seems appropriate to first give a concise step-by-step derivation of the test statistic λ.

If the underlying populations follow multivariate normal distributions, then the likelihood function based on the k independent samples, when all parameters are unrestricted, is given by

$$\begin{aligned} &L(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k, \Sigma_1, \ldots, \Sigma_k) \\ &\quad = \prod_{\alpha=1}^{k} \left[(2\pi)^{pN_\alpha/2} \det(\Sigma_\alpha)^{N_\alpha/2} \right]^{-1} \\ &\qquad \times \exp\left[-\frac{1}{2} \sum_{\alpha=1}^{k} \operatorname{tr}\left(\Sigma_\alpha^{-1} A_\alpha\right) - \frac{1}{2} \sum_{\alpha=1}^{k} \operatorname{tr}\left\{ \Sigma_\alpha^{-1} (\bar{\boldsymbol{x}}_\alpha - \boldsymbol{\mu}_\alpha)(\bar{\boldsymbol{x}}_\alpha - \boldsymbol{\mu}_\alpha)' \right\} \right] \end{aligned}$$

where for $\alpha = 1, \ldots, k$ we have $\boldsymbol{\mu}_\alpha \in \mathbb{R}^p$ and $\Sigma_\alpha > 0$. It is well known that for any fixed value of Σ_α the likelihood function is maximized with respect to the $\boldsymbol{\mu}_\alpha$'s when $\widehat{\boldsymbol{\mu}}_\alpha = \overline{\boldsymbol{x}}_\alpha$.

Let Δ_α be a diagonal matrix of standard deviations for the population α. Further, let $P_\alpha = \Delta_\alpha^{-1} \Sigma_\alpha \Delta_\alpha^{-1}$ be the population correlation matrix for the population α. The natural logarithm of the likelihood function, after evaluation at $\widehat{\boldsymbol{\mu}}_\alpha = \overline{\boldsymbol{x}}_\alpha$, may then be written as

$$\begin{aligned} &\ln[L(\bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_k, \Delta_1 P_1 \Delta_1, \ldots, \Delta_k P_k \Delta_k)] \\ &\quad = -\frac{1}{2} Np \ln(2\pi) - \frac{1}{2} \sum_{\alpha=1}^{k} N_\alpha \ln[\det(P_\alpha \Delta_\alpha^2)] - \frac{1}{2} \sum_{\alpha=1}^{k} \operatorname{tr}(N_\alpha P_\alpha^{-1} G_\alpha R_\alpha G_\alpha) \end{aligned}$$

where $N = \sum_{\alpha=1}^{k} N_\alpha$ and $G_\alpha = \Delta_\alpha^{-1} D_\alpha$. Further, when the parameters are unrestricted, the likelihood function $L(\overline{\boldsymbol{x}}_1, \ldots, \overline{\boldsymbol{x}}_k, \Delta_1 P_1 \Delta_1, \ldots, \Delta_k P_k \Delta_k)$ is maximized when $-\ln[\det(P_\alpha \Delta_\alpha^2)] - \operatorname{tr}\left(P_\alpha^{-1} G_\alpha R_\alpha G_\alpha\right)$ is maximized for each α. This is true when

$$\begin{aligned} &\ln[\det(P_\alpha \Delta_\alpha^2)] + \operatorname{tr}\left(P_\alpha^{-1} G_\alpha R_\alpha G_\alpha\right) \\ &\qquad = \ln[\det(\Delta_\alpha P_\alpha \Delta_\alpha)] + \operatorname{tr}\left(\Delta_\alpha^{-1} P_\alpha^{-1} \Delta_\alpha^{-1} D_\alpha R_\alpha D_\alpha\right) \end{aligned}$$

is minimized for each α. This is achieved when $\Delta_\alpha P_\alpha \Delta_\alpha = D_\alpha R_\alpha D_\alpha$. From this it follows that the maximum value of $L(\overline{\boldsymbol{x}}_1, \ldots, \overline{\boldsymbol{x}}_k, \Delta_1 P_1 \Delta_1, \ldots, \Delta_k P_k \Delta_k)$, when

the parameters are unrestricted, is given by

$$\ln[L(\bar{\boldsymbol{x}}_1,\ldots,\bar{\boldsymbol{x}}_k,D_1R_1D_1,\ldots,D_kR_kD_k)] = -\frac{1}{2}Np[\ln(2\pi)+1]-\frac{1}{2}\sum_{\alpha=1}^{k}N_\alpha\ln[\det(R_\alpha D_\alpha^2)]. \quad (1)$$

Let P be the common value of the population correlation matrices under the null hypothesis of equality of correlation matrices. The reduced parameter space for the covariance matrices is the set of all covariance matrices that may be written as $\Delta_\alpha P$ where P is a correlation matrix and Δ_α is a diagonal matrix with positive elements on the diagonal. The restricted log likelihood function is written as

$$\ln[L(\bar{\boldsymbol{x}}_1,\ldots,\bar{\boldsymbol{x}}_k,\Delta_1P,\ldots,\Delta_kP)] = -\frac{1}{2}Np\ln(2\pi)-\frac{1}{2}\sum_{\alpha=1}^{k}N_\alpha\ln[\det(P\Delta_\alpha^2)]-\frac{1}{2}\sum_{\alpha=1}^{k}N_\alpha\operatorname{tr}\left(P^{-1}G_\alpha R_\alpha G_\alpha\right).$$

Let $P^{-1}=(\rho^{ij})$. Since Δ_α is a diagonal matrix,

$$\ln[\det(\Delta_\alpha)^2]=2\ln[\det(\Delta_\alpha)]=2\ln\left[\prod_{i=1}^{p}\sigma_{\alpha ii}\right]=2\sum_{i=1}^{p}\ln(\sigma_{\alpha ii})$$

Also, since $G_\alpha=\Delta_\alpha^{-1}D_\alpha$ is a diagonal matrix, we have

$$\operatorname{tr}\left(P^{-1}G_\alpha R_\alpha G_\alpha\right)=\sum_{i=1}^{p}\sum_{j=1}^{p}\rho^{ij}g_{\alpha j}r_{\alpha ij}g_{\alpha i}$$

Thus,

$$\ln[L(\bar{\boldsymbol{x}}_1,\ldots,\bar{\boldsymbol{x}}_k,\Delta_1P,\ldots,\Delta_kP)] = -\frac{1}{2}Np\ln(2\pi)-\frac{1}{2}\sum_{\alpha=1}^{k}N_\alpha\sum_{i=1}^{p}\ln(\sigma_{\alpha ii})-\frac{1}{2}\sum_{\alpha=1}^{k}N_\alpha\ln[\det(P)] - \frac{1}{2}\sum_{\alpha=1}^{k}N_\alpha\sum_{i=1}^{p}\sum_{j=1}^{p}\rho^{ij}g_{\alpha j}r_{\alpha ij}g_{\alpha i}$$

Since, $g_{\alpha i}=s_{\alpha ii}/\sigma_{\alpha ii}$, differentiation of $\ln[L(\overline{\boldsymbol{x}}_1,\ldots,\overline{\boldsymbol{x}}_k,\Delta_1P,\ldots,\Delta_kP)]$ with respect to $\sigma_{\alpha ii}$ yields

$$\frac{\partial}{\partial\sigma_{\alpha ii}}\ln[L(\overline{\boldsymbol{x}}_1,\ldots,\overline{\boldsymbol{x}}_k,\Delta_1P,\ldots,\Delta_kP)]=-\frac{N_\alpha}{2\sigma_{\alpha ii}}+\frac{N_\alpha}{2\sigma_{\alpha ii}}\sum_{j=1}^{p}g_{\alpha i}g_{\alpha j}\rho^{ij}r_{\alpha ij}$$

Further, setting this equal to zero gives $\sum_{j=1}^{p}g_{\alpha i}g_{\alpha j}\rho^{ij}r_{\alpha ij}-1=0$. Differentiating $\ln[L(\overline{\boldsymbol{x}}_1,\ldots,\overline{\boldsymbol{x}}_k,\Delta_1P,\ldots,\Delta_kP)]$ with respect to the matrix P using Lemma 6, we obtain

$$\frac{\partial}{\partial P}\ln[L(\overline{\boldsymbol{x}}_1,\ldots,\overline{\boldsymbol{x}}_k,\Delta_1P,\ldots,\Delta_kP)]=-\frac{1}{2}NP^{-1}+\frac{1}{2}\sum_{\alpha=1}^{k}N_\alpha P^{-1}G_\alpha R_\alpha G_\alpha P^{-1}$$

Setting this equal to zero, multiplying by 2, pre and post multiplying by P and dividing by N gives $P = \sum_{\alpha=1}^{k} N_\alpha G_\alpha R_\alpha G_\alpha / N$ so that $\sum_{\alpha=1}^{k} N_\alpha g_{\alpha i}^2 / N = 1$.

The likelihood ratio test statistic λ for testing $H : P_1 = \cdots = P_k$ is now derived as

$$\lambda = \prod_{\alpha=1}^{k} \frac{\det(R_\alpha D_\alpha^2)^{N_\alpha/2}}{\det(\widehat{P}\widehat{\Delta}_\alpha^2)^{N_\alpha/2}}$$

where $\widehat{P}$ and $\widehat{\Delta}_\alpha^2$ are solutions of $\widehat{P} = \sum_{\alpha=1}^{k} N_\alpha \widehat{\Delta}_\alpha^{-1} S_\alpha \widehat{\Delta}_\alpha^{-1} / N$ and $\sum_{j=1}^{p} \rho^{ij} s_{\alpha ij} - 1 = 0$, $i = 1, \ldots, p$, respectively.

To obtain an approximation of the likelihood ratio statistic we replace $\sigma_{\alpha ii}$ by its consistent estimator $\widehat{\sigma}_{\alpha ii}$. Then, it follows that $\widehat{g}_{\alpha ii} = s_{\alpha ii}/\widehat{\sigma}_{\alpha ii}$ and $\widehat{G}_\alpha = \operatorname{diag}(\widehat{g}_{\alpha 1}, \ldots, \widehat{g}_{\alpha p})$, and the estimator of P is given by $\widehat{P} = \sum_{\alpha=1}^{k} N_\alpha \widehat{G}_\alpha R_\alpha \widehat{G}_\alpha / N$. Thus, an approximation of the maximum of $\ln[L(\overline{\boldsymbol{x}}_1, \ldots, \overline{\boldsymbol{x}}_k, \Delta_1 P, \ldots, \Delta_k P)]$ is given as

$$-\frac{1}{2} Np[\ln(2\pi) + 1] - \frac{1}{2} \sum_{\alpha=1}^{k} N_\alpha \ln[\det(\widehat{\Delta}_\alpha)^2] - \frac{1}{2} N \ln[\det(\widehat{P})] \tag{2}$$

As the sample size goes to infinity, $s_{\alpha ii}/\widehat{\sigma}_{\alpha ii}$ converges in probability to 1 so that $\widehat{G}_\alpha$ converges in probability to I_p. This suggest further approximation of (2) as

$$-\frac{1}{2} Np[\ln(2\pi) + 1] - \frac{1}{2} \sum_{\alpha=1}^{k} N_\alpha \ln[\det(D_\alpha)^2] - \frac{1}{2} N \ln\left[\det\left(\sum_{\alpha=1}^{k} \frac{N_\alpha}{N} R_\alpha\right)\right] \tag{3}$$

Now, using (1) and (3), the likelihood ratio statistic is approximated as

$$\tilde{\lambda} = \frac{\prod_{\alpha=1}^{k} \det(R_\alpha)^{N_\alpha/2}}{\det(\sum_{\alpha=1}^{k} N_\alpha R_\alpha / N)^{N/2}} \tag{4}$$

Further, replacing N_α by n_α above, an approximated modified likelihood ratio statistic is derived as

$$M = \frac{\prod_{\alpha=1}^{k} \det(R_\alpha)^{n_\alpha/2}}{\det(\sum_{\alpha=1}^{k} n_\alpha R_\alpha / n)^{n/2}} = \frac{\prod_{\alpha=1}^{k} \det(R_\alpha)^{n_\alpha/2}}{\det(\overline{R})^{n/2}} \tag{5}$$

Since $-2 \ln M = \sum_{\alpha=1}^{k} n_\alpha \ln\{\det(\overline{R}) / \det(R_\alpha)\} = L^*$, the statistic proposed by Kullback may be thought of as an approximated modified likelihood ratio statistic.

3. Preliminaries

Let the vectors $\boldsymbol{x}_{\alpha 1}, \ldots, \boldsymbol{x}_{\alpha N_\alpha}$ be a random sample of size n_α for $\alpha = 1, \ldots, k$ from k multivariate populations of dimensionality p and finite fourth moments. The characteristic function for the population α is given by $\phi_\alpha^*(\boldsymbol{t}) = E[\exp(\iota \boldsymbol{t}' \boldsymbol{x})]$

where $\iota = \sqrt{-1}$ and $\boldsymbol{t} = (t_1, \ldots, t_p)'$. The log characteristic function for population α may be written as

$$\ln[\phi^*_\alpha(\boldsymbol{t})] = \sum_{r_1+\cdots+r_p=1}^{\infty} \kappa^*_\alpha(r_1, \ldots, r_p) \prod_{j=1}^{p} \frac{(\iota t_j)^{r_j}}{r_j!}, \quad r_j \in \boldsymbol{I}^+ \tag{6}$$

where $\boldsymbol{I}^+$ is the set of non-negative integers. The cumulants of the distribution are the coefficients $\kappa^*_\alpha(r_1, \ldots, r_p)$. If $r_1 + \cdots + r_p = m$, then the associated cumulant is of order m. The relationship between the cumulants of a distribution and the characteristic function provide a convenient method for deriving the asymptotic distribution of statistic whose asymptotic expectations can be derived.

The cumulants of order m are functions of the moments of order m or lower. Thus if the m^{th} order moment is finite, so is the m^{th} order cumulant. Let $\mu_i = E(X_i)$, $\mu_{ij} = E(X_iX_j)$, $\mu_{ijk} = E(X_iX_jX_k)$, and $\mu_{ijk\ell} = E(X_iX_jX_kX_\ell)$ and κ_i, κ_{ij}, κ_{ijk}, and $\kappa_{ijk\ell}$ be the corresponding cumulants. Then, Kaplan (1952) gives the following relationship:

$$\begin{aligned}
\kappa_i &= \mu_i, \\
\kappa_{ij} &= \mu_{ij} - \mu_i\mu_j, \\
\kappa_{ijk} &= \mu_{ijk} - (\mu_i\mu_{jk} + \mu_j\mu_{ik} + \mu_k\mu_{ij}) + 2\mu_i\mu_j\mu_k, \\
\kappa_{ijk\ell} &= \mu_{ijk\ell} - \sum^{4} \mu_i\mu_{jk\ell} - \sum^{3} \mu_{ij}\mu_{k\ell} + 2\sum^{6} \mu_i\mu_j\mu_{k\ell} - 6\mu_i\mu_j\mu_k\mu_\ell
\end{aligned}$$

where the summations are over the possible ways of grouping the subscripts, and the number of terms resulting is written over the summation sign.

Define the random matrix V_α as

$$V_\alpha = \sqrt{n_\alpha}\left(\frac{1}{n_\alpha}\Delta_\alpha^{-1}A_\alpha\Delta_\alpha^{-1} - P_\alpha\right) \tag{7}$$

Then, the random matrices $V_\alpha^{(0)}$, $V_\alpha^{(1)}$ and $V_\alpha^{(2)}$ are defined as

$$V_\alpha^{(0)} = \text{diag}(v_{\alpha 11}, v_{\alpha 22}, \ldots, v_{\alpha pp}) \tag{8}$$

$$V_\alpha^{(1)} = V_\alpha - \frac{1}{2}V_\alpha^{(0)}P_\alpha - \frac{1}{2}P_\alpha V_\alpha^{(0)} \tag{9}$$

and

$$V_\alpha^{(2)} = \frac{1}{4}V_\alpha^{(0)}P_\alpha V_\alpha^{(0)} - \frac{1}{2}V_\alpha V_\alpha^{(0)} - \frac{1}{2}V_\alpha^{(0)}V_\alpha + \frac{3}{8}(V_\alpha^{(0)})^2P_\alpha + \frac{3}{8}P_\alpha(V_\alpha^{(0)})^2 \tag{10}$$

Konishi (1979*a*, 1979*b*) has shown that

$$R_\alpha = P_\alpha + \frac{1}{\sqrt{n_\alpha}}V_\alpha^{(1)} + \frac{1}{n_\alpha}V_\alpha^{(2)} + O_p(n_\alpha^{-3/2})$$

The pooled estimate of the common correlation matrix is

$$\overline{R} = \sum_{\alpha=1}^{k} \omega_\alpha R_\alpha$$

so that

$$\overline{R} = \overline{P} + \frac{1}{\sqrt{n}}\overline{V}^{(1)} + \frac{1}{n}\overline{V}^{(2)} + O_p(n^{-3/2})$$

where $\omega_\alpha = n_\alpha/n$, $\overline{P} = \sum_{\alpha=1}^{k} \omega_\alpha P_\alpha$, $\overline{V}^{(1)} = \sum_{\alpha=1}^{k} \sqrt{\omega_\alpha}\ V_\alpha^{(1)}$ and $\overline{V}^{(2)} = \sum_{\alpha=1}^{k} V_\alpha^{(2)}$. The limiting distribution of $V_\alpha = \sqrt{n_\alpha}\left(\Delta_\alpha^{-1} A_\alpha \Delta_\alpha^{-1}/n_\alpha - P_\alpha\right)$ is normal with means 0 and covariances that depend on the fourth order cumulants of the parent population (Anderson 2003, p. 88).

Since Δ_α is a diagonal matrix of population standard deviations, $\Delta_\alpha^{-1}\boldsymbol{x}_{\alpha 1}, \ldots, \Delta_\alpha^{-1}\boldsymbol{x}_{\alpha N_\alpha}$ may be thought of as N_α observations from a population with finite fourth order cumulants and characteristic function given by

$$\ln[\phi_\alpha(\boldsymbol{t})] = \sum_{r_1+\cdots+r_p=1}^{\infty} \kappa_\alpha(r_1, \ldots, r_p) \prod_{j=1}^{p} \frac{(\iota t_j)^{r_j}}{r_j!},\ r_j \in \boldsymbol{I}^+ \tag{11}$$

where the standardized cumulants, $\kappa_\alpha(r_1, r_2, \ldots, r_p)$, are derived from the expression (6) as

$$\kappa_\alpha(r_1, r_2, \ldots, r_p) = \frac{\kappa_\alpha^*(r_1, r_2, \ldots, r_p)}{\sigma_{\alpha 11}\chi_{r_1}\sigma_{\alpha 22}\chi_{r_2}\cdots\sigma_{\alpha pp}\chi_{r_p}}$$

with $\chi_{r_j} = 1$ if $r_j = 0$, $\chi_{r_j} = 1/\sigma^{(\alpha)jj}$ if $r_j \neq 0$ and $\Sigma_\alpha^{-1} = (\sigma^{(\alpha)jj})$.

K-statistics are unbiased estimates of the cumulants of a distribution, and may be used to derive the moments of the statistics which are symmetric functions of the observations (Kendall and Stuart 1969). Kaplan (1952) gives a series of tensor formulaes for computing the expectations of various functions of the k-statistics associated with a sample of size N from a multivariate population. For the definition of the k-statistics, let $N^{(r)} = N(N-1)\cdots(N-r+1)$.

If $s_{i_1 i_2 \cdots i_\ell}$ denotes the product $X_{i_1}X_{i_2}\cdots X_{i_\ell}$ summed over the sample, the tensor formulae for the k-statistics may be shown to be as follows:

$$k_i = \frac{s_i}{N}, \quad k_{ij} = \frac{Ns_{ij} - s_i s_j}{N^{(2)}}, \quad k_{ijk} = \frac{N^2 s_{ijk} - N\sum^{3} s_i s_{jk} + 2s_i s_j s_k}{N^{(3)}}$$

$$k_{ijk\ell} = \frac{N(N+1)(Ns_{ijk\ell} - \sum^{4} s_i s_{jk\ell}) - N^{(2)}\sum^{3} s_{ij}s_{k\ell} + 2N\sum^{6} s_i s_j s_{k\ell} - 6s_i s_j s_k s_\ell}{N^{(4)}}$$

$$\begin{aligned}\kappa(ab, ij) &= E[(k_{ab} - \kappa_{ab})(k_{ij} - \kappa_{ij})] \\ &= \frac{\kappa_{abij}}{N} + \frac{\kappa_{ai}\kappa_{bj} + \kappa_{aj}\kappa_{bi}}{N-1}\end{aligned}$$

$$\begin{aligned}\kappa(ab, ij, pq) &= E[(k_{ab} - \kappa_{ab})(k_{ij} - \kappa_{ij})(k_{pq} - \kappa_{pq})] \\ &= \frac{\kappa_{abijpq}}{N^2} + \sum^{12} \frac{\kappa_{abip}\kappa_{jq}}{N(N-1)} + \sum^{4} \frac{(N-2)\kappa_{aip}\kappa_{bjq}}{N(N-1)^2} + \sum^{8} \frac{\kappa_{ai}\kappa_{bp}\kappa_{jq}}{(N-1)^2}\end{aligned}$$

The summations are over the possible ways of grouping the subscripts, and the number of terms resulting is written over the summation sign.

The matrix V_α is constructed from observations from the standardized distribution so that $v_{\alpha ij} = \sqrt{n_\alpha}(k_{\alpha ij} - \rho_{\alpha ij})$ where $k_{\alpha ij}$ is the related k-statistic for standardized population α. Kaplan's formulae may be applied to derive the following expressions for the expectations of elements of the matrices V_α (note that $\kappa_{\alpha ij} = \rho_{\alpha ij}$). We obtain

$$E(v_{\alpha ij}) = 0$$

$$E(v_{\alpha ij}v_{\alpha k\ell}) = \kappa_{\alpha ijk\ell} + \rho_{\alpha ik}\rho_{\alpha j\ell} + \rho_{\alpha i\ell}\rho_{\alpha jk} + O(n_\alpha^{-1})$$

and

$$E(v_{\alpha ij}v_{\alpha k\ell}v_{\alpha ab}) = \frac{1}{\sqrt{n_\alpha}}\Bigg[\kappa_{\alpha ijk\ell ab} + \sum^{12}\kappa_{\alpha ijka}\rho_{\alpha\ell b} + \sum^{4}\kappa_{\alpha ika}\kappa_{\alpha j\ell b} + \sum^{8}\rho_{\alpha ik}\rho_{\alpha ja}\rho_{\alpha\ell b}\Bigg] + O(n_\alpha^{-3/2})$$

The random matrices $V_\alpha^{(0)}$, $V_\alpha^{(1)}$ and $V_\alpha^{(2)}$ are defined in (8), (9), and (10), respectively. The expectations associated with these random matrices are given as

$$E(v_{\alpha ij}^{(1)}) = 0$$

$$E(v_{\alpha ij}^{(2)}) = \frac{1}{4}\rho_{\alpha ij}\kappa_{\alpha iijj} - \frac{1}{2}(\kappa_{\alpha iiij} + \kappa_{\alpha ijjj}) + \frac{3}{8}\rho_{\alpha ij}(\kappa_{\alpha iiii} + \kappa_{\alpha jjjj}) + \frac{1}{2}(\rho_{\alpha ij}^3 - \rho_{\alpha ij}) + O(n_\alpha^{-1})$$

$$\begin{aligned} E(v_{\alpha ij}^{(1)}v_{\alpha k\ell}^{(1)}) = {} & \kappa_{\alpha ijk\ell} - \frac{1}{2}(\rho_{\alpha ij}\kappa_{\alpha iik\ell} + \rho_{\alpha ij}\kappa_{\alpha jjk\ell} + \rho_{\alpha k\ell}\kappa_{\alpha ijkk} + \rho_{\alpha k\ell}\kappa_{\alpha ij\ell\ell}) \\ & + \frac{1}{4}\rho_{\alpha ij}\rho_{\alpha k\ell}(\kappa_{\alpha iikk} + \kappa_{\alpha ii\ell\ell} + \kappa_{\alpha jjkk} + \kappa_{\alpha jj\ell\ell}) \\ & - (\rho_{\alpha k\ell}\rho_{\alpha ik}\rho_{\alpha jk} + \rho_{\alpha k\ell}\rho_{\alpha i\ell}\rho_{\alpha j\ell} + \rho_{\alpha ij}\rho_{\alpha ik}\rho_{\alpha i\ell} + \rho_{\alpha ij}\rho_{\alpha jk}\rho_{\alpha j\ell}) \\ & + \frac{1}{2}\rho_{\alpha ij}\rho_{\alpha k\ell}(\rho_{\alpha ik}^2 + \rho_{\alpha i\ell}^2 + \rho_{\alpha jk}^2 + \rho_{\alpha i\ell}^2) \\ & + (\rho_{\alpha ik}\rho_{\alpha j\ell} + \rho_{\alpha i\ell}\rho_{\alpha jk}) + O(n_\alpha^{-1}) \end{aligned} \tag{12}$$

and

$$E(v_{\alpha ij}^{(1)}v_{\alpha k\ell}^{(1)}v_{\alpha ab}^{(1)}) = \frac{1}{\sqrt{n_\alpha}}\left(t_{\alpha 1} - \frac{1}{2}t_{\alpha 2} + \frac{1}{4}t_{\alpha 3} - \frac{1}{8}t_{\alpha 4}\right) + O(n_\alpha^{-3/2})$$

where

$$t_{\alpha 1} = \kappa_{\alpha ijk\ell ab} + \sum^{12}\kappa_{\alpha ijka}\kappa_{\alpha\ell b} + \sum^{4}\kappa_{\alpha ika}\kappa_{\alpha i\ell b} + \sum^{8}\rho_{\alpha ik}\rho_{\alpha ja}\rho_{\alpha\ell b}$$

$$t_{\alpha 2} = \sum^{3} \rho_{\alpha ij} \Bigg[\kappa_{\alpha iik\ell ab} + \kappa_{\alpha jjk\ell a} + \sum^{12} (\kappa_{\alpha iika} + \kappa_{\alpha jjka}) + \sum^{3} (\kappa_{\alpha ika} \kappa_{\alpha i\ell b} + \kappa_{\alpha jka} \kappa_{\alpha j\ell b}) + \sum^{8} (\rho_{\alpha ik} \rho_{\alpha ia} \rho_{\alpha \ell b} + \rho_{\alpha jk} \rho_{\alpha ja} \rho_{\alpha \ell b}) \Bigg]$$

$$t_{\alpha 3} = \sum^{3} \rho_{\alpha ij} \rho_{\alpha k\ell} \Bigg[\kappa_{\alpha iikkab} + \kappa_{\alpha ii\ell\ell ab} + \kappa_{\alpha jjkkab} + \kappa_{\alpha jj\ell\ell ab} + \sum^{12} (\kappa_{\alpha iika} \rho_{\alpha kb} + \kappa_{\alpha ii\ell a} \rho_{\alpha \ell b} + \kappa_{\alpha jjka} \rho_{\alpha kb} + \kappa_{\alpha jj\ell a} \rho_{\alpha \ell b}) + \sum^{3} (\kappa_{\alpha ika} \kappa_{\alpha ikb} + \kappa_{\alpha i\ell a} \kappa_{\alpha i\ell b} + \kappa_{\alpha jka} \kappa_{\alpha jkb} + \kappa_{\alpha j\ell a} \kappa_{\alpha j\ell b}) + \sum^{8} (\rho_{\alpha ik} \rho_{\alpha ia} \rho_{\alpha bk} + \rho_{\alpha i\ell} \rho_{\alpha ia} \rho_{\alpha \ell b} + \rho_{\alpha jk} \rho_{\alpha ja} \rho_{\alpha bk} + \rho_{\alpha j\ell} \rho_{\alpha ja} \rho_{\alpha \ell b}) \Bigg]$$

and

$$t_{\alpha 4} = \rho_{\alpha ij} \rho_{\alpha k\ell} \rho_{\alpha ab} \sum^{8} \Bigg[\kappa_{\alpha iikkaa} + \sum^{12} (\kappa_{\alpha iika} \rho_{\alpha ka}) + \sum^{3} (\kappa_{\alpha ika} \kappa_{\alpha ikb}) + \sum^{8} (\rho_{\alpha ik} \rho_{\alpha ia} \rho_{\alpha ka}) \Bigg]$$

Lemma 1. *The diagonal elements of* $V_\alpha^{(1)}$ *are zero.*

Proof. Using (9) and the fact that $V_\alpha^{(0)}$ is a diagonal matrix, we have

$$v_{\alpha ij}^{(1)} = v_{\alpha ij} - \frac{1}{2} \rho_{\alpha ij} (v_{\alpha ii} + v_{\alpha jj})$$

The result follows by taking $j = i$ above and noting that diagonal elements of P_α are 1. □

Lemma 2. *The diagonal elements of* $V_\alpha^{(2)}$ *are zero.*

Proof. Using (10) and the fact that $V_\alpha^{(0)}$ is a diagonal matrix, we get

$$v_{\alpha ij}^{(2)} = \frac{1}{4} v_{\alpha ii}^{(0)} \rho_{\alpha ij} v_{\alpha jj}^{(0)} - \frac{1}{2} v_{\alpha ij} (v_{\alpha jj} + v_{\alpha ii}) + \frac{3}{8} \rho_{\alpha ij} (v_{\alpha ii}^2 + v_{\alpha jj}^2)$$

The result follows by substituting $j = i$ above and observing that $\rho_{\alpha ii} = 1$. □

4. Asymptotic Expansion of L^*

In order to derive the asymptotic distribution for L^* the statistic is first expanded in terms of other random variables (see Konishi and Sugiyama 1981).

The statistic L^* may be written as $L^* = ng(R_1 \ldots, R_k)$ where $g(R_1, \ldots, R_k) = \ln[\det(\overline{R})] - \sum_{\alpha=1}^{k} \omega_\alpha \ln[\det(R_\alpha)]$. Let

$$B_\alpha = \frac{1}{\sqrt{n_\alpha}} P_\alpha^{-1} V_\alpha^{(1)} + \frac{1}{n_\alpha} P_\alpha^{-1} V_\alpha^{(2)}$$

Since, P_α, $V_\alpha^{(1)}$ and $V_\alpha^{(2)}$ are all positive definite, so is B_α. This insures that the eigenvalues of B_α exist and are positive. Also, as n_α becomes large, the elements in B_α become small so that the characteristic roots may be assumed to be less than one. Using Lemma 5,

$$\begin{aligned}\ln[\det(R_\alpha)] &= \ln[\det(P_\alpha + P_\alpha B_\alpha)] + O_p(n_\alpha^{-3/2}) \\ &= \ln[\det(P_\alpha)] + \operatorname{tr}(B_\alpha) - \frac{1}{2}\operatorname{tr}(B_\alpha B_\alpha) + O_p(n_\alpha^{-3/2})\end{aligned}$$

Now, $B_\alpha B_\alpha = n_\alpha^{-1} P_\alpha^{-1} V_\alpha^{(1)} P_\alpha^{-1} V_\alpha^{(1)} + O_p(n_\alpha^{-3/2})$ so that

$$\begin{aligned}\ln[\det(R_\alpha)] = \ln[\det(P_\alpha)] &+ \frac{1}{\sqrt{n_\alpha}} \operatorname{tr}(P_\alpha^{-1} V_\alpha^{(1)}) + \frac{1}{n_\alpha} \operatorname{tr}(P_\alpha^{-1} V_\alpha^{(2)}) \\ &- \frac{1}{2n_\alpha} \operatorname{tr}\left(P_\alpha^{-1} V_\alpha^{(1)} P_\alpha^{-1} V_\alpha^{(1)}\right) + O_p(n_\alpha^{-3/2})\end{aligned}$$

A similar expansion for $\ln[\det(\overline{R})]$ may be obtained by defining $\overline{B}$ by

$$\overline{B} = \frac{1}{\sqrt{n}} \sum_{\alpha=1}^{k} \sqrt{\omega_\alpha} \overline{P}^{-1} V_\alpha^{(1)} + \frac{1}{n} \sum_{\alpha=1}^{k} \overline{P}^{-1} V_\alpha^{(2)}$$

Then

$$\begin{aligned}\ln[\det(\overline{R})] &= \ln[\det(\overline{P} + \overline{PB})] + O_p(n^{-3/2}) \\ &= \ln[\det(\overline{P})] + \operatorname{tr}(\overline{B}) - \frac{1}{2}\operatorname{tr}(\overline{BB}) + O_p(n^{-3/2})\end{aligned}$$

Since $\overline{BB} = n^{-1} \sum_{\alpha=1}^{k} \sum_{\beta=1}^{k} \sqrt{\omega_\alpha \omega_\beta} \overline{P}^{-1} V_\alpha^{(1)} \overline{P}^{-1} V_\beta^{(1)} + O_p(n^{-3/2})$,

$$\begin{aligned}\ln[\det(\overline{R})] = \ln[\det(\overline{P})] &+ \frac{1}{\sqrt{n}} \sum_{\alpha=1}^{k} \sqrt{\omega_\alpha} \operatorname{tr}(\overline{P}^{-1} V_\alpha^{(1)}) + \frac{1}{n} \sum_{\alpha=1}^{k} \operatorname{tr}(\overline{P}^{-1} V_\alpha^{(2)}) \\ &- \frac{1}{2n} \sum_{\alpha=1}^{k} \sum_{\beta=1}^{k} \sqrt{\omega_\alpha \omega_\beta} \operatorname{tr}(\overline{P}^{-1} V_\alpha^{(1)} \overline{P}^{-1} V_\beta^{(1)}) + O_p(n^{-3/2})\end{aligned}$$

Combining these expressions yields

$$\begin{aligned}g(R_1, \ldots, R_k) = \ln[\det(\overline{P})] &- \sum_{\alpha=1}^{k} \omega_\alpha \ln[\det(P_\alpha)] + \frac{1}{\sqrt{n}} \sum_{\alpha=1}^{k} \sqrt{\omega_\alpha} \operatorname{tr}(H_\alpha V_\alpha^{(1)}) \\ &+ \frac{1}{n} \sum_{\alpha=1}^{k} \operatorname{tr}(H_\alpha V_\alpha^{(2)}) + \frac{1}{2} \sum_{\alpha=1}^{k} \frac{\omega_\alpha}{n_\alpha} \operatorname{tr}(P_\alpha^{-1} V_\alpha^{(1)} P_\alpha^{-1} V_\alpha^{(1)})\end{aligned}$$

$$- \frac{1}{2n} \sum_{\alpha=1}^{k} \sum_{\beta=1}^{k} \sqrt{\omega_\alpha \omega_\beta}\, \mathrm{tr}(\overline{P}^{-1} V_\alpha^{(1)} \overline{P}^{-1} V_\beta^{(1)}) + O_p(n^{-3/2})$$

where $H_\alpha = (h_{\alpha ij}) = (\overline{P}^{-1} - P_\alpha^{-1}) = H'_\alpha$. Let $G(R_1, \ldots, R_k) = \sqrt{n}[g(R_1, \ldots, R_k) - g(P_1, \ldots, P_k)]$. Then, since $\sqrt{n}(\omega_\alpha / n_\alpha) = (\sqrt{n})^{-1}$, we obtain

$$\begin{aligned} G(R_1, \ldots, R_k) = & \sum_{\alpha=1}^{k} \sqrt{\omega_\alpha}\, \mathrm{tr}(H_\alpha V_\alpha^{(1)}) + \frac{1}{\sqrt{n}} \sum_{\alpha=1}^{k} \mathrm{tr}(H_\alpha V_\alpha^{(2)}) \\ & + \frac{1}{2\sqrt{n}} \sum_{\alpha=1}^{k} \mathrm{tr}(P_\alpha^{-1} V_\alpha^{(1)} P_\alpha^{-1} V_\alpha^{(1)}) \\ & - \frac{1}{2\sqrt{n}} \sum_{\alpha=1}^{k} \sum_{\beta=1}^{k} \sqrt{\omega_\alpha \omega_\beta}\, \mathrm{tr}(\overline{P}^{-1} V_\alpha^{(1)} \overline{P}^{-1} V_\beta^{(1)}) + O_p(n^{-1}) \end{aligned} \quad (13)$$

Theorem 1. *The expression $G(R_1, \ldots, R_k)$ may be written as*

$$\begin{aligned} G(R_1, \ldots, R_k) = & \sum_{\alpha=1}^{k} \sum_{i<j} \sqrt{\omega_\alpha}\, \overline{h}_{\alpha ij} v_{\alpha ij}^{(1)} + \frac{1}{\sqrt{n}} \sum_{\alpha=1}^{k} \sum_{i<j} \overline{h}_{\alpha ij} v_{\alpha ij}^{(2)} \\ & + \frac{1}{\sqrt{n}} \sum_{\alpha=1}^{k} \sum_{i<j} \sum_{k<\ell} q_\alpha(ij, k\ell) v_{\alpha ij}^{(1)} v_{\alpha k\ell}^{(1)} \\ & - \frac{1}{\sqrt{n}} \sum_{\alpha=1}^{k} \sum_{\beta=1}^{k} \sum_{i<j} \sum_{k<\ell} \sqrt{\omega_\alpha \omega_\beta}\, q(ij, k\ell) v_{\alpha ij}^{(1)} v_{\beta k\ell}^{(1)} + O_p(n^{-1}) \end{aligned}$$

where $P_\alpha^{-1} = (\rho_\alpha^{ij})$, $\overline{P}^{-1} = (\overline{\rho}^{ij})$, $\overline{h}_{\alpha ij} = 2(\overline{\rho}^{ij} - \rho_\alpha^{ij})$, $q_\alpha(ij, k\ell) = \rho_\alpha^{i\ell}\rho_\alpha^{jk} + \rho_\alpha^{ik}\rho_\alpha^{j\ell}$ *and* $q(ij, k\ell) = \overline{\rho}^{i\ell}\overline{\rho}^{jk} + \overline{\rho}^{ik}\overline{\rho}^{j\ell}$.

Proof. Using results on matrix algebra, we have

$$\sum_{\alpha=1}^{k} \sqrt{\omega_\alpha}\, \mathrm{tr}(H_\alpha V_\alpha^{(1)}) = \sum_{\alpha=1}^{k} \sqrt{\omega_\alpha} \sum_{i=1}^{p} \sum_{j=1}^{p} h_{\alpha ji} v_{\alpha ij}^{(1)}$$

and since H_α is symmetric, application of Lemma 3 yields

$$\sum_{\alpha=1}^{k} \sqrt{\omega_\alpha}\, \mathrm{tr}(H_\alpha V_\alpha^{(1)}) = \sum_{\alpha=1}^{k} \sqrt{\omega_\alpha} \sum_{i<j} (h_{\alpha ji} + h_{\alpha ij}) v_{\alpha ij}^{(1)} = \sum_{\alpha=1}^{k} \sqrt{\omega_\alpha} \sum_{i<j} \overline{h}_{\alpha ij} v_{\alpha ij}^{(1)}$$

In an entirely similar manner,

$$\sum_{\alpha=1}^{k} \mathrm{tr}(H_\alpha V_\alpha^{(2)}) = \sum_{\alpha=1}^{k} \sum_{i<j} \overline{h}_{\alpha ij} v_{\alpha ij}^{(2)}$$

Using Lemma 4, results on matrix algebra and the symmetry of $V_\alpha^{(1)}$, we have

$$\begin{aligned}
&\frac{1}{2}\sum_{\alpha=1}^{k}\operatorname{tr}(P_\alpha^{-1}V_\alpha^{(1)}P_\alpha^{-1}V_\alpha^{(1)}) \\
&\quad=\frac{1}{2}\sum_{\alpha=1}^{k}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{\ell=1}^{p}\rho_\alpha^{i\ell}\rho_\alpha^{jk}v_{\alpha ij}^{(1)}v_{\alpha k\ell}^{(1)} \\
&\quad=\frac{1}{2}\sum_{\alpha=1}^{k}\sum_{i<j}\sum_{k<\ell}(\rho_\alpha^{i\ell}\rho_\alpha^{jk}+\rho_\alpha^{j\ell}\rho_\alpha^{ik}+\rho_\alpha^{ik}\rho_\alpha^{j\ell}+\rho_\alpha^{i\ell}\rho_\alpha^{jk})v_{\alpha ij}^{(1)}v_{\alpha k\ell}^{(1)} \\
&\quad=\sum_{\alpha=1}^{k}\sum_{i<j}\sum_{k<\ell}q_\alpha(ij,k\ell)v_{\alpha ij}^{(1)}v_{\alpha k\ell}^{(1)}
\end{aligned}$$

In a similar manner,

$$\begin{aligned}
&\frac{1}{2}\sum_{\alpha=1}^{k}\sum_{\beta=1}^{k}\sqrt{\omega_\alpha\omega_\beta}\operatorname{tr}(\overline{P}^{-1}V_\alpha^{(1)}\overline{P}^{-1}V_\beta^{(1)}) \\
&\qquad=\sum_{\alpha=1}^{k}\sum_{\beta=1}^{k}\sum_{i<j}\sum_{k<\ell}\sqrt{\omega_\alpha\omega_\beta}\,q(ij,k\ell)v_{\alpha ij}^{(1)}v_{\beta k\ell}^{(1)}
\end{aligned}$$

Combining these expansions in (13) completes the proof. □

Corollary 1. *In the special case $p=2$, $G(R_1,\ldots,R_k)$ may be written as*

$$\begin{aligned}
G(R_1,\ldots,R_k)&=2\sum_{\alpha=1}^{k}\sqrt{\omega_\alpha}\left(\frac{\rho_\alpha}{1-\rho_\alpha^2}-\frac{\overline{\rho}}{1-\overline{\rho}^2}\right)v_{\alpha 12}^{(1)} \\
&\quad+\frac{2}{\sqrt{n}}\sum_{\alpha=1}^{k}\left(\frac{\rho_\alpha}{1-\rho_\alpha^2}-\frac{\overline{\rho}}{1-\overline{\rho}^2}\right)v_{\alpha 12}^{(2)}+\frac{1}{\sqrt{n}}\sum_{\alpha=1}^{k}\frac{1+\rho_\alpha^2}{(1-\rho_\alpha^2)^2}(v_{\alpha 12}^{(1)})^2 \\
&\quad-\frac{1}{\sqrt{n}}\sum_{\alpha=1}^{k}\sum_{\beta=1}^{k}\frac{1+\overline{\rho}^2}{(1-\overline{\rho}^2)^2}v_{\alpha 12}^{(1)}v_{\beta 12}^{(1)}+O_p(n^{-1}).
\end{aligned}$$

Proof. For $p=2$, $\sum_{i<j}$ consists of single term corresponding to $i=1, j=2$. Also, $P_\alpha=\left(\begin{smallmatrix}1 & \rho_\alpha\\ \rho_\alpha & 1\end{smallmatrix}\right)$ so that $P_\alpha^{-1}=(1-\rho_\alpha^2)^{-1}\left(\begin{smallmatrix}1 & -\rho_\alpha\\ -\rho_\alpha & 1\end{smallmatrix}\right)$. Similarly, $\overline{P}^{-1}=(1-\overline{\rho}^2)^{-1}\left(\begin{smallmatrix}1 & -\overline{\rho}\\ -\overline{\rho} & 1\end{smallmatrix}\right)$. Thus, the off diagonal element of H_α is given by $\rho_\alpha(1-\rho_\alpha^2)^{-1}-\overline{\rho}(1-\overline{\rho}^2)^{-1}$. Further, $q_\alpha(12,12)=\rho_\alpha^{12}\rho_\alpha^{21}+\rho_\alpha^{11}\rho_\alpha^{22}=(1+\rho_\alpha^2)/(1-\rho_\alpha^2)^2$ and $q(12,12)=(1+\overline{\rho}^2)/(1-\overline{\rho}^2)^2$. The result follows by using these values in the theorem. □

5. Asymptotic Null Distribution of L^*

In this section we derive asymptotic distribution of the statistic L^* when the null hypothesis is true.

Define the $k \times k$ matrix W as $W = (w_{ij})$ where $w_{ii} = 1 - \omega_i$ and for $i \neq j$, $w_{ij} = -\sqrt{\omega_i \omega_j} = w_{ji}$, $1 \leq i, j, \leq k$. The matrix W has rank $k-1$ and each of its non-zero eigenvalues is equal to 1.

Theorem 2. *Let the k correlation matrices $R_1, \ldots, R_k$ be based on independent samples of sizes $N_1, \ldots, N_k$, respectively, with finite fourth order cumulants. Define the $kp(p-1)/2 \times 1$ vector $\mathbf{v}^{(1)}$ by*

$$\boldsymbol{v}^{(1)} = (v^{(1)}_{1,1,2}, v^{(1)}_{1,1,3}, \ldots, v^{(1)}_{1,p-1,p}, v^{(1)}_{2,1,2}, v^{(1)}_{2,1,3}, \ldots, v^{(1)}_{2,p-1,p}, \ldots, v^{(1)}_{k,1,2}, v^{(1)}_{k,1,3}, \ldots, v^{(1)}_{k,p-1,p})'$$

where $V^{(1)}_\alpha$ is as defined in (9). Let $Q = (q(ij, k\ell))$ be the $p(p-1)/2 \times p(p-1)/2$ matrix of coefficients defined in Theorem 1.

Let T_α be the asymptotic dispersion matrix of $V^{(1)}_\alpha$ with entry $(ij, k\ell)$ equal to $E(v^{(1)}_{\alpha ij} v^{(1)}_{\alpha k\ell})$ given in (12). Then, the asymptotic dispersion matrix of $\mathbf{v}^{(1)}$ is

$$T^* = \begin{pmatrix} T_1 & 0 & \cdots & 0 \\ 0 & T_2 & \cdots & 0 \\ \vdots & \vdots & & \\ 0 & 0 & \cdots & T_k \end{pmatrix}$$

Under the null hypothesis

$$L^* = \sum_{i=1}^{p(p-1)(k-1)/2} \lambda_i y_i + O_p(n^{-1/2})$$

where $y_1, \ldots, y_{p(p-1)(k-1)/2}$ are independent, $y_i \sim \chi^2_1$, $1 \leq i \leq p(p-1)(k-1)/2$ and $\lambda_1, \ldots, \lambda_{p(p-1)(k-1)/2}$ are the eigenvalues of $T^*(Q \otimes W)$. If the standardized fourth order cumulants of the populations are all equal, then $T_\alpha = T$ for $\alpha = 1, \ldots, k$ and

$$L^* = \sum_{i=1}^{p(p-1)/2} \theta_i u_i + O_p(n^{-1/2}),$$

where $u_1, \ldots, u_{p(p-1)/2}$ are independent, $u_i \sim \chi^2_{k-1}$ and $\theta_1, \ldots, \theta_{p(p-1)/2}$ are the eigenvalues of TQ.

Proof. Under the null hypothesis we have $P_\alpha = \overline{P}$ for $\alpha = 1, \ldots, k$ so that $g(P_1, \ldots, P_k) = 0$, $h_{\alpha ij} = 0$ and $q_\alpha(ij, k\ell) = q(ij, k\ell) = \rho^{i\ell}\rho^{jk} + \rho^{ik}\rho^{j\ell}$ for all α. Since $g(R_1, \ldots, R_k) = \ln[\det(\overline{R})] - \sum_{\alpha=1}^k \omega_\alpha \ln[\det(\overline{R}_\alpha)] = n^{-1}L^*$, using Theorem 1, one obtains

$$\begin{aligned} L^* &= n g(R_1, \ldots, R_k) = n[g(R_1, \ldots, R_k) - g(P_1, \ldots, P_k)] \\ &= \sum_{\alpha=1}^k \sum_{i<j} \sum_{k<\ell} q(ij, k\ell) v^{(1)}_{\alpha ij} v^{(1)}_{\alpha k\ell} \\ &\quad - \sum_{\alpha=1}^k \sum_{\beta=1}^k \sum_{i<j} \sum_{k<\ell} \sqrt{\omega_\alpha \omega_\beta}\, q(ij, k\ell) v^{(1)}_{\alpha ij} v^{(1)}_{\beta k\ell} + O_p(n^{-1/2}) \end{aligned}$$

$$= \sum_{\alpha=1}^{k}\sum_{\beta=1}^{k} w_{\alpha\beta} \sum_{i<j}\sum_{k<\ell} q(ij,k\ell) v_{\alpha ij}^{(1)} v_{\beta k\ell}^{(1)} + O_p(n^{-1/2})$$
$$= (\boldsymbol{v}^{(1)})'(Q \otimes W)\boldsymbol{v}^{(1)} + O_p(n^{-1/2})$$

Since Q is of rank $p(p-1)/2$ and W is of rank $k-1$, the matrix $Q \otimes W$ is of rank $p(p-1)(k-1)/2$. From (8) and (9) it is clear that elements of $V_\alpha^{(1)}$ are linear functions of elements of V_α and the limiting distribution of V_α is normal with means 0 and covariances that depend on the fourth order cumulants of the parent population. Therefore, $\boldsymbol{v}^{(1)}$ is asymptotically normal with means zero and dispersion matrix T^*. Thus, $L^* = \sum_{i=1}^{p(p-1)(k-1)/2} \lambda_i y_i + O_p(n^{-1/2})$.

If the standardized fourth order cumulants are the same for each underlying population, then $T^* = T \otimes I$. Further, $(T \otimes I)(Q \otimes W) = TQ \otimes W$ has as its eigenvalues $\theta_i \epsilon_j$, $i = 1, \ldots, p(p-1)/2$, $j = 1, \ldots, k$ where θ_i are the eigenvalues of TQ and ϵ_j are the eigenvalues of W. Since there are $p(p-1)/2$ non-zero eigenvalues of $(T \otimes I)(Q \otimes W)$ each occurring with multiplicity $k-1$, we have $L^* = \sum_{i=1}^{p(p-1)/2} \theta_i u_i + O_p(n^{-1/2})$. □

Corollary 2. *Let the k sample correlation coefficients $r_1, r_2, \ldots, r_k$ be based on independent samples of sizes $N_1, N_2, \ldots, N_k$ from bivariate populations with finite fourth order cumulants. Let ρ be the hypothesized common correlation coefficient. Define the $k \times 1$ vector $\boldsymbol{v}^{(1)}$ by*

$$\boldsymbol{v}^{(1)} = (v_1^{(1)}, \ldots, v_k^{(1)})'$$

where $v_\alpha^{(1)} = v_{\alpha 12} - \rho(v_{\alpha 11} + v_{\alpha 22})$ as defined in (9). Let

$$t_\alpha = (1-\rho^2)^2 + \frac{1}{4}\rho^2(\kappa_{\alpha 1111} + \kappa_{\alpha 2222}) + \left(1 + \frac{1}{2}\rho^2\right)\kappa_{\alpha 1122} - \rho(\kappa_{\alpha 1113} + \kappa_{\alpha 1222})$$

and define $T^* = \operatorname{diag}(t_1, \ldots, t_k)$. Under the null hypothesis the statistic L^* is asymptotically expanded as

$$L^* = \frac{1+\rho^2}{(1-\rho^2)^2} \sum_{i=1}^{k-1} \lambda_i y_i + O_p(n^{-1/2})$$

where $y_1, \ldots, y_{k-1}$ are independent, $y_i \sim \chi_1^2$ and $\lambda_1, \ldots, \lambda_{,k-1}$ are the eigenvalues of T^*W. If the standardized fourth order cumulants are equal, then

$$t_\alpha = (1-\rho^2)^2 + \frac{1}{4}\rho^2(\kappa_{1111} + \kappa_{2222}) + \left(1 + \frac{1}{2}\rho^2\right)\kappa_{1122} - \rho(\kappa_{1113} + \kappa_{1222})$$

for $\alpha = 1, 2, \ldots, k$ and

$$L^* = \left[(1-\rho^2)^2 + \frac{1}{4}\rho^2(\kappa_{1111} + \kappa_{2222}) + \left(1 + \frac{1}{2}\rho^2\right)\kappa_{1122} - \rho(\kappa_{1113} + \kappa_{1222})\right] \frac{1+\rho^2}{(1-\rho^2)^2}\chi_{k-1}^2 + O_p(n^{-1/2})$$

Proof. As shown in Corollary 1, when $p = 2$, Q is a scalar. If ρ is the common correlation coefficient, then $Q = (1+\rho^2)/(1-\rho^2)^2$. The asymptotic variance of $v_{\alpha 12}^{(1)}$ is given in (12). Upon simplification,

$$E(v_{\alpha 12}^{(1)} v_{\alpha 12}^{(1)}) = t_\alpha = (1-\rho^2)^2 + \frac{1}{4}\rho^2(\kappa_{\alpha 1111} + \kappa_{\alpha 2222}) + \left(1 + \frac{1}{2}\rho^2\right)\kappa_{\alpha 1122} - \rho(\kappa_{\alpha 1113} + \kappa_{\alpha 1222}) + O_p(n^{-1/2}) \tag{14}$$

so that T^* is the asymptotic covariance matrix of $\mathbf{v}^{(1)}$. Further, $T^*(Q \otimes W) = [(1+\rho^2)/(1-\rho^2)^2]T^*W$. Thus $L^* = [(1+\rho^2)/(1-\rho^2)^2] \sum_{i=1}^{k-1} \lambda_i y_i + O_p(n^{-1/2})$, where λ_i are the eigenvalues of T^*W. If the standardized fourth order cumulants are identical, $T = tI$, so that there is one eigenvalue of TQ with multiplicity k. This eigenvalue is merely $t(1+\rho^2)/(1-\rho^2)^2$ and the result follows immediately from Theorem 2. □

Corollary 3. *Let the k sample correlation coefficients $r_1, r_2, \ldots, r_k$ be based on independent samples of sizes $N_1, N_2, \ldots, N_k$ from bivariate populations which are elliptically contoured with a common curtosis of 3κ and common correlation coefficient ρ. Then*

$$L^* = \left[(1-\rho^2)^2 + (1+2\rho^2)\kappa\right] \frac{1+\rho^2}{(1-\rho^2)^2} \chi_{k-1}^2 + O_p(n^{-1/2})$$

Proof. For elliptically contoured distributions (Muirhead 1982, Anderson 2003, Gupta and Varga 1993) the fourth order cumulants are such that $\kappa_{iiii} = 3\kappa_{iijj} = 3\kappa$ for $i \neq j$ and all other cumulants are zero (Waternaux 1984). Substituting this into the expression for t in Corollary 2 yields $t = (1-\rho^2)^2 + (1+2\rho^2)\kappa$. The result then follows from Corollary 2. □

Corollary 4. *Let the k sample correlation coefficients $r_1, \ldots, r_k$ be based on independent samples of sizes $N_1, \ldots, N_k$ from bivariate normal populations with a common correlation coefficient ρ. Then*

$$L^* = (1+\rho^2)\chi_{k-1}^2 + O_p(n^{-1/2})$$

Proof. Normal distributions are special case of elliptically contoured distributions. The fourth order cumulants are all zero (Anderson 2003). The result follows by setting $\kappa = 0$ in Corollary 3. □

6. An Example

This example is included to demonstrate the procedure to be used when testing the equality of correlation matrices by using the statistic L^*. The data represent random samples from three trivariate populations each with identical correlation matrix P given by

$$P = \begin{pmatrix} 1.0 & 0.3 & 0.2 \\ 0.3 & 1.0 & -0.3 \\ 0.2 & -0.3 & 1.0 \end{pmatrix}$$

Since the statistic L^* is an approximation of the modified likelihood ratio statistic for samples from multivariate normal populations, it is particularly suited to populations that are *near* normal. The contaminated normal model has been chosen to represent such a distribution.

Samples of size 25 from contaminated normal populations with mixing parameter $\epsilon = 0.1$ and $\sigma = 2$ were generated using the SAS system. These data are tabulated in Gupta, Johnson and Nagar (2012). The density of a contaminated normal model is given by

$$\phi_\epsilon(\mathbf{x}, \sigma, \Sigma) = (1-\epsilon)\phi(\mathbf{x}, \Sigma) + \epsilon\phi(\mathbf{x}, \sigma\Sigma), \quad \sigma > 0, \quad 0 < \epsilon < 1$$

where $\phi(\mathbf{x}, \Sigma)$ is the density of a multivariate normal distribution with zero mean vector and covariance matrix Σ.

If the data were known to be from three normal populations all that would be required at this point would be the sample sizes and the matrix of corrected sums of squares and cross products. A key element, however, of the modified likelihood ratio procedure is that this assumption need not be made, but the fourth order cumulant must be estimated. To do this the k-statistics are calculated using Kaplan's formulae summarized in Section 3. The computations are made considerably easier by standardizing the data so that all of the first order sums are zero.

The computation using original (or standardized) data yields the following estimates of the individual correlation matrices:

$$R_1 = \begin{pmatrix} 1.0000 & 0.5105 & 0.3193 \\ 0.5105 & 1.0000 & -0.3485 \\ 0.3193 & -0.3485 & 1.0000 \end{pmatrix}, \quad \det(R_1) = 0.4024$$

$$R_2 = \begin{pmatrix} 1.0000 & 0.1758 & 0.2714 \\ 0.1758 & 1.0000 & -0.2688 \\ 0.2714 & -0.2688 & 1.0000 \end{pmatrix}, \quad \det(R_2) = 0.7975$$

$$R_3 = \begin{pmatrix} 1.0000 & 0.2457 & 0.3176 \\ 0.2457 & 1.0000 & -0.0331 \\ 0.3176 & -0.0331 & 1.0000 \end{pmatrix}, \quad \det(R_3) = 0.8325$$

Since each sample is of size 25, $\omega_i = 1/3$ for $i = 1, 2, 3$ and the pooled correlation matrix is merely the average of these three matrices:

$$\overline{R} = \begin{pmatrix} 1.0000 & 0.3107 & 0.3028 \\ 0.3107 & 1.0000 & -0.2168 \\ 0.3028 & -0.2168 & 1.0000 \end{pmatrix}, \quad \det(\overline{R}) = 0.7240$$

The value of the test statistic is now easily calculated as

$$\begin{aligned} L^* &= 72\ln(0.7240) - 24[\ln(0.4024) + \ln(00.7975) + \ln(0.8325)] \\ &= 8.7473 \end{aligned}$$

The null hypothesis is to be rejected if the value of the test statistic is too large. The next step of the procedure is to estimate the coefficients in the linear combination of chi-square variables that make up the actual distribution under the null hypothesis. The most arduous part is the computation of the estimates of fourth order cumulants.

Since the data are standardized, the formula for the k-statistic for the four way product $x_i \times x_j \times x_k \times x_\ell$ simplifies to

$$k_{ijk\ell} = \frac{1}{N^{(4)}}[N^2(N+1)s_{ijk\ell} - N(N-1)(s_{ij}s_{k\ell} + s_{ik}s_{j\ell} + s_{i\ell}s_{jk})]$$

where $N^{(4)} = N(N-1)(N-2)(N-3)$. Using this to estimate the cumulant corresponding to $x_1^2 x_2^2$ yields $k_{1122} = 0.6670$. The computation for other fourth order cumulant are performed similarly. The resulting estimates are then pooled across population to yield an estimate of the common fourth order cumulants used in building the tau matrix (it is possible, of course, to drop the assumption of common fourth order cumulants and use the nine by nine matrix that would result if each separate tau matrix were joined in a block diagonal matrix). The estimates of the fourth order cumulants are summarized in the Table 1.

The pooled correlation matrix and these estimates are now used to build the estimated covariance matrix $V^{(1)}$. The entry corresponding to $v_{ij}^{(1)} v_{k\ell}^{(1)}$ is given by

$$\begin{aligned} k_{ijk\ell} &- \frac{1}{2}(r_{ij}k_{iik\ell} + r_{ij}k_{jjk\ell} + r_{k\ell}k_{ijkk} + r_{k\ell}k_{ij\ell\ell}) \\ &+ \frac{1}{4}r_{ij}r_{k\ell}(k_{iikk} + k_{ii\ell\ell} + k_{jjkk} + k_{jj\ell\ell}) \\ &- (r_{k\ell}r_{ik}r_{jk} + r_{k\ell}r_{i\ell}r_{j\ell} + r_{ij}r_{ik}r_{i\ell} + r_{ij}r_{jk}r_{j\ell}) \\ &+ \frac{1}{2}r_{ij}r_{k\ell}(r_{ik}^2 + r_{i\ell}^2 + r_{jk}^2 + r_{j\ell}^2) + r_{ik}r_{j\ell} + r_{i\ell}r_{jk} \end{aligned}$$

where r_{ij} is the pooled estimate of the correlation value and $k_{ijk\ell}$ is the corresponding pooled fourth order cumulant. The entry corresponding to $12, 13$ is given by $t_{12,13} = -0.3065$. Similar calculations yield the following covariance matrix corresponding to $(v_{\alpha 12}^{(1)}, v_{\alpha 13}^{(1)}, v_{\alpha 23}^{(1)})'$,

$$T = \begin{pmatrix} 1.0150 & -0.3065 & 0.1800 \\ -0.3065 & 0.7242 & 0.3974 \\ 0.1800 & 0.3974 & 0.8179 \end{pmatrix}$$

To complete the example, the inverse of the pooled correlation matrix is used to estimate the matrix Q. The entry corresponding to the element $ij, k\ell$ is given by $r^{ik}r^{j\ell} + r^{i\ell}r^{jk}$ where $R^{-1} = (r^{ij})$. These matrices are as follows:

$$\overline{R}^{-1} = \begin{pmatrix} 1.3163 & -0.5198 & -0.5113 \\ -0.5198 & 1.2546 & 0.4294 \\ -0.5113 & 0.4294 & 1.2479 \end{pmatrix}$$

Table 1: Estimated fourth order cumulants

Variables	Population 1	Population 2	Population 3	Pooled
1111	0.9077	0.1181	0.9355	0.6538
1112	0.7765	-0.0387	-0.0565	0.2271
1113	-0.3015	0.7008	0.0677	0.1105
1122	0.6670	0.3595	-0.3663	0.2201
1123	-0.3917	0.3519	-0.1333	-0.0574
1133	-0.1848	0.6608	-0.7475	-0.0905
1222	0.4896	-0.7128	-0.0178	-0.0803
1223	-0.3005	0.1637	-0.2243	-0.1204
1233	-0.0980	0.6343	-0.1394	0.1323
1333	-0.3430	0.3973	-0.0773	-0.0077
2222	-0.0787	-0.9989	0.8134	-0.0881
2223	-0.2543	0.0750	0.1887	0.0032
2233	0.3800	-0.1764	-0.5454	-0.1139
2333	-0.8386	0.8496	0.2869	0.0993
3333	0.9130	-0.9196	1.3068	0.4334

$$Q = \begin{pmatrix} 1.9217 & 0.8310 & -0.8647 \\ 0.8310 & 1.9041 & -0.8682 \\ -0.8647 & 0.8682 & 1.7500 \end{pmatrix}$$

Most eigenvalues extraction routines require that the matrix being analyzed be symmetric. Let A be the Cholesky decomposition of Q, that is $Q = A'A$ where A is an upper triangular matrix. Then the eigenvalues of TQ are the same as the eigenvalues of ATA' which is clearly symmetric. In this case

$$A = \begin{pmatrix} 1.3863 & 0.5995 & -0.6237 \\ 0 & 1.2429 & -0.3977 \\ 0 & 0 & 1.0967 \end{pmatrix}$$

$$ATA' = \begin{pmatrix} 1.4111 & -0.2877 & -0.0246 \\ -0.2877 & 0.8552 & 0.1849 \\ -0.0246 & 0.1849 & 0.9837 \end{pmatrix}$$

and the eigenvalue of this matrix are $1.55, 1.0473$ and 0.6527. Using Theorem 2, the distribution of the statistic is estimated to be that of $Y = (1.55)\chi^2_2 + (1.0473)\chi^2_2 + (0.6527)\chi^2_2$ where each of the chi-square variate is independent. Using Lemma 7 the cumulative probability value associated with 8.7473 is obtained as 0.7665 so that the observed significance level is 0.2335. Thus, if the test is performed at the $\alpha = 0.1$ level of significance the conclusion reached is that there is insufficient evidence to reject the null hypothesis that the samples are from populations with identical correlation matrices.

Acknowledgements

The authors wish to thank three anonymous referees for their careful reading of the manuscript and their fruitful comments and suggestions.

[Recibido: junio de 2012 — Aceptado: julio de 2013]

References

Aitkin, M. (1969), 'Some tests for correlation matrices', *Biometrika* **56**, 443–446.

Aitkin, M. A., Nelson, W. C. & Reinfurt, K. H. (1968), 'Tests for correlation matrices', *Biometrika* **55**, 327–334.

Ali, M. M., Fraser, D. A. S. & Lee, Y. S. (1970), 'Distribution of the correlation matrix', *Journal of Statistical Research* **4**, 1–15.

Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis*, Wiley Series in Probability and Statistics, third edn, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.

Browne, M. W. (1978), 'The likelihood ratio test for the equality of correlation matrices', *The British Journal of Mathematical and Statistical Psychology* **31**(2), 209–217.
*http://dx.doi.org/10.1111/j.2044-8317.1978.tb00585.x

Cole, N. (1968*a*), The likelihood ratio test of the equality of correlation matrices, Technical Report 1968-65, The L. L. Thurstone Psychometric Laboratory, University of North Carolina, Chapel Hill, North Carolina.

Cole, N. (1968*b*), On testing the equality of correlation matrices, Technical Report 1968-66, The L. L. Thurstone Psychometric Laboratory, University of North Carolina, Chapel Hill, North Carolina.

Gleser, L. J. (1968), 'On testing a set of correlation coefficients for equality: Some asymptotic results', *Biometrika* **55**, 513–517.

Gupta, A. K., Johnson, B. E. & Nagar, D. K. (2012), Testing equality of several correlation matrices, Technical Report 12-08, Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, Ohio.

Gupta, A. K. & Nagar, D. K. (2000), *Matrix Variate Distributions*, Vol. 104 of *Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics*, Chapman & Hall/CRC, Boca Raton, FL.

Gupta, A. K. & Varga, T. (1993), *Elliptically Contoured Models in Statistics*, Vol. 240 of *Mathematics and its Applications*, Kluwer Academic Publishers Group, Dordrecht.
*http://dx.doi.org/10.1007/978-94-011-1646-6

Jennrich, R. I. (1970), 'An asymptotic χ^2 test for the equality of two correlation matrices', *Journal of the American Statistical Association* **65**, 904–912.

Kaplan, E. L. (1952), 'Tensor notation and the sampling cumulants of k-statistics', *Biometrika* **39**, 319–323.

Revista Colombiana de Estadística
Diciembre 2013, volumen 36, no. 2, pp. 261 a 271

Estimation of Variance Components in Linear Mixed Models with Commutative Orthogonal Block Structure

Estimación de las componentes de varianza en modelos lineales mixtos con estructura de bloques ortogonal conmutativa

Sandra S. Ferreira[1,a], Dário Ferreira[2,b], Célia Nunes[3,c], João T. Mexia[4,d]

[1]Department of Mathematics and Center of Mathematics, Faculty of Sciences, University of Beira Interior, Covilhã, Portugal

[2]Department of Mathematics and Center of Mathematics, Faculty of Sciences, University of Beira Interior, Covilhã, Portugal

[3]Department of Mathematics and Center of Mathematics, Faculty of Sciences, University of Beira Interior, Covilhã, Portugal

[4]Department of Mathematics and Center of Mathematics and Its Applications, Faculty of Science and Technology, New University of Lisbon, Covilhã, Portugal

Abstract

Segregation and matching are techniques to estimate variance components in mixed models. A question arising is whether segregation can be applied in situations where matching does not apply. Our motivation for this research relies on the fact that we want an answer to that question and to explore this important class of models that can contribute to the development of mixed models. That is possible using the algebraic structure of mixed models. We present two examples showing that segregation can be applied in situations where matching does not apply.

***Key words*:** Commutative Jordan algebra, Mixed model, Variance components.

[a]Professor. E-mail: sandraf@ubi.pt
[b]Professor. E-mail: dario@ubi.pt
[c]Professor. E-mail: celian@ubi.pt
[d]Professor. E-mail: jtm@fct.unl.pt

Resumen

La segregación y el emparejamiento son técnicas para estimar las componentes de varianza en modelos mixtos. Una pregunta que ha surgido es si la segregación puede ser aplicada en situaciones en las que el emparejamiento no es aplicable. Nuestra motivación para esta investigación se basa en el hecho de que se quiere una respuesta a esta pregunta y se quiere explorar esta importante clase de modelos con el fin de contribuir al desarrollo de los modelos mixtos. Esto es posible utilizando la estructura algebraica de los modelos mixtos con estructura de bloques ortogonal conmutativa. Se presentan dos ejemplos que muestran que la segregación puede ser aplicada en situaciones donde el emparejamiento no es aplicable.

***Palabras clave*:** álgebra conmutativa Jordan, componentes de varianza, modelo mixto.

1. Introduction

Mixed models have orthogonal block structure, OBS, when their variance covariance matrices are orthogonal all the linear combinations of known pairwise projection matrices, POOPM, add up to $\mathbf{I}_n$ with non negative coefficients. These models play an important role in design of experiments (Houtman & Speed 1983, Mejza 1992) and were introduced by Nelder (1965*a*, 1965*b*), continuing to play an important part in the theory of randomized block designs (see Caliński & Kageyama 2000, Caliński & Kageyama 2003).

A direct generalization of this class of models is that of models whose variance covariance matrices are linear combinations of known POOPM, we say these models to have generalized orthogonal block structure, GOBS. Moreover if the orthogonal projection matrix $\mathbf{T}$ on the space spanned by the mean vectors commutes with these POOPM the model, (see Fonseca, Mexia & Zmyślony 2008) will have commutative orthogonal block structure, COBS. Then, (see Zmyślony 1978), its least square estimators, LSE, for estimable vectors will be best linear unbiased estimators, BLUE, whatever the variance components.

In what follows, we will present techniques for the estimation of variance components in COBS. These techniques will be based on the algebraic structure of the models then being quite distinct from other techniques that require normality. Moreover it has interesting developments, namely these related to model segregation.

The next Section presents the algebraic structure of the models considering commutative Jordan algebras. Then we discuss, in section 3, the techniques for the estimation of variance components: Matching and segregation. Segregation displays the possibility of using the algebraic structure in estimation. Thus, in subsections 3.1 and 3.2, we present two models in which this technique has to be used to complete the structure based on estimation of variance components. Lastly, we present some final remarks.

2. Algebras and Models

Commutative Jordan Algebras, CJA, (of symmetric matrices) are linear spaces constituted by symmetric matrices that commute and containing the square of this matrices. Each CJA $\mathcal{A}$ has a principal unique basis (see, Seely 1971), $pb(\mathcal{A})$, constituted by pairwise orthogonal projection matrices. Any orthogonal projection matrix belonging to $\mathcal{A}$ will be the sum of matrices in $pb(\mathcal{A})$.

Moreover, given a family $\boldsymbol{W}$ of symmetric matrices that commute, there is a minimal CJA $\mathcal{A}(\boldsymbol{W})$ containing $\boldsymbol{W}$ (see, Fonseca et al. 2008).

Consider the model

$$\mathbf{Y} = \sum_{i=0}^{w} \mathbf{X}_i \boldsymbol{\beta_i} \tag{1}$$

where $\boldsymbol{\beta}_0$ is fixed and the $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_w$ are independent, with null mean vectors and variance covariance matrices

$$\begin{cases} \boldsymbol{\mu} = \mathbf{X}_0 \boldsymbol{\beta_0} \\ \mathbf{V}(\boldsymbol{\theta}) = \sum_{i=1}^{w} \theta_i \mathbf{M}_i \end{cases} \tag{2}$$

with $\mathbf{M}_i = \mathbf{X}_i \mathbf{X}_i^{'}, i = 1, \ldots, w$. When the matrices in $\{\boldsymbol{T}, \boldsymbol{M}, \ldots, \boldsymbol{M}_w\}$ commute we have the CJA $\mathcal{A}(\boldsymbol{W})$ with principal basis

$$\boldsymbol{Q} = \{\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_m\}.$$

We can order the matrices in $\boldsymbol{Q}$ to have $\boldsymbol{T} = \sum_{j=1}^{z} \mathbf{Q}_j$. Moreover

$$\mathbf{M}_i = \sum_{j=1}^{m} b_{i,j} \mathbf{Q}_j, i = 1, \ldots, w,$$

so that

$$\mathbf{V}(\boldsymbol{\theta}) = \sum_{i=1}^{w} \boldsymbol{\theta}_i \mathbf{M}_i = \sum_{j=1}^{m} \boldsymbol{\gamma}_j \mathbf{Q}_j = \boldsymbol{V}(\boldsymbol{\gamma})$$

with $\boldsymbol{\gamma}_j = \sum_{i=1}^{w} b_{i,j} \boldsymbol{\theta}_i, j = 1, \ldots, m$, thus the model will have COBS since its variance covariance matrices are linear combinations of known POOPM that commute with the $\mathbf{Q}_1, \ldots, \mathbf{Q}_m$, belonging jointly to $\mathcal{A}(\boldsymbol{W})$.

3. Segregation and Matching

Since $R(\boldsymbol{Q}_j) \subseteq R(\boldsymbol{T})$, $j = 1, \ldots, z$ we can estimate directly the $\gamma_{z+1}, \ldots, \gamma_m$, for which we have the unbiased estimators

$$\widetilde{\boldsymbol{\gamma}}_j = \frac{\|\mathbf{Q}_j \mathbf{Y}\|^2}{r(\mathbf{Q}_j)}, j = z+1, \ldots, m. \tag{3}$$

Partitioning matrix $\boldsymbol{B} = [b_{i,j}]$ as $[\mathbf{B}_1 \quad \mathbf{B}_2]$, where $\mathbf{B}_1$ has z columns, and taking $\boldsymbol{\gamma_1}= (\gamma_1, \ldots, \gamma_z)'$, $\boldsymbol{\gamma_2}= (\gamma_{z+1}, \ldots, \gamma_m)'$, and $\boldsymbol{\sigma^2} = (\sigma_1^2, \ldots, \sigma_w^2)'$, with $w \leq m - z$, we have

$$\boldsymbol{\gamma_l} = \mathbf{B}_l'\boldsymbol{\sigma}^2, l = 1, 2. \tag{4}$$

When the column vectors of $\mathbf{B}_2^{'}$ are linearly independent we have

$$\boldsymbol{\sigma^2} = (\mathbf{B}_2')^+\boldsymbol{\gamma_2}, \tag{5}$$

as well as

$$\boldsymbol{\gamma_1} = \mathbf{B}_1'(\mathbf{B}_2')^+\boldsymbol{\gamma_2}, \tag{6}$$

allowing the estimation of $\boldsymbol{\sigma^2}$ and $\boldsymbol{\gamma_1}$, through $\boldsymbol{\gamma_2}$. It may be noted that if the matrices $\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_m$ can be ordered in such a way that the transition matrix is

$$\boldsymbol{B} = \begin{bmatrix} B_{1,1} & 0 \\ B_{2,1} & B_{2,2} \end{bmatrix},$$

with $B_{1,1}$ a $z \times z$ matrix, the model is said to be segregated, see Ferreira, Ferreira & Mexia (2007) and Ferreira, Ferreira, Nunes & Mexia (2010). It can be pointed out that, in that case, sub-matrices $B_{1,1}$ and $B_{2,2}$ are regular.

When $\mathbf{B}_1$ is a sub-matrix of $\mathbf{B}_2$, $\mathbf{B}_1'$ will be a sub-matrix of $\mathbf{B}_2'$ and so $\boldsymbol{\gamma_1}$ will be a sub-vector of $\boldsymbol{\gamma_2}$, see Mexia, Vaquinhas, Fonseca & Zmyslony (2010). In this case the match have between the components of $\boldsymbol{\gamma_1}$ and some components of $\boldsymbol{\gamma_2}$. When this happens we say that the model has matching. Thus $\boldsymbol{\gamma_1}$ and

$$\boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\gamma_1'} & \boldsymbol{\gamma_2'} \end{bmatrix}',$$

can be directly estimated from $\boldsymbol{\gamma_2}$. If the row vectors of $\mathbf{B}$ are linearly independent, we have

$$\boldsymbol{\sigma}^2 = (\mathbf{B}')^+\boldsymbol{\gamma}, \tag{7}$$

and we can also estimate $\boldsymbol{\sigma}^2$. Requiring the row vectors of $\mathbf{B}$ to be linearly independent is less restrictive than requiring the row vectors of $\mathbf{B}_2$ to be linearly independent.

Below we introduce two examples which show that segregation can be applied in situations where matching does not apply.

3.1. Segregation without Matching: Stair Nesting

We choose to present an example with stair nesting instead of the usual nesting because stair nesting designs are unbalanced and use fewer observations than the balanced case, and in addition, the degrees of freedom for all factors are more evenly distributed, as was shown by Fernandes, Mexia, Ramos & Carvalho (2011). Cox & Solomon (2003) suggested that having u factors, we will have u steps where each step corresponds to one factor of the model.

In order to describe the branching in such models, we can consider $u+1$ steps. The first step, with index 0, has $a_0 = c_0 = u$ branches, one per factor. In the second step, with index 1, we have $c_1 = a(1) + u - 1$ branches, $a(1)$ the number of "active" levels for the first factor and $u-1$ the number of the remaining factors. We point out that the branch for the first factors concerns its "active" levels. For the third step, with index 2, we have $c(2) = a(1) + a(2) + u - 2$, where $a(1)$ represents the number of "active" levels for the first two factors resulting from the branching for the first factor; $a(2)$ is the number levels for the second factor and $u-2$, the number of the remaining factors. In this way, for the $(i+1)$-th step, with index i, we have $c(i) = \sum_{h=1}^{i} a(h) + u - i, i = 3, \ldots, u$ branches. $a(1), \ldots, a(i)$ are the number of "active" levels for the first i factors and $u-i$ the number of remaining factors. These designs are also studied in Fernandes, Ramos & Mexia (2010) and some results of nesting may be seen, for example, in Bailey (2004).

The model for stair nesting designs is given by

$$\mathbf{Y} = \sum_{i=0}^{u} \mathbf{X}_i \boldsymbol{\beta_i}, \tag{8}$$

with

$$\begin{cases} \boldsymbol{X}_0 = D(\mathbf{1}_{a(1)}, \ldots, \mathbf{1}_{a(i)}, \mathbf{1}_{a(i+1)}, \ldots, \mathbf{1}_{a(u)}) \\ \vdots \\ \boldsymbol{X}_i = D(\boldsymbol{I}_{a(1)}, \ldots, \boldsymbol{I}_{a(i)}, \mathbf{1}_{a(i+1)}, \ldots, \mathbf{1}_{a(u)}), i = 1, \ldots, u-1 \\ \vdots \\ \boldsymbol{X}_u = D(\boldsymbol{I}_{a(1)}, \ldots, \boldsymbol{I}_{a(i)}, \boldsymbol{I}_{a(i+1)}, \ldots, \boldsymbol{I}_{a(u)}) \end{cases} \tag{9}$$

where $D(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_u)$ is the block diagonal matrix with principal blocks $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_u$ and $\mathbf{1}_{a(s)}$ is the vector with all $a(s)$ components equal to 1.

In this approach we will assume that $\boldsymbol{\beta_0} = \mathbf{1}_u \boldsymbol{\mu}$, where $\boldsymbol{\mu}$ is the general mean value and the vectors $\boldsymbol{\beta_i}, i = 1, \ldots, u$, are independent normal with null mean vectors and variance-covariance matrix $\sigma_i^2 \mathbf{I}_{c(i)}, i = 1, \ldots, u$, and

$$c(i) = \sum_{h=1}^{i} a(h) + u - i, i = 1, \ldots, u$$

Hence $\mathbf{Y}$ is normal distributed with mean vector $\boldsymbol{\mu} = \mathbf{1}_n \mu$, and variance-covariance matrix $\mathbf{V} = \sum_{i=1}^{u} \sigma_i^2 \mathbf{M}_i$, where $\mathbf{M}_i = \mathbf{X}_i \mathbf{X}_i^{'}, i = 1, \ldots, u$, we have

$$\begin{cases} \boldsymbol{M}_0 = D(\boldsymbol{J}_{a(1)}, \ldots, \boldsymbol{J}_{a(i)}) \\ \qquad\qquad \vdots \\ \boldsymbol{M}_i = D(\boldsymbol{I}_{a(1)}, \ldots, \boldsymbol{I}_{a(i)}, \boldsymbol{J}_{a(i+1)}, \ldots, \boldsymbol{J}_{a(u)}), \quad i = 1, \ldots, u-1 \\ \qquad\qquad \vdots \\ \boldsymbol{M}_u = D(\boldsymbol{I}_{a(1)}, \ldots, \boldsymbol{I}_{a(i)}, \boldsymbol{I}_{a(i+1)}, \ldots, \boldsymbol{I}_{a(u)}) \end{cases} \tag{10}$$

with $\boldsymbol{J}_s = \mathbf{1}_s\mathbf{1}_s'$. Now, the orthogonal projection matrix on $r(\boldsymbol{X}_0)$, will be $\mathbf{T}$ given by

$$\mathbf{T} = D\left(\frac{1}{a(1)}\boldsymbol{J}_{a(1)}, \ldots, \frac{1}{a(i)}\boldsymbol{J}_{a(i)}, \frac{1}{a(i+1)}\boldsymbol{J}_{a(i+1)}, \ldots, \frac{1}{a(u)}\boldsymbol{J}_{a(u)}\right) \tag{11}$$

Moreover, with $\boldsymbol{K}_{a(i)} = \boldsymbol{I}_{a(i)} - \frac{1}{a(i)}\boldsymbol{J}_{a(i)}$ and $\mathbf{0}_{a(i)}$ the null $a(i) \times a(i)$ matrix, $i = 1, \ldots, u$, taking

$$\begin{cases} \boldsymbol{Q}_i = D(\mathbf{0}_{a(1)}, \ldots, \frac{1}{a(i)}\boldsymbol{J}_{a(i)}, \ldots, \mathbf{0}_{a(u)}), & i = 1, \ldots, u \\ \boldsymbol{Q}_{i+u} = D(\mathbf{0}_{a(1)}, \ldots, \boldsymbol{K}_{a(i)}, \ldots, \mathbf{0}_{a(u)}), & i = 1, \ldots, u \end{cases} \tag{12}$$

we will have

$$\begin{cases} \mathbf{T} = \sum_{j=1}^{u} \boldsymbol{Q}_j \\ \mathbf{M}_i = \sum_{j=1}^{i}(\boldsymbol{Q}_j + \boldsymbol{Q}_{j+u}) + \sum_{j=i+1}^{u} a(j)\boldsymbol{Q}_j, \quad i = 1, \ldots, u-1. \\ \mathbf{M}_u = \sum_{j=1}^{u}(\boldsymbol{Q}_j + \boldsymbol{Q}_{j+u}) \end{cases} \tag{13}$$

So we have

$$\boldsymbol{B} = \left[\begin{array}{cc} \boldsymbol{B_1} & \boldsymbol{B_2} \end{array}\right],$$

with

$$\boldsymbol{B_1} = \begin{bmatrix} 1 & a(2) & \ldots & a(u) \\ 1 & 1 & \ldots & a(u) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \ldots & a(u) \\ 1 & 1 & \ldots & 1 \end{bmatrix}, \qquad \boldsymbol{B_2} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 1 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \ldots & 0 \\ 1 & 1 & \ldots & 1 \end{bmatrix},$$

so we have segregation but we do not have matching.

Let us consider an example where $u = 3, a(1) = 3, a(2) = 2$ and $a(3) = 3$ "active" levels and the number of observations in the design is $n = 3 + 2 + 3 = 8$. So, we have $g(1) = 2, g(2) = 1$ and $g(3) = 2$ degrees of freedom for the first, second, and third factors, respectively. The design is shown in Figure 1.

The random effects model for stair nesting can be summarized as

$$\mathbf{Y} = \sum_{i=0}^{3} \mathbf{X}_i\boldsymbol{\beta}_i \tag{14}$$

where $a(1) = 3, a(2) = 2$ and $a(3) = 3$ are the levels for the 3 factors that nest. We make the same assumptions on the random effects as we did in the section 3.1,

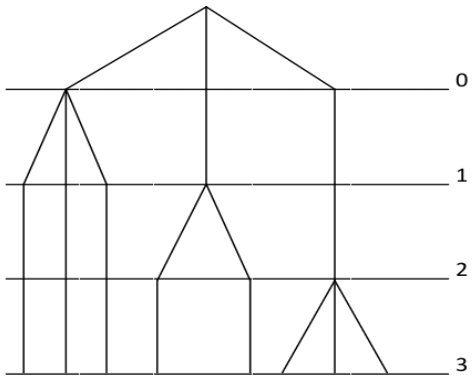


FIGURE 1: Stair nested design.

where

$$\begin{cases} \mathbf{X}_0 = D(\mathbf{1}_3, \mathbf{1}_2, \mathbf{1}_3) \\ \mathbf{X}_1 = D(\mathbf{I}_3, \mathbf{1}_2, \mathbf{1}_3) \\ \mathbf{X}_2 = D(\mathbf{I}_3, \mathbf{I}_2, \mathbf{1}_3) \\ \mathbf{X}_3 = D(\mathbf{I}_3, \mathbf{I}_2, \mathbf{I}_3) \end{cases} \tag{15}$$

From formula (13) we obtain

$$\begin{cases} \mathbf{M}_1 = D(\mathbf{I}_3, \mathbf{J}_2, \mathbf{J}_3) \\ \mathbf{M}_2 = D(\mathbf{I}_3, \mathbf{I}_2, \mathbf{J}_3) \\ \mathbf{M}_3 = D(\mathbf{I}_3, \mathbf{I}_2, \mathbf{I}_3) \end{cases} \tag{16}$$

Considering $m = 6, z = 3$, we have the pairwise orthogonal projection matrices

$$\begin{cases} \mathbf{Q}_1 = \{\frac{1}{3}\mathbf{J}_3, \mathbf{0}_2, \mathbf{0}_3\} \\ \mathbf{Q}_2 = \{\mathbf{0}_3, \frac{1}{2}\mathbf{J}_2, \mathbf{0}_3\} \\ \mathbf{Q}_3 = \{\mathbf{0}_3, \mathbf{0}_2, \frac{1}{3}\mathbf{J}_3\} \\ \mathbf{Q}_4 = \{\boldsymbol{K}_3, \mathbf{0}_2, \mathbf{0}_3\} \\ \mathbf{Q}_5 = \{\mathbf{0}_3, \boldsymbol{K}_2, \mathbf{0}_3\} \\ \mathbf{Q}_6 = \{\mathbf{0}_3, \mathbf{0}_2, \boldsymbol{K}_3\} \end{cases}$$

and the matrices

$$\begin{cases} \mathbf{M}_1 = \mathbf{Q}_1 + a(2)\mathbf{Q}_2 + a(3)\mathbf{Q}_3 + \mathbf{Q}_4 \\ \mathbf{M}_2 = \mathbf{Q}_1 + \mathbf{Q}_2 + a(3)\mathbf{Q}_3 + \mathbf{Q}_4 + \mathbf{Q}_5 \\ \mathbf{M}_3 = \mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{Q}_3 + \mathbf{Q}_4 + \mathbf{Q}_5 + \mathbf{Q}_6 \end{cases}$$

It follows readily that

$$\mathbf{B} = \begin{bmatrix} 1 & a(2) & a(3) & 1 & 0 & 0 \\ 1 & 1 & a(3) & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

considering

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B_1} & \boldsymbol{B_2} \end{bmatrix}$$

where

$$\boldsymbol{B_1} = \begin{bmatrix} 1 & a(2) & a(3) \\ 1 & 1 & a(3) \\ 1 & 1 & 1 \end{bmatrix}$$

and

$$\boldsymbol{B_2} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

3.2. Segregation without Matching: Crossing

Let there be a first factor that crosses with a second that nests a third. The factors will have a, b and c levels, respectively. The first and the third factors have random effects and the second has fixed effects.

The mean vector will then be

$$\boldsymbol{\mu} = (\mathbf{1}_a \otimes \mathbf{1}_b \otimes \mathbf{1}_c)\,\mu + (\mathbf{1}_a \otimes \mathbf{I}_b \otimes \mathbf{1}_c)\,\boldsymbol{\beta}\,(2)$$

where $\boldsymbol{\beta}\,(2)$ is the fixed vector of the effects for the second factor and $\otimes$ represent the Kronecker matrix product.

The random effects part of the model will be

$$(\mathbf{I}_a \otimes \mathbf{1}_b \otimes \mathbf{1}_c)\,\boldsymbol{\beta}\,(1) + (\mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{1}_c)\,\boldsymbol{\beta}\,(1,2) + (\mathbf{1}_a \otimes \mathbf{I}_b \otimes \mathbf{I}_c)\,\boldsymbol{\beta}\,(3) +$$

$$+ (\mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{I}_c)\,\boldsymbol{\beta}\,(1,3)\,,$$

where $\boldsymbol{\beta}\,(1)$, $\boldsymbol{\beta}\,(1,2)$, $\boldsymbol{\beta}\,(3)$ and $\boldsymbol{\beta}\,(1,3)$ correspond to the effects of the first factor, to the interactions of the first and second factors, to the effects of the third factor and to the interactions between the first and the third factors. As usual, we assume these vectors to be independent, homoscedastic and represent the corresponding

variance components by $\sigma^2(1)$, $\sigma^2(1,2)$, $\sigma^2(3)$ and $\sigma^2(1,3)$. So the variance-covariance matrix will be given by

$$\mathbf{V} = \sigma^2(1)\,\mathbf{I}_a \otimes \mathbf{J}_b \otimes \mathbf{J}_c + \sigma^2(1,2)\,\mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{J}_c + \sigma^2(3)\,\mathbf{J}_a \otimes \mathbf{I}_b \otimes \mathbf{I}_c + \sigma^2(1,3)\,\mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{I}_c.$$

In this case the matrices in the principal basis will be

$$\begin{cases} \mathbf{Q}_1 = \frac{1}{a}\mathbf{J}_a \otimes \frac{1}{b}\mathbf{J}_b \otimes \frac{1}{c}\mathbf{J}_c \\ \mathbf{Q}_2 = \mathbf{K}_a \otimes \frac{1}{b}\mathbf{J}_b \otimes \frac{1}{c}\mathbf{J}_c \\ \mathbf{Q}_3 = \frac{1}{a}\mathbf{J}_a \otimes \mathbf{K}_b \otimes \frac{1}{c}\mathbf{J}_c \\ \mathbf{Q}_4 = \mathbf{K}_a \otimes \mathbf{K}_b \otimes \frac{1}{c}\mathbf{J}_c \\ \mathbf{Q}_5 = \frac{1}{a}\mathbf{J}_a \otimes \frac{1}{b}\mathbf{J}_b \otimes \mathbf{K}_c \\ \mathbf{Q}_6 = \mathbf{K}_a \otimes \frac{1}{b}\mathbf{J}_b \otimes \mathbf{K}_c \end{cases}$$

Moreover the orthogonal projection matrix on Ω will be

$$\mathbf{T} = \frac{1}{a}\mathbf{J}_a \otimes \mathbf{I}_b \otimes \frac{1}{c}\mathbf{J}_c = \mathbf{Q}_1 + \mathbf{Q}_3.$$

We will also have

$$\begin{cases} \mathbf{I}_a \otimes \mathbf{J}_b \otimes \mathbf{J}_c = bc\mathbf{Q}_1 + bc\mathbf{Q}_2 \\ \mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{J}_c = c\mathbf{Q}_1 + c\mathbf{Q}_2 + c\mathbf{Q}_3 + c\mathbf{Q}_4 \\ \mathbf{J}_a \otimes \mathbf{I}_b \otimes \mathbf{I}_c = a\mathbf{Q}_1 + a\mathbf{Q}_3 + a\mathbf{Q}_5 \\ \mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{I}_c = \mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{Q}_3 + \mathbf{Q}_4 + \mathbf{Q}_5 + \mathbf{Q}_6 \end{cases}$$

Therefore

$$\mathbf{V} = \sum_{j=1}^{6} \gamma_j \mathbf{Q}_j,$$

with

$$\begin{cases} \gamma_1 = bc\sigma^2(1) + c\sigma^2(1,2) + a\sigma^2(3) + \sigma^2(1,3) \\ \gamma_2 = bc\sigma^2(1) + c\sigma^2(1,2) + \sigma^2(1,3) \\ \gamma_3 = c\sigma^2(1,2) + a\sigma^2(3) + \sigma^2(1,3) \\ \gamma_4 = c\sigma^2(1,2) + \sigma^2(1,3) \\ \gamma_5 = a\sigma^2(3) + \sigma^2(1,3) \\ \gamma_6 = \sigma^2(1,3) \end{cases}$$

Now γ_1 and γ_3 are different from all other canonical variance components so there is no matching. Despite this we have

$$\begin{cases} \sigma^2(1,3) = \gamma_6 \\ \sigma^2(3) = \frac{\gamma_5 - \gamma_6}{a} \\ \sigma^2(1,2) = \frac{\gamma_4 - \gamma_6}{c} \\ \sigma^2(1) = \frac{\gamma_2 - c\sigma^2(1,2) - \sigma^2(1,3)}{bc} = \frac{\gamma_2 - \gamma_4}{bc} \end{cases}$$

so all variance components either usual or canonic can be estimated.

4. Final Remarks

COBS models consider important cases. In the second example in Section 3 we presented an example of a balanced crossing which, (see Fonseca, Mexia & Zmyślony 2003, Fonseca, Mexia & Zmyślony 2007) can be extended to apply to all models with balanced cross nesting, thus including a wide variety of well behaved models.

The first example in section 3, that of stair nesting, displays a different model also with COBS. Besides the algebraic structure enables us to obtain unbiased estimators without normality. The LSE for estimable vectors are BLUE, whatever the variance components.

Acknowledgements

This work was partially supported by the center of Mathematics, University of Beira Interior, under the project PEst-OE/MAT/UI0212/2011.

We thank the anonymous referees and the Editor for useful comments and suggestions on a previous version of the paper, which helped to improve substantially the initial manuscript.

[Recibido: octubre de 2012 — Aceptado: septiembre de 2013]

References

Bailey, R. A. (2004), *Association Schemes: Designed Experiments, Algebra and Combinatorics*, Cambridge University Press, Cambridge.

Caliński, T. & Kageyama, S. (2000), *Block Designs: A Randomization Approach Vol. I: Analysis*, Springer-Verlag, New York.

Caliński, T. & Kageyama, S. (2003), *Block Designs: A Randomization Approach Vol. II: Analysis*, Springer-Verlag, New York.

Cox, D. & Solomon, P. (2003), *Components of Variance*, Chapman and Hall, New York.

Fernandes, C., Mexia, J., Ramos, P. & Carvalho, F. (2011), 'Models with stair nesting', *AIP Conference Proceedings - Numerical Analysis and Applied Mathematics* **1389**, 1627–1630.

Fernandes, C., Ramos, P. & Mexia, J. (2010), 'Algebraic structure of step nesting designs', *Discussiones Mathematicae. Probability and Statistics* **30**, 221–235.

Ferreira, S. S., Ferreira, D. & Mexia, J. T. (2007), 'Cross additivity in balanced cross nesting models', *Journal of Statistical Theory and Practice* (3), 377–392.

Ferreira, S. S., Ferreira, D., Nunes, C. & Mexia, J. T. (2010), 'Nesting segregated mixed models', *Journal of Statistical Theory and Practice* **4**(2), 233–242.

Fonseca, M., Mexia, J. T. & Zmyślony, R. (2003), 'Estimators and tests for variance components in cross nested orthogonal models', *Discussiones Mathematicae - Probability and Statistics* **23**(3), 175–201.

Fonseca, M., Mexia, J. T. & Zmyślony, R. (2007), 'Jordan algebras generating pivot variables and orthogonal normal models', *Journal of Interdisciplinary Mathematics* (10), 305–326.

Fonseca, M., Mexia, J. T. & Zmyślony, R. (2008), 'Inference in normal models with commutative orthogonal block structure', *Acta et Commentationes Universitatis Tartuensis de Mathematica* (12), 3–16.

Houtman, A. & Speed, T. (1983), 'Balance in designed experiments with orthogonal block structure', *Annals of Statistics* **11**(4), 1069–1085.

Mejza, S. (1992), 'On some aspects of general balance in designed experiments', *Statistica* **52**, 263–278.

Mexia, J. T., Vaquinhas, R., Fonseca, M. & Zmyslony, R. (2010), 'COBS: Segregation, Matching, Crossing and Nesting', *Latest Trends on Applied Mathematics, Simulation, Modeling, 4th International Conference on Applied Mathematics, Simulation, Modelling (ASM'10)* pp. 249–255.

Nelder, J. (1965*a*), 'The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance', *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **283**(1393), 147–162.

Nelder, J. (1965*b*), 'The analysis of randomized experiments with orthogonal block structure. II. Treatment structure and the general analysis of variance', *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **273**(1393), 163–178.

Seely, J. (1971), 'Quadratic subspaces and completeness', *The Annals of Mathematical Statistics* **42**, 710–721.

Zmyślony, R. (1978), 'A characterization of best linear unbiased estimators in the general linear model', *Mathematical Statistics and Probability Theory* **2**, 365–373.

Kendall, M. G. & Stuart, A. (1969), *The Advanced Theory of Statistics*, Vol. 1 of *Third edition*, Hafner Publishing Co., New York.

Konishi, S. (1978), 'An approximation to the distribution of the sample correlation coefficient', *Biometrika* **65**(3), 654–656.
*http://dx.doi.org/10.1093/biomet/65.3.654

Konishi, S. (1979*a*), 'Asymptotic expansions for the distributions of functions of a correlation matrix', *Journal of Multivariate Analysis* **9**(2), 259–266.
*http://dx.doi.org/10.1016/0047-259X(79)90083-6

Konishi, S. (1979*b*), 'Asymptotic expansions for the distributions of statistics based on the sample correlation matrix in principal component analysis', *Hiroshima Mathematical Journal* **9**(3), 647–700.
*http://projecteuclid.org/getRecord?id=euclid.hmj/1206134750

Konishi, S. & Sugiyama, T. (1981), 'Improved approximations to distributions of the largest and the smallest latent roots of a Wishart matrix', *Annals of the Institute of Statistical Mathematics* **33**(1), 27–33.
*http://dx.doi.org/10.1007/BF02480916

Kullback, S. (1967), 'On testing correlation matrices', *Applied Statistics* **16**, 80–85.

Kullback, S. (1997), *Information Theory and Statistics*, Dover Publications Inc., Mineola, NY. Reprint of the second (1968) edition.

Modarres, R. (1993), 'Testing the equality of dependent variances', *Biometrical Journal* **35**(7), 785–790.
*http://dx.doi.org/10.1002/bimj.4710350704

Modarres, R. & Jernigan, R. W. (1992), 'Testing the equality of correlation matrices', *Communications in Statistics. Theory and Methods* **21**(8), 2107–2125.
*http://dx.doi.org/10.1080/03610929208830901

Modarres, R. & Jernigan, R. W. (1993), 'A robust test for comparing correlation matrices', *Journal of Statistical Computation and Simulation* **43**(3–4), 169–181.

Muirhead, R. J. (1982), *Aspects of Multivariate Statistical Theory*, John Wiley & Sons Inc., New York. Wiley Series in Probability and Mathematical Statistics.

Schott, J. R. (2007), 'Testing the equality of correlation matrices when sample correlation matrices are dependent', *Journal of Statistical Planning and Inference* **137**(6), 1992–1997.
*http://dx.doi.org/10.1016/j.jspi.2006.05.005

Siotani, M., Hayakawa, T. & Fujikoshi, Y. (1985), *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook*, American Sciences Press Series in Mathematical and Management Sciences, 9, American Sciences Press, Columbus, OH.

Waternaux, C. M. (1984), 'Principal components in the nonnormal case: The test of equality of q roots', *Journal of Multivariate Analysis* **14**(3), 323–335. *http://dx.doi.org/10.1016/0047-259X(84)90037-X

Appendix

Lemma 3. *Let $V = (v_{ij})$ be a $p \times p$ symmetric matrix with zero on the diagonal and let $C = (c_{ij})$ be a $p \times p$ symmetric matrix. Then*

$$\operatorname{tr}(CV) = \sum_{i=1}^{p}\sum_{j=1}^{p} c_{ij}v_{ij} = 2\sum_{i<j} c_{ij}v_{ij}$$

Proof. The proof is obtained by noting that $v_{jj} = 0$ and $c_{ij} = c_{ji}$. □

Lemma 4. *Let $V_\alpha = (v_{\alpha ij})$ and $V_\beta = (v_{\beta ij})$ be $p \times p$ symmetric matrices with zero on the diagonal. Then*

$$\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{\ell=1}^{p} c_{ijk\ell}v_{\alpha ij}v_{\beta k\ell} = \sum_{i<j}\sum_{k<\ell}(c_{ijk\ell} + c_{ij\ell k} + c_{jik\ell} + c_{ji\ell k})v_{\alpha ij}v_{\beta k\ell}.$$

Proof. Using Lemma 3, the sum may be written as

$$\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k<\ell}(c_{ijk\ell} + c_{ij\ell k})v_{\alpha ij}v_{\beta k\ell}$$

The proof is obtained by applying Lemma 3 second time. □

Lemma 5. *Let A be a real symmetric matrix with eigenvalues that are less than one in absolute value, then*

$$-\ln[\det(I - A)] = \operatorname{tr}(A) + \frac{1}{2}\operatorname{tr}(A^2) + \frac{1}{3}\operatorname{tr}(A^3) + \cdots$$

Proof. See Siotani, Hayakawa and Fujikoshi (1985). □

Lemma 6. *Let R be a correlation of dimension p. Then*

$$\frac{\partial}{\partial P}\ln[\det R] = R^{-1}$$

and

$$\frac{\partial}{\partial P}\operatorname{tr}(R^{-1}B) = R^{-1}BR^{-1}$$

where B is a symmetric non-singular matrix of order p.

Proof. See Siotani, Hayakawa and Fujikoshi (1985). □

Lemma 7. *Let Y_1, Y_2 and Y_3 be independent random variables, $Y_i \sim \chi^2_2$, $i = 1, 2, 3$. Define $Y = \alpha_1 Y_1 + \alpha_2 Y_2 + \alpha_3 Y_3$ where α_1, α_2 and α_3 are constants, $\alpha_1 > \alpha_2 > \alpha_3 > 0$. Then, the cumulative distribution function $F_Y(y)$ of Y is given by*

$$F_Y(y) = \sum_{i=1}^{3} C_i \left[1 - \exp\left(-\frac{y}{2\alpha_i}\right)\right], \quad y > 0,$$

where $C_1 = \alpha_1^2/(\alpha_1 - \alpha_3)(\alpha_1 - \alpha_2)$, $C_2 = -\alpha_2^2/(\alpha_2 - \alpha_3)(\alpha_1 - \alpha_2)$ and $C_3 = \alpha_3^2/(\alpha_2 - \alpha_3)(\alpha_1 - \alpha_3)$

Proof. We get the desired result by inverting the moment generating function $M_Y(t) = \sum_{i=1}^{3} C_i(1 - 2\alpha_i t)^{-1}$, $2\alpha_1 t < 1$. □

Revista Colombiana de Estadística
Diciembre 2013, volumen 36, no. 2, pp. 273 a 286

Detection of Influential Observations in Semiparametric Regression Model

Detección de observaciones influenciales en modelos de regresión semiparamétricos

Semra Türkan[a], Öniz Toktamis[b]

Department of Statistics, The Faculty of Science, Hacettepe University, Ankara, Turkey

Resumen

In this article, we consider the semiparametric regression model and examine influential observations which have undue effects on the estimators for this model. One of the approaches to measure the influence of an individual observation is to delete the observation from the data. The most common measure based on this approach is Cook's distance. Recently, Daniel Peña introduced a new measure based on this approach. Pena's measure is able to detect high leverage outliers, which could be undetected by Cook's distance, in large data sets in linear regression model. The Cook's distances for parameter vector, unknown smooth function and response variable in semiparametric regression model are expressed by authors as functions of the residuals and leverages. Following the study of them we derive a type of Pena's measure as functions of the residuals and leverages for the same model. We compare the performance of these measures as to detection of influential observations using real data, artificial data and simulation. The results show that the performance of Pena's measure is better than Cook's distance to detect high leverage outliers in large data sets in the semiparametric regression model such as in the linear regression model.

***Palabras clave*:** Cook's distance, High leverage outliers, Pena's measure, Semiparametric regression.

Abstract

En este articulo, se consideran modelos de regresión semiparamétrica y se examinan observaciones influenciales que pueden tener efectos sobre los estimadores para este modelo. Una de las formas de medir la influencia de una observación individual es borrando la observación en el conjunto de

[a]Doctor. E-mail: sturkan@hacettepe.edu.tr
[b]Emeritus professor. E-mail: oniz@hacettepe.edu.tr

datos. La medida más común bajo esta idea es la distancia de Cook. Recientemente, Daniel Peña introdujo una nueva medida basada en estas ideas. Las distancias de Cook para el vector de parámetros, la función de suavizamiento y la variable respuesta en modelos de regresión semiparamétrica han sido expresadas por otros autores como funciones de los residuales y los puntos de apalancamiento. Se deriva en este artículo, una medida del tipo de la de Peña como función de los residuales y puntos de apalancamiento para el mismo modelo. Se compara el desempeño de estas medidas para la detección de observaciones influenciales usando datos reales y bajo simulación. Los resultados muestran que la medida de Peña es mejor que la distancia de Cook para detectar outliers y puntos de apalancamiento en conjuntos de datos grandes en los modelos de regresión semiparamétrica tales como el modelo de regresión lineal.

***Key words*:** distancia de Cook, outliers, puntos de apalancamiento, medida de Peña, regresión semiparamétrica.

1. Introduction

One or few observations could have serious effects on estimators. When an observation is omitted from the analysis, the fitted equation may change hardly at all. In this situation, the observation is considered as an influential observation. Hence, the detection of these observations has received a great deal of attention in the last decades. Numerous influence measures have been developed to detect these observations. Firstly, Cook (1977) introduced Cook's distance, which is based on deleting the observations one after another and measuring their effects in linear regression. Following the study of Cook (1977), most of ideas of detecting influential observations based on the deleting approach have developed. In recent years, Pena's measure is one of these ideas.

The study of influential observations has been extended to other statistical models using similar ideas such as in linear regression. However, most of the influence measures are concerned with parametric regression models. In recent years, the detection of influential observations in the nonparametric regression and semiparametric regression have been studied (see Thomas 1991, Kim 1996, Kim & Kim 1998, Kim, Park & Kim 2001, Zhu & Wei 2001, Kim, Park & Kim 2002, Zhang, Mei & Zhang 2007).

In this article, we consider the influence of individual cases on estimators in the semiparametric regression model and adjust the Pena's measure (Pena 2005) for this model. We compare the Pena's measure and some types of Cook's distances suggested by Kim et al. (2002) as to the success of detection of high leverages outliers in the semiparametric regression model.

The study is organized as follows. In Section 2, the semiparametric regression model is introduced. In Section 3, the formulas of Cook's distances for semiparametric regression model are given. In Section 4, Pena's measure formula for semiparametric regression is derived. In Section 5, the success of these measures

to detect influential observations, particularly high leverages outliers in large data, is analyzed via real data, artificial data and simulation.

2. Semiparametric Regression

Consider a semiparametric regression model with k explanatory variables

$$y_i = \mathbf{z_i}^T\boldsymbol{\beta} + m(x_i) + \varepsilon_i,\ (1 \leq i \leq n)$$

where y_i's are outcomes, $\mathbf{z_i}$ is a $k \times 1$ vector related to parametric component, x_i is a scalar, $\boldsymbol{\beta}$ is the $k \times 1$ vector of unknown parameters and m is a smooth unknown function. There are many approaches to estimate $\boldsymbol{\beta}$ and $\boldsymbol{m}$. The Speckman approach is one of them. Here, we follow the Speckman approach.

Let $\widetilde{\mathbf{Z}} = (\mathbf{I} - \mathbf{S})\mathbf{Z}$ and $\widetilde{\mathbf{y}} = (\mathbf{I} - \mathbf{S})\mathbf{y}$ where $\mathbf{S}$ is a smoother matrix. The local polynomial and the spline estimators are two classes of smoothers in semiparametric regression. Here, we use a local polynomial estimator. Hence, the $(1 \times n)$ jth row vector of $\mathbf{S}$ could be defined as $\mathbf{S}_{xj} = \mathbf{t}^T(\mathbf{X}_x^T\mathbf{W}_x\mathbf{X}_x)^{-1}\mathbf{X}_x^T\mathbf{W}_x$ where $\mathbf{X}_x$ is the $n \times (p+1)$ matrix with its ijth element equal to $(x_i - x)^{j-1}$, $\mathbf{W}_x = Diag(K_h(x_i - x))$ is the weight matrix with $K_h(.) = K(.\,|h)/h$ being a kernel function and h bandwidth controlling the size of the local neighborhood and $\mathbf{t}^T = \mathbf{t}_x^T(x) = (1, x - x, \ldots, (x - x)^p)$ is a vector. Here, it is assumed that K is a symmetric probability density function. The estimators of $\boldsymbol{\beta}$ and $\mathbf{m}$ suggested in Speckman (1988) are given by

$$\widehat{\boldsymbol{\beta}} = \left(\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{Z}}\right)^{-1}\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{y}} \tag{1}$$

$$\widehat{\mathbf{m}}(x) = \mathbf{S}\left(\mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\beta}}\right) = \mathbf{S}(\mathbf{I} - \widehat{\mathbf{H}})\mathbf{y} = \mathbf{H}^*\mathbf{y} \tag{2}$$

where $\widehat{\mathbf{H}} = (\mathbf{I} - \mathbf{S})^{-1}\widetilde{\mathbf{Z}}\left(\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{Z}}\right)^{-1}\widetilde{\mathbf{Z}}^T(\mathbf{I} - \mathbf{S})$ and $\mathbf{H}^* = \mathbf{S}\left(\mathbf{I} - \widehat{\mathbf{H}}\right)$. The vector of fitted values could be expressed from (1) and (2) as below

$$\begin{aligned}\widehat{\mathbf{y}} &= \mathbf{Z}\widehat{\boldsymbol{\beta}} + \widehat{\mathbf{m}}(x)\\ &= \breve{\mathbf{H}}\mathbf{y}\end{aligned} \tag{3}$$

where $\breve{\mathbf{H}}$ is considered as hat matrix in linear regression model defined $\breve{\mathbf{H}} = \widehat{\mathbf{H}} + \mathbf{H}^*$. The residual vector is given by

$$\breve{\mathbf{e}} = \mathbf{y} - \widehat{\mathbf{y}} = (\mathbf{I} - \breve{\mathbf{H}})\mathbf{y}$$

which will be used in defining and interpreting Cook's distances in the semiparametric regression model.

3. Cook's Distance

Firstly, we briefly review the derivation of Cook's distance in the linear regression model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{y}$ is a response vector, $\mathbf{X}$ is a $n \times k$ matrix of

known covariates, $\boldsymbol{\beta}$ is a vector of unknown parameters, and $\boldsymbol{\varepsilon}$ is a vector of errors with mean zero and a common unknown variance σ^2. y_i and $\mathbf{x}_i^T$ denote the ith row of $\mathbf{y}$ and $\mathbf{X}$, respectively, and using the subscript $(-i)$ means that the ith observation is deleted. Hence, $\mathbf{X}_{-i}$ denotes the matrix $\mathbf{X}$ with$i$th row deleted. Let $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ be the least squares estimator of $\boldsymbol{\beta}$, $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{H}\mathbf{y}$ where $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is the hat matrix and $s^2 = \mathbf{e}^T\mathbf{e}/(n-k)$ is estimation of σ^2.

Cook's distance for measuring the influence of the ith observation is defined by

$$C_i = (\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{-i})^T(\mathbf{X}^T\mathbf{X})(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{-i})/s^2 tr(\mathbf{H})$$

Using the fact,

$$\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{-i} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i e_i/(1-h_{ii})$$

the Cook's distance can be written as leverage values and residuals

$$C_i = \frac{1}{tr(\mathbf{H})s^2}\frac{e_i^2 h_{ii}}{(1-h_{ii}^2)} \tag{4}$$

where h_{ii} is the diagonal elements of $\mathbf{H}$ and e_i is the element of residual vector $\mathbf{e} = \mathbf{y} - \widehat{\mathbf{y}}$. The trace of $\mathbf{H}$ is defined to be the sum of the elements on the main diagonal of $\mathbf{H}$. As a projection matrix, $\mathbf{H}$ is symmetric and idempotent ($\mathbf{H}^2 = \mathbf{H}$), the eigenvalues of a projection matrix are either zero or one and the number of non zero eigenvalues is equal to the rank of the matrix. In this case, $rank(\mathbf{H}) = rank(\mathbf{X}) = k$ and hence, $trace(\mathbf{H}) = k$ which means that $tr(\mathbf{H}) = \sum_{i=1}^{n} h_{ii} = k$.

3.1. Cook's Distance for $\widehat{\beta}$ in Semiparametric Regression

An influence measure for ith observation on $\widehat{\boldsymbol{\beta}}$ may be defined as a type of Cook's distance in linear regression by

$$\widetilde{C_i} = \frac{(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{-i})^T(\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{Z}})(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{-i})}{s^2 tr(\widetilde{\mathbf{H}})} \tag{5}$$

Note that $tr(\widetilde{\mathbf{H}}) = \sum_{i=1}^{n} \widetilde{h}_{ii} = k$ as in linear regression. Equation (5) can be expressed as a function of the ith residual and leverage such as in (4) for semiparametric regression model as below

$$\widetilde{C}_i = \frac{1}{s^2 k}\frac{\widetilde{h}_{ii}\widetilde{e}_i^2}{(1-\widetilde{h}_{ii})^2} \tag{6}$$

where $\widetilde{e}_i$ is the ith component of residual vector $\widetilde{\mathbf{e}} = \mathbf{y} - \widetilde{\mathbf{y}}$ and $\widetilde{h}_{ii}$ is the ith diagonal component of $\widetilde{\mathbf{H}} = \widetilde{\mathbf{Z}}(\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{Z}})^{-1}\widetilde{\mathbf{Z}}^T$ related to parametric component of semiparametric regression model (Kim et al. 2002).

3.2. Cook's Distance for $\widehat{\mathbf{m}}$ in Semiparametric Regression

An influence measure for ith observation on $\widehat{\mathbf{m}}$ may be defined as a type of Cook's distance utilizing (2) by

$$C_i^* = \frac{\{\widehat{m}(x_i) - \widehat{m}_{-i}(x_i)\}}{s^2 tr(\mathbf{H}^*)}$$

It can be expressed as a function of the ith residual and leverage such as in (4)

$$C_i^* = \frac{(h_{ii}^* e_i^*)^2}{(1 - h_{ii}^*)^2 s^2 tr(\mathbf{H}^*)} \tag{7}$$

where e_i^* is the ith component of residual vector $\mathbf{e}^* = (\mathbf{I} - \mathbf{H}^*)\mathbf{y}$ and h_{ii}^* is the ith diagonal component of $\mathbf{H}^*$ related to the nonparametric component of the semiparametric regression model (Kim et al. 2002).

3.3. Cook's Distance for $\widehat{\mathbf{y}}$ in Semiparametric Regression

An influence measure for ith observation on $\widehat{\mathbf{y}}$ may be defined as a type of Cook's distance utilizing (3) such as in linear regression by

$$\breve{C}_i = \frac{(\widehat{\mathbf{y}} - \widehat{\mathbf{y}}_{-i})^T (\widehat{\mathbf{y}} - \widehat{\mathbf{y}}_{-i})}{s^2 tr(\breve{\mathbf{H}})}$$

It can be expressed as a function of the ith residual and leverage such as in (4) for $\widehat{y}$

$$\breve{C}_i = \frac{\breve{h}_{ii} \breve{e}_i^2}{(1 - \breve{h}_{ii})^2 s^2 tr(\breve{\mathbf{H}})} \tag{8}$$

where $\breve{e}_i$ is the ith component of residual vector $\breve{\mathbf{e}} = \mathbf{y} - \widehat{\mathbf{y}} = (\mathbf{I} - \breve{\mathbf{H}})\mathbf{y}$ and $\breve{h}_{ii}$ is the ith diagonal component of $\breve{\mathbf{H}}$ (Kim et al. 2002).

4. Pena's Measure

Pena (2005) introduced a new measure to determine the influence of an observation based on how this observation is being influenced by the rest of the data. That is, the predicted change when each observation in the data is deleted is measured for each observation. In this way, the sensitivity of each observation to changes in the data is measured. Pena (2005) showed that this type of influential analysis is able to indicate features in the data, such as clusters of high leverage outliers. Pena's measure has some advantages over Cook's distance. In a sample without outliers or high leverage observations, all of the cases have the the same expected sensitivity with respect to the entire sample. This is an advantage over Cook's distance which has an expected value that depends heavily

on the leverage of the case. For large sample sizes with many predictors, the distribution of the Pena's measure will be approximately normal. This is advantage over Cook's distance which has a complicated asymptotical distribution. The sample contaminated by a group of similar outliers with high leverages, this measure could discriminate between outliers and good observations while Cook's distance fails to detect these observations. In addition, Pena's measure can be useful for identifying intermediate-leverage outliers that are not detected by Cook's distance (Pena 2005).

In the regression model, Pena's measure is defined as

$$S_i = \frac{\mathbf{s}_i^T \mathbf{s}_i}{p s^2_{\widehat{(y_i)}}} \tag{9}$$

where $\mathbf{s}_i = (\widehat{y}_i - \widehat{y}_{i(1)}, \ldots, \widehat{y}_i - \widehat{y}_{i(n)})$ is a vector and $\widehat{y}_{i(j)}$ is the ith fitted value when the jth observation is deleted. Using the facts, the difference $\widehat{y}_i - \widehat{y}_{i(j)}$ is obtained as

$$\widehat{y}_i - \widehat{y}_{i(j)} = \mathbf{x}_i^T \widehat{\boldsymbol{\beta}} - \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}_{-j} = \frac{h_{jj} e_j}{1 - h_{jj}} \text{ and } s^2_{\widehat{(y_i)}} = s^2 h_{ii} \tag{10}$$

Pena's measure can be expressed as a function of the ith residual and leverage from (10)

$$S_i = \frac{1}{p s^2 h_{ii}} = \sum_{j=1}^{n} \frac{h_{ji}^2 e_j^2}{(1 - h_{jj})^2} \tag{11}$$

Pena (2005) stated that S_i would be large if it exceeds median (S_i)+$4.5MAD(S_i)$ where $MAD(S_i) = median\{|S_i - median(S_i)|\}/0,6745$. Pena's measure is very effective in detection of high leverage outliers that can not be detected by Cook's distance in large data sets. Also, it is very simple to compute (Türkan, S. and Toktamis, Ö. 2012).

4.1. Pena's Measure for Semiparametric Regression

In this study, we derived Pena's measure formula for the semiparametric regression model. The fitted values vector in (3) can be written as

$$\begin{aligned} \widehat{\mathbf{y}} &= \mathbf{Z}\boldsymbol{\beta} + \widehat{\mathbf{m}}(x) \\ &= \widetilde{\mathbf{Z}}\widehat{\boldsymbol{\beta}} + \mathbf{S}\mathbf{y} \end{aligned} \tag{12}$$

Using ith row vector of $\mathbf{S}$ in (12), $\mathbf{S}_{xi} = \mathbf{t}^T(\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x$, the ith fitted value, $\widehat{y}_i$, can be written

$$\widehat{y}_i = \widetilde{\mathbf{z}}_i^T \widehat{\boldsymbol{\beta}} + \mathbf{t}_{x_i}(x_i) \widehat{\boldsymbol{\beta}}_{x_i}$$

where $\widehat{\boldsymbol{\beta}}_x = (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \mathbf{y}$ and $t_x(x_i) = (1, (x_i - x), \ldots, (x_i - x)^p)$. The ith fitted value when jth observation is deleted, $\widehat{y}_{i,-j}$, can be expressed as below:

$$\widehat{y}_{i,-j} = \widetilde{\mathbf{z}}_i^T \widehat{\boldsymbol{\beta}}_{-j} + \mathbf{t}_{x_i}(x_i) \widehat{\boldsymbol{\beta}}_{x_i,-j} \tag{13}$$

Utilizing Sherman-Morrison-Woodbury (SMW) theorem, $\widehat{y}_i - \widehat{y}_{i,-j}$ can be obtained as a function of the ith residuals and leverages

$$\widehat{y}_i - \widehat{y}_{i,-j} = \frac{\widetilde{h}_{jj}\widetilde{e}_j}{1-\widetilde{h}_{jj}} + \frac{h_{x_i}(j,j)e_{x_i(j)}}{1-h_{x_i}(j,j)} \tag{14}$$

where $\widetilde{h}_{ij} = \widetilde{\mathbf{z}}_i^T(\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{Z}})^{-1}\widetilde{\mathbf{z}}_j$ and $h_{x_i}(i,i) = (\mathbf{X}_{x_i}^T\mathbf{W}_{x_i}\mathbf{X}_{x_i})^{-1}K_h(0)$ are diagonal elements of $\widetilde{\mathbf{H}} = \widetilde{\mathbf{Z}}(\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{Z}})^{-1}\widetilde{\mathbf{Z}}^T$ and $\mathbf{H}_x = \mathbf{X}_x(\mathbf{X}_x^T\mathbf{W}_x\mathbf{X}_x)^{-1}\mathbf{X}_x^T\mathbf{W}_x$, respectively. From (14), Pena's measure for semiparametric regression model can be obtained as

$$\begin{aligned}\widetilde{S}_i &= \frac{\mathbf{s}_i^T\mathbf{s}_i}{tr(\breve{\mathbf{H}})var(\widehat{y}_i)} \\ &= \frac{1}{tr(\breve{\mathbf{H}})var(\widehat{y}_i)}\sum_{j=1}^{n}\left(\frac{\widetilde{h}_{jj}\widetilde{e}_j}{1-\widetilde{h}_{jj}} + \frac{h_{x_i}(j,j)e_{x_i(j)}}{1-h_{x_i}(j,j)}\right)^2\end{aligned} \tag{15}$$

(see Türkan 2012)

5. Application

In this section, we compare the performance of our adjusted Pena's measure with adjusted Cook's distances in the semiparametric regression model to identify influential observations via actual data, artificial data and a simulation.

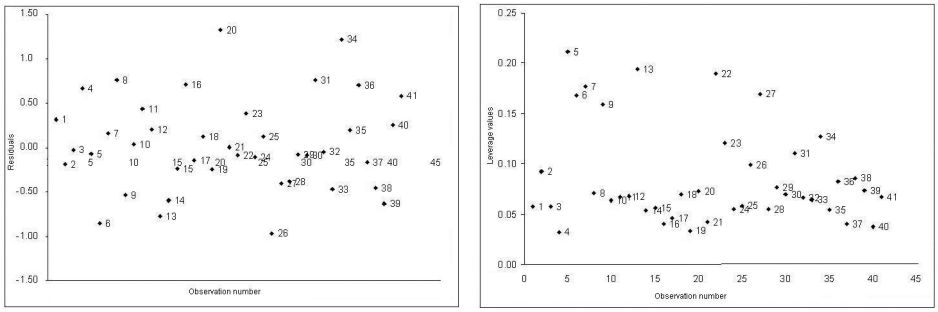
5.1. Actual Data

We consider actual data related to diabetes. The response variable is the logarithm of C-peptide concentration (y) at diagnosis and two predictors are age (x) and base deficit (z) (Kim et al. 2002). The data set contains 41 observations. There is a linear relationship between the logarithm of C-peptide concentration and base deficit, however, there is a nonlinear relationship between the logarithm of C-peptide concentration and age. Hence, the semiparametric regression model, $y_i = \mathbf{z}_i^T\boldsymbol{\beta} + m(x_i) + \varepsilon$, is used. Following the study of Kim et al. (2002), the local linear smoother was used and the bandwidth $h = 5.6$ was selected minimizing cross-validation (CV) criterion ($CV = \sum\{e_i/(1-h_{ii})\}^2$). Table 1 shows the estimates of both parametric and nonparametric components.

Figure 1 displays index plots of leverages values $\breve{h}_{ii}$ and residuals $\breve{e}_i$.

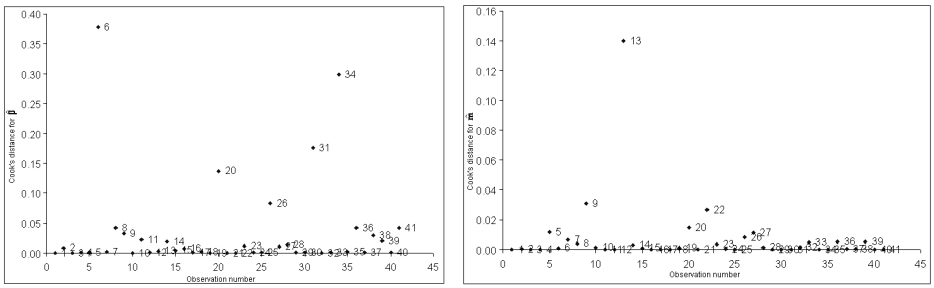
As seen from Figure 1(a), observations 20 and 34 are considered as outliers but these observations are not considered as high leverage from Figure 1(b) that the values of $\breve{h}_{ii}$ are not close to 1. Hence, it is said that there is no high leverage outlier in the data.

Figure 2 displays an index plot of influence measures ($\widetilde{C}, C_i^*, \breve{C}_i$ and $\widetilde{S}_i$) for this data.

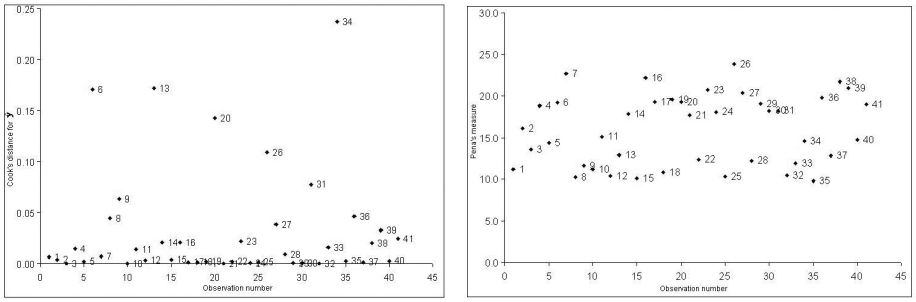


(a) (b)

FIGURE 1: (a) index plot of residuals, $\breve{e}_i$ (b) index plot of leverages values, $\breve{h}_{ii}$



(a) (b)



(c) (d)

FIGURE 2: Plots for diabetes data: (a) index plot of Cook's distance for $\widehat{\boldsymbol{\beta}}$, $\widetilde{C}_i$ (b) index plot of Cook's distance for $\widehat{\mathbf{m}}$, C_i^*, (c) index plot of Cook's distance for, $\widehat{\mathbf{y}}$, $\breve{C}_i$ (d) index plot of Pena's measure $\widetilde{S}_i$.

Table 1: Estimates of parametric and nonparametric components

Estimates of Parametric Component		Estimates of Nonparametric Component	
0.008	0.111	4.950	4.450
-0.501	0.312	5.206	5.345
0.339	0.261	05.279	5.319
-0.055	0.329	5.282	5.168
-0.539	-0.327	4.563	5.343
-0.711	0.286	5.332	5.342
-0.280	0.330	5.341	5.253
0.298	-0.430	5.003	5.295
0.366	0.323	4.617	5.327
0.033	-0.573	4.912	5.297
-0.369	0.181	5.156	4.941
0.213	-0.063	4.950	4.912
-0.079	-0.477	4.435	4.852
0.256	0.251	5.316	5.089
0.309	0.319	5.156	5.338
-0.133	0.210	5.309	5.257
-0.249	-0.407	5.282	5.329
0.404	0.251	5.191	5.338
0.036	-0.159	5.298	5.212
0.307	-0.382	5.333	5.289
0.176		5.304	

From Figure 2, according to Cook's distances ($\widetilde{C}$, C_i^* and $\breve{C}_i$) adjusted by Kim et al. (2002), observations 6, 34, 31, 20 and 26 are considered the five most influential observations on $\widehat{\boldsymbol{\beta}}$, observations 22, 13, 23, 26, 20 are considered the five most influential observations on $\widehat{\mathbf{m}}$ and observations 34, 6, 20, 26, 13 are considered the five most influential observations on $\widehat{\mathbf{y}}$. As seen from Figure 1(a), 1(b), there are no high leverage outliers in the data. Therefore, according to our adjusted Pena's measure $\widetilde{S}_i$, which is not useful in situations there are the outliers with low leverage, no observation is considered influential.

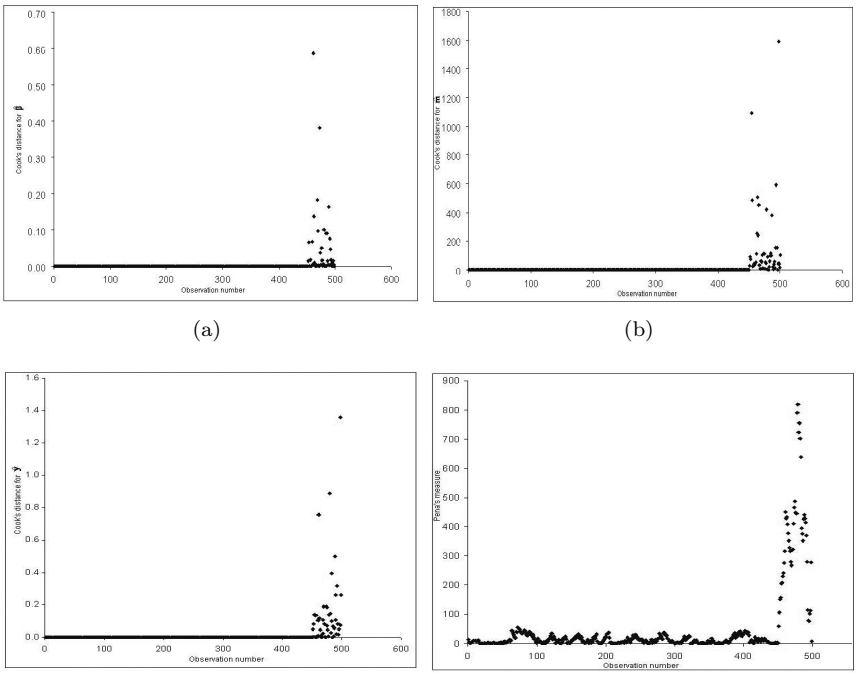
5.2. Artificial Data

Since we illustrate the performance of adjusted Pena's measure $\widetilde{S}_i$, an artificial data set with high leverage outliers is generated for semiparametric regression. We generate the data set using the model in the study of Kim et al. (2002)

$$y_i = 0.5\mathbf{z}_i + (x_i - 0.5)^2 + \varepsilon_i$$

We generate the 500 observations in which the last 50 observations would be high leverage outliers. For this reason, the first 450 of x_i from $U(0,1)$ and $z_i = i/450$ where ε_i is generated from $N(0, 0.02)$. The remaining 50 of x_i are generated from $U(5,10)$ and $z_i = i/50$ where ε_i is generated from $N(5,2)$. We suspect the last 50 observations for high leverage outliers. Figure 3 shows that the index plots of $\widetilde{C}$, C_i^*, $\breve{C}_i$ and $\widetilde{S}_i$.

As seen from Figure 3, $\widetilde{S}_i$ perfectly identifies 50 observations (observations $451 - 500$) as high leverage outliers. It is said that $\widetilde{S}_i$ is very useful for identifying high leverage outliers in semiparametric regression as in linear regression. In addition, $\widetilde{S}_i$ is clearly better than Cook's distances $(\widetilde{C}_i, C_i^*, \breve{C}_i)$ to detect high leverage outliers in large data as mentioned before.



(a) (b)

(c) (d)

FIGURE 3: Plots for Diabetes data: (a) index plot of Cook's distance for $\widehat{\boldsymbol{\beta}}$, $\widetilde{C}_i$ (b) index plot of Cook's distance for $\widehat{\mathbf{m}}$, C_i^*, (c) index plot of Cook's distance for, $\widehat{\mathbf{y}}$, $\breve{C}_i$ (d) index plot of Pena's measure $\widetilde{S}_i$.

5.3. Simulation Results

Here, we present a Monte Carlo simulation study that is designed to compare the performance of adjusted Pena's measure for semiparametric regression model. We generate the data sets from the same model in the previous section. We consider three different sample sizes, $n = 50$, 100, 250 with two different levels of influential observations ($i.e, \gamma = 10\%, 20\%$). The comparison of influence measures $(\widetilde{C}, C_i^*, \breve{C}_i$ and $\widetilde{S}_i)$ in semiparametric regression is carried out by the following steps:

1. Generation of the data with certain percentage of high leverages (X's outliers): For this purpose, we generate the first $n(1-\gamma)\%$ of x_i from $U(0,1)$

and $z_i = i/(n(1-\gamma)\%)$ where ε_i is generated from $N(0, 0.02)$. The remaining $n\gamma\%$ of x_i are generated from $U(5, 10)$ and $z_i = i/(n\gamma\%)$ where ε_i is generated from $N(0, 0.02)$.

2. Generation of the data with certain percentage of both high leverages (X's outliers) and outliers (Y's outliers): For this purpose, we generate the first $n(1-\gamma)\%$ of x_i from $U(0, 1)$ and $z_i = i/(n(1-\gamma)\%)$ where ε_i is generated from $N(0, 0.02)$. The remaining $n\gamma\%$ of x_i are generated from $U(5, 10)$ and $z_i = i/(n\gamma\%)$ where ε_i is generated from $N(5, 2)$.

3. Generation of the data with certain percentage of both intermediate-leverages and outliers (Y's outliers): For this purpose, we generate the first $n(1-\gamma)\%$ of x_i from $U(0, 1)$ and $z_i = i/(n(1-\gamma)\%)$ where ε_i is generated from $N(0, 0.02)$. The remaining $n\gamma\%$ of x_i are generated from $U(1, 3)$ and $z_i = i/(n\gamma\%)$ where ε_i is generated from $N(5, 2)$.

4. Generation of the data with certain percentage of low outliers: For this purpose, we generate the first $n(1-\gamma)\%$ of x_i from $U(0, 1)$ and $z_i = i/(n(1-\gamma)\%)$ where ε_i is generated from $N(0, 0.02)$. The remaining $n\gamma\%$ of x_i are generated from $U(1, 3)$ and $z_i = i/(n\gamma\%)$ where ε_i is generated from $N(1, 0.2)$.

5. Each measure is computed from each of the 100 replications.

6. Make comparison of detection of influential observations by using correct determination rate of each measure (i.e., total number of influential observations identified divided by total number of influential observations).

Table 2-5 show the correct determination rate of each measure ($\widetilde{C}, C_i^*, \breve{C}_i$ and $\widetilde{S}_i$) for different shows sizes and percentages of influential observations from 100 replications. From Table 2, adjusted Pena's measure, $\widetilde{S}_i$, performs similar results with Cook's distance $\breve{C}_i$ for $\widehat{y}$ to identify the high leverages for all the sample size. But, it is better than $\widetilde{C}_i$, C_i^* for all situations. From Table 3, adjusted Pena's measure, $\widetilde{S}_i$ clearly performs better than Cook's distances for $\widehat{\boldsymbol{\beta}}$, $\widehat{\mathbf{m}}$ and $\widehat{\mathbf{y}}$ ($\widetilde{C}_i$, C_i^*, $\breve{C}_i$) to detect high leverages outliers in large data. As seen from Table 3, almost all high leverage outliers could correctly be detected by $\widetilde{S}_i$ for $n = 250$. From Table 4, adjusted Pena's measure $\widetilde{S}_i$ successfully identifies intermediate leverage outliers that are not detected by Cook's distance for $n = 100$ and $n = 250$. From Table 5, adjusted Pena's measure $\widetilde{S}_i$ fails to detect low outliers with no high leverage as expected.

Table 2: The correct determination rate of high leverages (X's outliers).

Sample Size	Percentages of influential observations	Correct determination of measures (in percentages)			
		$\widetilde{C}_i$	C_i^*	$\breve{C}_i$	$\widetilde{S}_i$
n=50	10%	33	60	60	68
	20%	16	19	39	36
n=100	10%	23	11	39	45
	20%	17	14	38	35
n=250	10%	49	50	69	72
	20%	43	17	75	76

$\widetilde{C}_i$: Cook's distance for $\widehat{\boldsymbol{\beta}}$; C_i^*: Cook's distance for $\widehat{\mathbf{m}}$; $\breve{C}_i$: Cook's distance for $\widehat{\mathbf{y}}$; $\widetilde{S}_i$: Adjusted Pena's measure

Table 3: The correct determination rate of both high leverages (X's outliers) and outliers (Y's outliers).

Sample size	Percentages of influential observations	Correct determination of measures (in percentages)			
		$\widetilde{C}_i$	C_i^*	$\breve{C}_i$	$\widetilde{S}_i$
n=50	10%	51	70	72	80
	20%	46	44	68	84
n=100	10%	49	66	75	91
	20%	45	23	65	92
n=250	10%	52	52	71	98
	20%	44	19	62	98

$\widetilde{C}_i$: Cook's distance for $\widehat{\boldsymbol{\beta}}$; C_i^*: Cook's distance for $\widehat{\mathbf{m}}$; $\breve{C}_i$: Cook's distance for $\widehat{\mathbf{y}}$; $\widetilde{S}_i$: Adjusted Pena's measure.

Table 4: The correct determination rate of both intermediate leverages (X's outliers) and outliers (Y's outliers).

Sample size	Percentages of influential observations	Correct determination of measures (in percentages)			
		$\widetilde{C}_i$	C_i^*	$\breve{C}_i$	$\widetilde{S}_i$
n=50	10%	40	48	81	82
	20%	32	34	70	86
n=100	10%	32	39	77	86
	20%	23	27	66	89
n=250	10%	20	31	73	94
	20%	14	17	63	96

$\widetilde{C}_i$: Cook's distance for $\widehat{\boldsymbol{\beta}}$; C_i^*: Cook's distance for $\widehat{\mathbf{m}}$; $\breve{C}_i$: Cook's distance for $\widehat{\mathbf{y}}$; $\widetilde{S}_i$: Adjusted Pena's measure.

Table 5: The Correct Determination Rate of low outliers.

Sample size	Percentages of influential observations	Correct determination of measures (in percentages)			
		$\widetilde{C}_i$	C_i^*	$\breve{C}_i$	$\widetilde{S}_i$
n=50	10%	51	38	51	21
	20%	28	18	33	22
n=100	10%	39	43	47	13
	20%	25	19	30	4
n=250	10%	33	29	43	13
	20%	23	12	31	1

$\widetilde{C}_i$: Cook's distance for $\widehat{\boldsymbol{\beta}}$; C_i^*: Cook's distance for $\widehat{\mathbf{m}}$; $\breve{C}_i$: Cook's distance for $\widehat{\mathbf{y}}$; $\widetilde{S}_i$: Adjusted Pena's measure.

6. Conclusions

In this paper, we derived Pena's measure formula for semiparametric regression. The numerical examples and simulation study show that the proposed Pena's measure $\widetilde{S}_i$ performs very effectively in the identification of high leverage outliers and intermediate-leverage outliers in large data sets that are not clearly detected by adjusted Cook's distances for semiparametric regression model.

[Recibido: marzo de 2013 — Aceptado: junio de 2013]

References

Cook, R. (1977), 'Detection of influential observations in linear regression', *Technometrics* **19**, 15–18.

Kim, C. (1996), 'Cook's distance in spline smoothing', *Statistics and Probability Letters* **31**, 139–144.

Kim, C. & Kim, W. (1998), 'Some diagnostics results in nonparametric density estimation', *Communications in Statistics - Theory and Methods* **27**, 291–303.

Kim, C., Park, B. & Kim, W. (2001), 'Cook's distance in local polynomial regression', *Statistics & Probability Letters* **54**, 33–40.

Kim, C., Park, B. & Kim, W. (2002), 'Influential diagnostics in semiparametric regression models', *Statistics & Probability Letters* **60**, 49–58.

Pena, D. (2005), 'A new statistic for influence in linear regression', *Technometrics* **47**, 1–12.

Speckman, P. (1988), 'Kernel smoothing in partial linear models', *Journal of the Royal Statistical Society. Series B* **50**(3), 413–436.

Thomas, W. (1991), 'Influence diagnostics for the cross-validated smoothing parameter in spline smoothing', *Journal of the American Statistical Association* **86**(415), 693–698.

Türkan, S. (2012), Analysis of influential observation in semiparametric regression model, Doctoral Thesis, Hacettepe University, Faculty of Science. Department of Statistics, Ankara.

Türkan, S. and Toktamis, Ö. (2012), 'Detection of influential observations in ridge regression and modified ridge regression', *Model Assisted Statistics and Applications* **7**, 91–97.

Zhang, C., Mei, C. & Zhang, J. (2007), 'Influence diagnostics in partially varying-coefficient models', *Acta Mathematicae Applicatae Sinica* **23**(4), 619–628.

Zhu, Z. & Wei, B. (2001), 'Influence analysis in semiparametric nonlinear regression models', *Acta Mathematicae Applicatae Sinica* **24**(4), 568–581.

Revista Colombiana de Estadística
Diciembre 2013, volumen 36, no. 2, pp. 287 a 303

Censored Bimodal Symmetric-Asymmetric Alpha-Power Model

Modelo bimodal censurado simétrico-asimétrico alpha-potencia

Hugo S. Salinas[1,a], Guillermo Martínez-Flórez[2,b], Germán Moreno-Arenas[3,c]

[1]Departamento de Matemáticas, Facultad de Ingeniería, Universidad de Atacama, Copiapó, Chile

[2]Departamento de Matemáticas y Estadística, Facultad de Ciencias, Universidad de Córdoba, Córdoba, Colombia

[3]Escuela de Matemáticas, Universidad Industrial de Santander, Bucaramanga, Colombia

Abstract

We introduce the censored bimodal symmetric-asymmetric alpha-power models to adjust censored data with bimodality and high levels of skewness and kurtosis. The moments corresponding are computed, the maximum likelihood estimation for the model parameters is considered and the observed information matrix is derived. We show the appropriateness of the proposed models through two applications with censored real data related to HIV-1 RNA measurement.

***Key words*:** AART, alpha-power model, bimodality, censorship, cumulative distribution, HIV-1 RNA, limit of detection, power-normal model.

Resumen

Se introducen los modelos potencia alfa simétricos asimétricos bimodales censurados con el fin de ajustar datos censurados con bimodalidad y altos niveles de sesgo y curtosis. Los momentos correspondientes son calculados, se considera la estimación máximo verosímil para los parámetros del modelo y se deriva la matriz de información observada. Se muestra la utilidad de los modelos propuestos a través de dos aplicaciones con datos censurados reales relacionados con la medición de HIV-1 RNA.

***Palabras clave*:** AART, bimodalidad, censura, distribución acumulada, HIV-1 RNA, límite de detección, modelo alfa potencia, modelo normal potencia.

[a]Associate professor. E-mail: hugo.salinas@uda.cl
[b]Professor. E-mail: gmartinez@correo.unicordoba.edu.co
[c]Professor. E-mail: gmorenoa@uis.edu.co

1. Introduction

In epidemiological studies where biomarkers are the main outcomes, it is common to have detection limits below which it is not possible to determine the specific values. For instance, in highly active antiretroviral therapy (HAART), the number of viral load measurements in patients with HIV has a lower detection limit when using ultrasensitive tests.

The quantitative measurements in people with HIV may be highly left censored with a high percentage below the detection limit, depending on the method used for each measurement. For example, in Bucaramanga City, Colombia, the viral load measurements are conducted by different laboratories, and the HIV-1 RNA quantification is performed by three different methods: Versant bDNA 3.0® (Bayer), LCx HIV® (Abbott) and Amplicor HIV Monitor v1.5® (Roche), all of which have a detection limit of 50 copies per mL. In order to model the percentage of individuals below the detection limit, an asymmetric bimodal model may be necessary for this type of variable.

The analysis of viral load, HIV-RNA, (scale $\log_{10}$) is used to measure the effectiveness of HAART therapy which suppresses HIV-1 RNA to undetectable levels, thereby reducing the morbidity and mortality rates of HIV. Li, Chu, Gallant, Hoover, Mack, Chmiel & Muñoz (2006) found that $\log_{10}$(HIV-1 RNA) has two modal values in its distribution, corresponding to the optimal and suboptimal response to HAART, and it can be modeled with a mixture of two normal distributions in the presence of left censoring. In other cases, the bimodal behavior is also considered as the variable has a high (or low) degree of asymmetry and kurtosis in at least some partial distributions that compose the bimodal behavior.

In general a random variable Y, which has a part of its probability at discrete points and the rest spread over several intervals, has a mixture distribution.

When data are censored, the observed variable Y is a mixture of a continuous latent process Y^* and a selection mechanism (censoring or truncation) modeling in binary form. This idea was popularized by Tobin (1958) and the resulting model is known as the Tobit model, which is defined in terms of the latent variable $Y_i = Y_i^* I(Y_i^* > c)$, for some constant c, where $I(\cdot)$ is the indicator function and Y^* has a certain distribution, e.g., normal Tobin (1958) or Student-t of Arellano-Valle, Castro, González-Farías & Muñoz-Gajardo (2012) or generalized normal of Martínez-Flórez, Bolfarine & Gómez (2013).

Until the last two decades of the twentieth century, the inferential processes assumed the normality of the data under study. This assumption is unrealistic for many variables, and the inferential processes are inadequate. In these situations many authors choose to transform the variables in order to attain data symmetry or normality. This transformation leads to unsatisfactory results because the interpretation of results becomes cumbersome. The data becomes more difficult to interpret when there are several variables with different types of transformations. In view of these deficiencies, more flexible models have been developed that are able to accommodate different degrees of asymmetry and kurtosis. Previous work in this area include Azzalini (1985), Henze (1986), Durrans (1992), Fernández &

Steel (1998), Mudholkar & Hutson (2000), Pewsey (2000), Eugene, Lee & Famoye (2002), Jones (2004), Gómez, Venegas & Bolfarine (2007) and Arnold, Gómez & Salinas (2009).

For bimodal data, extensions for asymmetric cases have been studied by Kim (2005), Gómez et al. (2007) and Arnold et al. (2009), among others. Kim (2005) introduces the bimodal skew-normal called the *two-pieces skew-normal model*. An asymmetric extension of this model was presented by Arnold et al. (2009) who defined *the extended two-pieces skew-normal model*. Gómez, Elal-Olivero, Salinas & Bolfarine (2009) also studied a bimodal skew-normal model for certain values of the shape parameter, and this distribution is called *skew-flexible-normal*. Other works in this area have been published by Elal-Olivero, Gómez & Quintana (2009) and Bolfarine, Gómez & Rivas (2011).

In this paper, we present a new distribution for adjusted censored data with bimodality and high levels of skewness and kurtosis. The paper is structured as follows. In Section 2, we introduce the censored bimodal symmetric alpha-power distribution, moments, estimation and inference for model parameters. In Section 4, we introduce the censored bimodal asymmetric alpha-power distribution, moments, estimation and inference for model parameters; we derive the information matrix and discuss likelihood ratio tests for some hypotheses of interest. In Section 6, the appropriateness of this model is illustrated using two applications involving real data. Finally, some concluding remarks are presented in Section 7.

2. Censored bimodal symmetric alpha-power model

Based on the works by Durrans (1992) and Kim (2005), Bolfarine, Martínez-Flórez & Salinas (2012) introduced the bimodal symmetric alpha-power model, whose probability density is

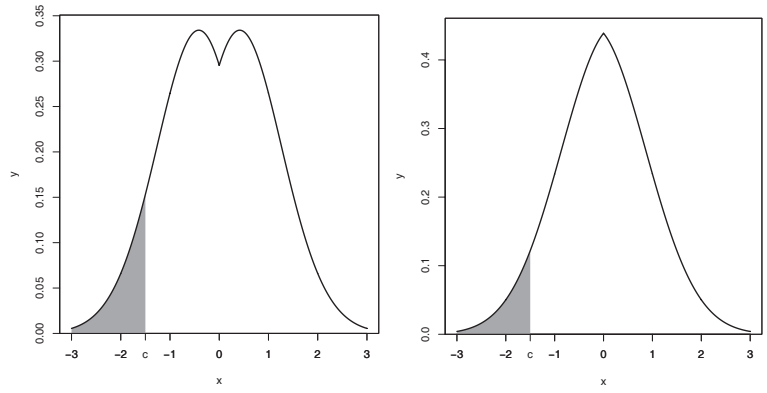
$$\varphi(z;\alpha) = \alpha c_\alpha f(z) \left\{F(|z|)\right\}^{\alpha-1}, \quad -\infty < z < \infty \tag{1}$$

where $\alpha \in \mathbb{R}^+$, F is an absolutely continuous density function with density function $f = dF$ symmetric around zero and $c_\alpha = \frac{2^{\alpha-1}}{2^\alpha - 1}$ is the normalizing constant. We use the notation $Z \sim BSP(\alpha)$.

Now, consider a random variable $Y^* \sim BSP(\alpha)$ where $(Y_1^*, Y_2^*, \ldots, Y_n^*)$ is a random sample of size n and point of censorship equal to c. Values of Y^* greater than the constant c are mapped to themselves, whereas values of Y^* less than or equal to the constant c are mapped to c. Then, without loss of generality for $c = 0$, the observed value is $Y_i = D_i Y_i^*$, $i = 1, 2, \ldots, n$, where $D_i = I(Y_i^* > 0)$. Here we have a random sample that is left censored. We say that Y follows a censored BSP distribution. We denote this variable by $Y \sim CBSP(\alpha)$. The generalization to right censoring or when the point of censorship is different from zero is trivial.

For a random variable $Y \sim CBSP(\alpha)$ with $\alpha \in \mathbb{R}^+$, the location-scale extension is defined as the distribution of the random variable $X = \xi + \eta Y$ for $\xi \in \mathbb{R}$ and $\eta > 0$. We use the notation $X \sim CBSP(\xi, \eta, \alpha)$.

From equation (1), when $f = \phi$ and $F = \Phi$ are the standard normal density and cumulative distribution functions, respectively, we obtain the bimodal power-normal density function and use the notation $Z \sim BPN(\alpha)$. Similarly, we obtain the censored bimodal power-normal density function $Y \sim CBPN(\alpha)$ and the location-scale extension $X \sim CBPN(\xi, \eta, \alpha)$. The density function of the random variable $Y \sim CBPN(\alpha)$ is symmetric and unimodal for $0 < \alpha \leq 1$ and bimodal for $\alpha > 1$. Figure 1 depicts plots for the random variable $Y \sim CBPN$ with a point of censorship $c \neq 0$ and two values of α.



(a) (b)

FIGURE 1: Densities of $CBPN(0, 1, \alpha)$ censored at the left (grey color): (a) $\alpha = 1.75$ and (b) $\alpha = 0.75$.

The moments of the random variable CBSP are given as functions of the incomplete moments of the alpha-power model which are defined as

$$\mu_r(x) = \int_x^\infty \alpha z^r f(z) \{F(z)\}^{\alpha-1} dz, \quad r = 0, 1, 2, \ldots,$$

The r-th moment of the random variable $X \sim CBSP(\xi, \eta, \alpha)$ is then given by

$$\mathbb{E}(X^r) = c_\alpha \sum_{k=0}^{r} \binom{r}{k} \xi^{r-k} \eta^k \mu_k(0)$$

3. Inference to CBSP Model

The contribution of the censored and uncensored observations to the log-likelihood function is as follows: if $Y_i = 0$, then $\mathbb{P}(Y_i = 0) = \mathbb{P}(X_i \leq 0) = c_\alpha \left[1 - \left\{F\left(\frac{\xi}{\eta}\right)\right\}^\alpha\right]$, and for the non-nulls Y_i's we have that they are distributed as the respective X_i's.

Assume that n independent and identically distributed observations $x_1, x_2, \ldots, x_n$ are available from $BSP(\xi, \eta, \alpha)$. We denote by $\sum_0$ the sum over the censored

observations and by $\sum_1$ the sum over uncensored observations. The log-likelihood function of (ξ, η, α) based on $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ is given by

$$\ell(\xi, \eta, \alpha; \boldsymbol{x}) = \sum_0 \left(\log(c_\alpha) + \log\left[1 - \left\{F\left(\frac{\xi}{\eta}\right)\right\}^\alpha\right]\right) + \sum_1 [\log(\alpha) + \log(c_\alpha) - \log(\eta) + \log(f(z_i)) + (\alpha - 1)\log(F(|z_i|))]$$

where $z_i = \frac{x_i - \xi}{\eta}$. Hence, assuming that f' exists, the score function defined as the first derivative of the log-likelihood function, with respect to all parameters is given by:

$$U(\xi) = -\frac{\alpha}{\eta} \sum_0 \frac{\left\{F\left(\frac{\xi}{\eta}\right)\right\}^{\alpha-1} f\left(\frac{\xi}{\eta}\right)}{1 - \left\{F\left(\frac{\xi}{\eta}\right)\right\}^\alpha} - \frac{1}{\eta} \sum_1 \left\{ \frac{f'(z_i)}{f(z_i)} + (\alpha - 1)\text{sgn}(z_i) \frac{f(|z_i|)}{F(|z_i|)} \right\}$$

$$U(\eta) = \frac{\alpha\xi}{\eta^2} \sum_0 \frac{\left\{F\left(\frac{\xi}{\eta}\right)\right\}^{\alpha-1} f\left(\frac{\xi}{\eta}\right)}{1 - \left\{F\left(\frac{\xi}{\eta}\right)\right\}^\alpha} - \frac{1}{\eta} \sum_1 \left\{ 1 + z_i \frac{f'(z_i)}{f(z_i)} + (\alpha - 1)|z_i| \frac{f(|z_i|)}{F(|z_i|)} \right\}$$

and

$$U(\alpha) = \sum_0 \left\{ -\frac{\log(2)}{2^\alpha - 1} - \frac{\left\{F\left(\frac{\xi}{\eta}\right)\right\}^\alpha \log\left[F\left(\frac{\xi}{\eta}\right)\right]}{1 - \left\{F\left(\frac{\xi}{\eta}\right)\right\}^\alpha} \right\} + \sum_1 \left\{ \frac{1}{\alpha} - \frac{\log 2}{2^\alpha - 1} + \log[F(|z_i|)] \right\}$$

The score equations are obtained by equating the derivatives presented above to zero. The maximum likelihood estimators are the solutions of the score equations, and clearly depend on the functions f and F. Model parameters are estimated using iterative algorithms that can be implemented in any statistical package. The elements of the observed information matrix are given in Appendix.

4. Censored Bimodal Asymmetric Alpha-Power model

The CBPN model is an alternative when data are censored and have a bimodal and symmetrical distribution; however, in case that the asymmetrical distributions are not adequate, we introduce another model for censored data whose distribution function is bimodal and asymmetric. The following lemma given by Azzalini (1985) will be essential to achieve this model.

Lemma 1. *Let f_0 be a probability density function symmetric about zero and G be a distribution function such that G' exists and is a probability density function symmetric about zero. Then $f_Z(z;\beta) = 2f_0(z)G(\beta z)$ is a probability density function for $z, \beta \in \mathbb{R}$.*

Based on this lemma and given that the density function of a random variable $BSP(\alpha)$ is symmetric about zero, then for G, which is a distribution function such that G' is a probability density function symmetric about zero, then

$$\varphi(z;\alpha,\beta) = 2\alpha c_\alpha f(z)\left\{F(|z|)\right\}^{\alpha-1} G(\beta z), \quad -\infty < z < \infty \tag{2}$$

is a probability density function, such that $\alpha \in \mathbb{R}^+$ and $\beta \in \mathbb{R}$. The parameter β controls asymmetric behavior. We denote by $Z \sim BAsP(\alpha,\beta)$.

The location-scale extension for a random variable $Z \sim BAsP(\alpha,\beta)$ is defined as the distribution of the random variable $X = \xi + \eta Z$, where $\xi \in \mathbb{R}$ is the location parameter and $\eta > 0$ for the scale parameter. We denote by $X \sim BAsP(\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\xi, \eta, \alpha, \beta)$. Thus, redefining the random variable latent $Y_i = X_i I(X_i > 0)$ we obtain a censored random variable, which we denote by $Y \sim CBAsP(\boldsymbol{\theta})$.

When $F = G = \Phi$ in equation (2) naturally follows the *censored bimodal asymmetric alpha-power normal model*, which we denote by $CBAsN(\boldsymbol{\theta})$, this distribution is bimodal for $\alpha > 1$ and unimodal for $0 < \alpha \leq 1$, while the parameter β controls asymmetric behavior.

Figure 2 depicts plots for the random variable $Y \sim CBAsN(\boldsymbol{\theta})$ with point of censorship $c \neq 0$ and two values of β.

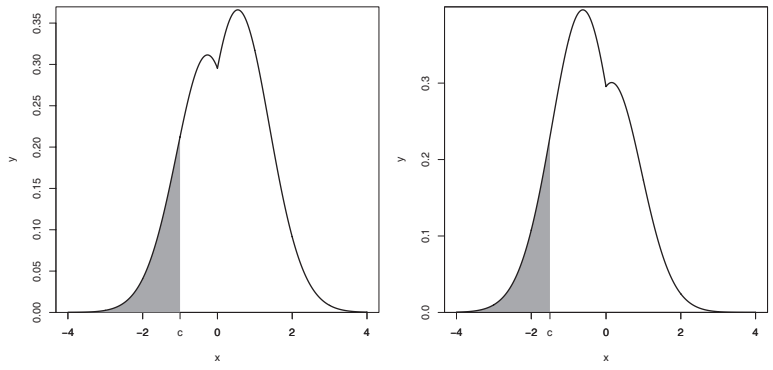


FIGURE 2: Density of $CBAsN(0, 1, 1.75, \beta)$ censored at the left (grey color). (a) $\beta = 0.25$ and (b) $\beta = -0.45$.

5. Inference to CBAsN Model

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n obtained from the $CBAsN(\boldsymbol{\theta})$ distribution with unknown parameter vector $\boldsymbol{\theta}$. The contribution of the i-th

observation to the likelihood is given by $\mathbb{P}(Y = 0) = \mathbb{P}(X \leq 0) = \alpha c_\alpha\, A_c(\boldsymbol{\theta})$ where $A_c(\boldsymbol{\theta}) = \int_{\frac{\xi}{\eta}}^{\infty} \phi(z)\{\Phi(z)\}^{\alpha-1}\{1-\Phi(\beta z)\}dz$.

The log-likelihood function of $\boldsymbol{\theta}$ based on $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ is given by

$$\begin{aligned}\ell(\boldsymbol{\theta};\boldsymbol{y}) &= \sum_{1} [\log(2\alpha c_\alpha) - \log(\eta) + \log(\phi(z_i)) + (\alpha-1)\log(\Phi(|z_i|)) + \log(\Phi(\beta z_i))] \\ &+ \sum_{0} [\log(\alpha c_\alpha) + \log A_c(\boldsymbol{\theta})]\end{aligned}$$

where $z_i = \frac{y_i - \xi}{\eta}$. The first derivatives of the log-likelihood function with respect to the parameters are given by:

$$\begin{aligned}U(\xi) &= -\frac{n_0 r_c(\boldsymbol{\theta})}{\eta A_c(\boldsymbol{\theta})} - \frac{1}{\eta}\sum_{1}\left\{-z_i + (\alpha-1)\mathrm{sgn}(z_i)\frac{\phi(|z_i|)}{\Phi(|z_i|)} + \beta\frac{\phi(\beta z_i)}{\Phi(\beta z_i)}\right\} \\ U(\eta) &= \frac{n_0 r_c(\boldsymbol{\theta})\xi}{\eta^2 A_c(\boldsymbol{\theta})} - \frac{1}{\eta}\sum_{1}\left\{1 - z_i^2 + (\alpha-1)|z_i|\frac{\phi(|z_i|)}{\Phi(|z_i|)} + \beta z_i\frac{\phi(\beta z_i)}{\Phi(\beta z_i)}\right\} \\ U(\alpha) &= n\left\{\frac{1}{\alpha} - \frac{\log 2}{2^\alpha - 1}\right\} + \frac{n_0 B_c(\boldsymbol{\theta})}{A_c(\boldsymbol{\theta})} + \sum_{1}\{\log[\Phi(|z_i|)]\}\end{aligned}$$

and

$$U(\beta) = \frac{n_0 D_c(\boldsymbol{\theta})}{A_c(\boldsymbol{\theta})} + \sum_{1} z_i\frac{\phi(\beta z_i)}{\Phi(\beta z_i)}$$

where

$$\begin{aligned}B_c(\boldsymbol{\theta}) &= \int_{\frac{\xi}{\eta}}^{\infty} \phi(z)\{\Phi(z)\}^{\alpha-1}\log(\Phi(z))\{1-\Phi(\beta z)\}dz, \\ r_c(\boldsymbol{\theta}) &= \phi\left(\frac{\xi}{\eta}\right)\left\{\Phi\left(\frac{\xi}{\eta}\right)\right\}^{\alpha-1}\left\{1-\Phi\left(\frac{\beta\xi}{\eta}\right)\right\}, \\ D_c(\boldsymbol{\theta}) &= \int_{\frac{\xi}{\eta}}^{\infty} z\phi(z)\{\Phi(z)\}^{\alpha-1}\{1-\Phi(\beta z)\}dz\end{aligned}$$

The maximum likelihood estimate $\widehat{\boldsymbol{\theta}} = (\hat{\xi}, \hat{\eta}, \hat{\alpha}, \hat{\beta})$ of $\boldsymbol{\theta}$ is obtained by setting $U(\xi) = U(\eta) = U(\alpha) = U(\beta) = 0$ and solving these equations numerically using iterative techniques. The elements of the observed information matrix are given in Appendix.

6. Illustrations

In this section we illustrate the usefulness of the proposed models by fitting a CBAsP distribution to some data sets. We use two real data sets to compare the fit of this model with censored normal (CN), censored skew-normal (CSN) and censored bimodal skew-normal (CBSN) distributions and with the parent distribution itself.

6.1. HIV-1 RNA Data Obtained from the Secretariat of Health of Bucaramanga City

The database was provided by Secretariat of Health, Department of Santander, Colombia, and corresponds to persons who are reported to the SIVIGILA system. This database maintains the absolute confidentiality of patient identification and contains the age, gender, date of admission to the SIVIGILA system, presence or absence of HAART treatment, CD-4 count and HIV-1 RNA plasma levels (viral load) of some patients. The database corresponds to 1275 persons infected with HIV, and who have been officially reported to the Surveillance and Epidemiology Service of Bucaramanga City. Tests used for the diagnosis of HIV infection in a particular person require a high degree of both sensitivity and specificity. In Colombia, this is achieved using an algorithm combining two tests for HIV antibodies. If antibodies are detected by an initial test based on the ELISA method, then a second test using the Western blot procedure is performed. The combination of these two methods is highly accurate. Patients are at different stages of treatment, 681 patients in the sample have had HAART therapy since 2007 and HIV-1 RNA plasma level (viral load) measurement, and there were 206 women and 475 men.

Because the measurements were performed at different laboratories, the HIV-1 RNA quantification could be performed by three different methods: Versant bDNA 3.0® (Bayer), LCx HIV® (Abbott) and Amplicor HIV Monitor v1.5® (Roche), all of which have a detection limit of 50 copies per mL. Descriptive statistics for $\log_{10}$(HIV-1 RNA) observations above the detection limit of 475 men in the example are mean=1.7350 and variance=1.7397. The skewness=0.5258 and kurtosis=2.1346 correspond to sample values above $\log_{10}(50)$. These statistics show that the data have a high positive bias and a low kurtosis compared to the normal model, which is an indication that the censored normal model is not an alternative to adjusting for viral loads. In addition to these characteristics, the histogram of Figure 3 shows that the behavior of the $\log_{10}$(HIV-1 RNA) variable is bimodal, and therefore the censored bimodal skew-normal model can be used to adjust $\log_{10}$(HIV-1 RNA) data. Furthermore, we adjust the censored normal (CN), censored skew-normal (CSN), censored bimodal symmetric skew-normal (CBPN) and censored bimodal asymmetric skew-normal (CBAsPN) models.

As can be seen in Figure 3-(a), the CSN model can accommodate to some degree the asymmetry that occurs in the observations, but it fails to explain the bimodality of the variable if it is adjusted for the CBPN and CBAsPN models.

To compare between the models considered above, we use the Akaike Information Criterion (AIC; Akaike 1974) and Bayesian Information Criterion (BIC; Schwarz 1978). Table 1 shows maximum likelihood estimates for the four adjusted models. According to the AIC and BIC criterions, the CBAsPN is a better fit for $\log_{10}$(HIV-1 RNA) data.

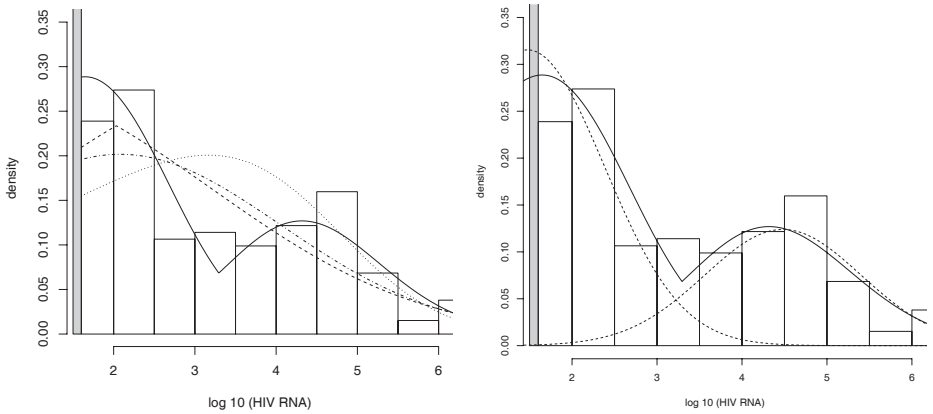
A parametric test to prove the bimodality hypothesis is given by $H_0 : \alpha = 1$ versus $H_1 : \alpha \neq 1$, which compares the CSN model with the CBAsPN model using the likelihood ratio statistics based on the ratio $\Lambda_1 = L_{CSN}(\widehat{\xi}, \widehat{\eta}, \widehat{\beta})/L_{CBAsPN}(\widehat{\xi}, \widehat{\eta},$

$\widehat{\alpha}, \widehat{\beta}$). Substituting the estimated values, we obtain $-2\log(\Lambda_1) = -2(-414.79 + 405.05) = 19.48$ which, when compared with the 95% critical value of $\chi^2_1 = 3.84$, indicate that the null hypotheses is clearly rejected. Then, the CBAsPN model is a good alternative for modeling $\log_{10}$(HIV-1 RNA) data.

Table 1: Maximum likelihood parameter estimates (Standard derivation in brackets) for CN, CSN, CBPN and CBAsPN models.

Model	CN	CSN	CBPN	CBAsPN
$\widehat{\xi}$	0.477(0.137)	1.689(1.147)	0.431(0.186)	1.692(0.085)
$\widehat{\eta}$	1.978(0.121)	2.362(0.767)	2.139(0.226)	1.549(0.120)
$\widehat{\alpha}$			0.396(0.576)	4.007(0.629)
$\widehat{\beta}$		–0.861 (1.013)		–0.595(0.100)
AIC	833.615	835.587	834.337	818.108
BIC	854.268	848.076	846.826	834.761

Additionally, we carry out the parametric test: $H_0 : (\alpha, \beta) = (1, 0)$ versus $H_1 : (\alpha, \beta) \neq (1, 0)$, which compares the CN model with the CBAsPN model. Using the statistic likelihood of ratio, $\Lambda_2 = L_{CN}(\widehat{\xi}, \widehat{\eta})/L_{CBAsPN}(\widehat{\xi}, \widehat{\eta}, \widehat{\alpha}, \widehat{\beta})$ leading to $-2\log(\Lambda_2) = -2(-414.81 + 405.05) = 19.52$, which is greater than the value of the chi-square with a 5% significance, $\chi^2_1 = 3.84$. This confirms that the best model to fit $\log_{10}$(HIV-1 RNA) data is the CBAsPN model. We can also observe that to some degree, the model adjusts the bimodality, but cannot adjust the asymmetry present in the observations of the viral load.



(a) (b)

Figure 3: (a) Histogram for $\log_{10}$(HIV-1 RNA): CBAsPN (solid line), CBPN (dashed line), CSN (dotted line) and CN (dashed line with points), (b) CBAsPN (solid line) CMN (dashed line).

Another model widely applied in such situations is the mixture model of two normal distributions (see Teck-Onn, Bakri, Morad & Hamid (2002), Chu, Moulton, Mack, Passaro, Barroso & Muñoz (2005), Li et al. (2006), Schneider, Margolick, Jacobson, Reddy, Martinez-Maza & Muñoz (2012), among others). The normal

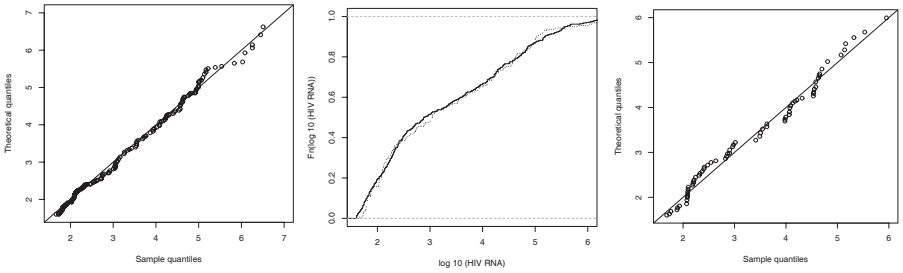
mixture model is given by

$$\rho(x;\mu_1,\sigma_1,\mu_2,\sigma_2,p) = pf_1(x,\mu_1,\sigma_1) + (1-p)f_2(x;\mu_2,\sigma_2)$$

where f_j is a normal distribution with parameters (μ_j,σ_j), $j=1,2$ and $0<p<1$. For data with detection limits, we denote them using the CMN$(\mu_1,\sigma_1,\mu_2,\sigma_2,p)$ model. Now we compare the CBAsPN with CMN$(\mu_1,\sigma_1,\mu_2,\sigma_2,p)$.

The estimated model is CMN$(1.48, 0.90, 4.48, 0.92, 0.71)$ with AIC=819.9 and BIC=840.7. This model has AIC and BIC greater than that of the CBAsPN model, so the CBAsPN model fits the data $\log_{10}$(HIV-1 RNA) better than the CMN model. Figure 3-(b) shows the estimated CBAsPN and CMN models. Furthermore, we studied the goodness of fit of the CBAsPN model getting the QQ-plot and cumulative distribution function from the MLE's.

The QQ-plot and the cumulative distribution function obtained from the estimated model are given in Figure 4(a)-(b): these show the good fit obtained in the estimated model. The total censored data corresponds to 39.92% of the sample under study, and the estimated percentage with the CBAsPN model is 39.50%, while in the CBPN model, it is 40.43%, which confirms the good fit of the CBAsPN model.



(a) (b) (c)

FIGURE 4: (a) QQ-plot men, (b) cumulative distribution function for men and (c) QQ-plot women.

These results indicate that the CBAsPN model is a suitable option for adjusting this type of information. In the case of HIV-infected women ($n = 106$) under HAART, 33.96% are below the detection limit. The estimated model was CBAsPN(1.6306, 1.8201, 2.8874, –0.5936), which estimated 32.95% of women below the detection limit. The QQ-plot given in Figure 4-(c) illustrates the good behavior of the CBAsPN model.

6.2. HIV-1 RNA Measuring by COBAS TaqMan

Plasma HIV-1 RNA was measured in 306 samples which were collected from 273 men in highly active antiretroviral therapy, with both Roche COBAS TaqMan (whose detection limit is 20 copies per mL) and Roche Amplicor (whose detection limit is 50 copies per mL) assays. See Schneider et al. (2012) for details.

The data used in this paper to illustrate the model are measurements made with the Roche TaqMan assay with $\log_{10}$(HIV-1 RNA). The non-censored information has a mean $\overline{y} = 1.3235$ and variance $s^2 = 1.5849$. Quantities $\sqrt{b_1} = 0.7012$ and $b_2 = 2.0054$ correspond to sample asymmetry and kurtosis coefficients for values above $\log_{10}(20)$, respectively. These statistics show that the data displays a high positive bias and a low kurtosis over the normal model. Figure 5 shows that the behavior of the variable under study is bimodal. Therefore, a censored bimodal asymmetric power-normal model may be used to adjust the $\log_{10}$(HIV-1 RNA) data. We adjusted the CSN and CBAsPN models.

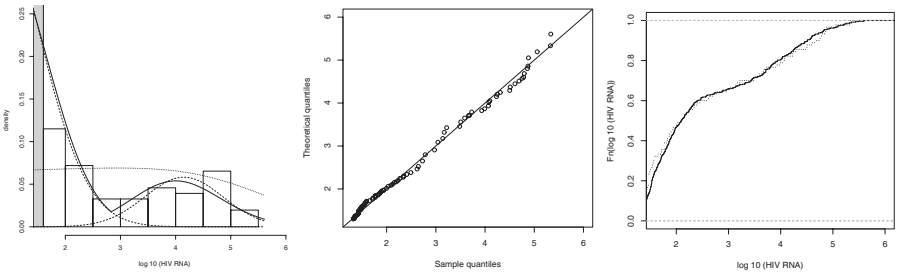
Table 2 shows maximum likelihood estimates of the proposed model. According to the AIC criterion, the model that best fits the $\log_{10}$(HIV-1 RNA) data is the CBAsPN normal model. The CSN model fails to capture the bimodality of the $\log_{10}$(HIV-1 RNA) data.

TABLE 2: Maximum likelihood parameter estimates (with (SD)) for CSN and CBAsPN models.

Model	$\widehat{\xi}$	$\widehat{\eta}$	$\widehat{\alpha}$	$\widehat{\beta}$	AIC
CSN	4.355(0.379)	11.121(1.371)		–9.637(3.274)	685.373
CBAsPN	1.531(0.090)	1.729(0.174)	6.400(0.901)	–1.175(0.148)	585.669

We can see that the estimate of α in the CSN model is significantly different from zero, which verifies the high degree of asymmetry present in the observations. Figure 5 shows that the CSN model adjusts to some extent the asymmetry present in the observations, but fails to explain the natural bimodality of the variable under study.

Again, we can prove the bimodality hypothesis $H_0 : \alpha = 1$ versus $H_1 : \alpha \neq 1$. Then, using the statistic likelihood of ratios, $\Lambda_3 = L_{CSN}(\widehat{\xi}, \widehat{\eta}, \widehat{\beta})/L_{CBAsPN}(\widehat{\xi}, \widehat{\eta}, \widehat{\alpha}, \widehat{\beta})$ and substituting the estimated values, we obtain $-2\log(\Lambda_3) = -2(-339.69 + 288.83) = 101.72$, which is greater than the value of the chi-square with 5% significance, $\chi^2_1 = 3.84$. Then the CBAsPN model is a good alternative for modelling $\log_{10}$(HIV-1 RNA) data.



(a) (b) (c)

FIGURE 5: (a) Histogram for $\log_{10}$(HIV-1 RNA), models: CBAsPN (solid line), CSN (dotted line) and CMN (dashed line), (b) QQ-plot and (c) cumulative distribution function for uncensored values.

We also obtained the estimate for the $CMN(0.577, 0.903, 4.15, 0.706, 0.897)$ model with AIC=585.27 (see Figure 5-(a)). There is no statistical difference between the AIC of the two models, and therefore, the two models have a similar fit. However, the CBAsPN model has fewer parameters, and is therefore less suitable than the CMN model.

Figure 5-(b)-(c) illustrate the QQ-plot and cumulative distribution function from the estimated model for uncensored data: these show the good fit of the estimated model. The total censored data corresponds to 70.58% of the study population, and the percentage estimated with the CBAsPN model is 70.69%, while with the CMN model, it is 70.74%, which illustrates the good fit of the CBAsPN model.

7. Concluding Remarks

We proposed two new distributions called the censored bimodal symmetric alpha-power and censored bimodal asymmetric alpha-power distributions. These distributions can adjust the skewness and bimodality of censored data. The inclusion of a new parameter can explain the asymmetric and bimodal behavior of an extended family of distributions, allowing a more flexible model than the censored normal, censored skew-normal models and censored mixture normal. The parameter estimation is approached by the maximum likelihood ratio and the observed information matrix is derived. Two real applications using data from HIV-infected persons illustrate the usefulness of the new models. The first application compares the censored normal, censored skew-normal and censored mixture normal with the two proposed models. The second application compares the censored skew-normal model and censored mixture normal with the CBAsPN model. The results show that the CBAsPN model fits much better to the viral load. The usefulness of the new models is tested with the likelihood ratio statistics and formal goodness-of-fit tests. The CBAsPN model has the potential to attract wider applications for censored data.

Acknowledgements

The authors acknowledge the comments and suggestions of the referees that helped to improve significantly our work. We also want to especially thank the Editor of this journal for having given suggestions and corrections of our manuscript. Moreno-Arenas thanks the Mobility Program of the Industrial University of Santander (Colombia) and the research of Salinas was supported by DIUDA 221229 (Chile).

[Recibido: enero de 2013 — Aceptado: agosto de 2013]

References

Akaike, H. A. (1974), 'A new look at statistical model identification', *IEEE Transaction on Automatic Control* **19**(6), 716–723.

Arellano-Valle, R., Castro, L., González-Farías, G. & Muñoz-Gajardo, K. (2012), 'Student-t censored regression model: Properties and inference', *Statistical Methods and Applications* **21**(4), 453–473.

Arnold, B. C., Gómez, H. W. & Salinas, H. S. (2009), 'On multiple constraint skewed models', *Statistics* **43**(3), 279–293.

Azzalini, A. (1985), 'A class of distributions which includes the normal ones', *Scandinavian Journal of Statistics* **12**(2), 171–178.

Bolfarine, H., Gómez, H. W. & Rivas, L. I. (2011), 'The log-bimodal-skew-normal model. A geochemical application', *Journal of Chemometrics* **25**(6), 329–332.

Bolfarine, H., Martínez-Flórez, G. & Salinas, H. S. (2012), 'Bimodal symmetric-asymmetric power-normal families', *Communications in Statistics-Theory and Methods* . DOI:10.1080/03610926.2013.765475.

Chu, H., Moulton, L. H., Mack, W. J., Passaro, D. J., Barroso, P. F. & Muñoz, A. (2005), 'Correlating two continuous variables subject to detection limits in the context of mixture distributions', *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **54**, 831–845.

Durrans, S. R. (1992), 'Distributions of fractional order statistics in hydrology', *Water Resources Research* **28**(6), 1649–1655.

Elal-Olivero, D., Gómez, H. W. & Quintana, F. A. (2009), 'Bayesian modeling using a class of bimodal skew-elliptical distributions', *Journal of Statistical Planning and Inference* **139**, 1484–1492.

Eugene, N., Lee, C. & Famoye, F. (2002), 'Beta-normal distribution and its applications', *Communications in Statistics–Theory and Methods* **31**(4), 497–512.

Fernández, C. & Steel, M. F. J. (1998), 'On Bayesian modeling of fat tails and skewness', *Journal of the American Statistical Association* **93**(441), 359–371.

Gómez, H. W., Elal-Olivero, D., Salinas, H. S. & Bolfarine, H. (2009), 'Bimodal extension based on the skew-normal distribution with application to pollen data', *Environmetrics* **22**(1), 50–62.

Gómez, H. W., Venegas, O. & Bolfarine, H. (2007), 'Skew-symmetric distributions generated by the distribution function of the normal distribution', *Environmetrics* **18**, 395–407.

Henze, N. (1986), 'A probabilistic representation of the skew-normal distribution', *Scandinavian Journal of Statistics* **13**(4), 271–275.

Jones, M. C. (2004), 'Families of distributions arising from distributions of order statistics', *TEST* **13**(1), 1–43.

Kim, H. J. (2005), 'On a class of two-piece skew-normal distribution', *Statistic* **39**(6), 537–553.

Li, X., Chu, H., Gallant, J. E., Hoover, D. R., Mack, W. J., Chmiel, J. S. & Muñoz, A. (2006), 'Bimodal virological response to antiretroviral therapy for HIV infection: an application using a mixture model with left censoring', *Journal of Epidemiology and Community Health* **60**(9), 811–818.

Martínez-Flórez, G., Bolfarine, H. & Gómez, H. W. (2013), 'The alpha-power tobit model', *Communications in Statistics-Theory and Methods* **42**(4), 633–643.

Mudholkar, G. S. & Hutson, A. D. (2000), 'The epsilon-skew-normal distribution for analyzing near-normal data', *Journal of Statistical Planning and Inference* **83**(2), 291–309.

Pewsey, A. (2000), 'Problems of inference for Azzalini's skew normal distribution', *Journal of Applied Statistics* **27**(7), 859–870.

Schneider, M., Margolick, J., Jacobson, L., Reddy, S., Martinez-Maza, O. & Muñoz, A. (2012), 'Improved estimation of the distribution of suppressed plasma HIV-1 RNA in men receiving effective antiretroviral therapy', *Journal of Acquired Immune Deficiency Syndromes* **59**(4), 389–392.

Schwarz, G. (1978), 'Estimating the dimension of a model', *Annals of Statistics* **6**(2), 461–464.

Teck-Onn, L., Bakri, R., Morad, Z. & Hamid, M. A. (2002), 'Bimodality in blood glucose distribution', *Diabetes Care* **25**(12), 2212–2217.

Tobin, J. (1958), 'Estimation of relationships for limited dependent variables', *Econometrica (The Econometric Society)* **26**(1), 24–36.

Appendix

Appendix A. Observed Information Matrix for CBSP Model

As is well known, the elements of the observed information matrix are computed as minus the second partial derivatives with respect to all parameters and are denoted by $j_{\xi\xi}, j_{\xi\eta}, \ldots, j_{\alpha\alpha}$. Assuming that f'' exists and making $w_i = \frac{f(|z_i|)}{F(|z_i|)}$, $s_i = \frac{f'(|z_i|)}{F(|z_i|)}$, $t_i = \frac{f''(z_i)}{f(z_i)}$, $v_i = \frac{f'(z_i)}{f(z_i)}$, $w_c = \frac{f\left(\frac{\xi}{\eta}\right)}{F\left(\frac{\xi}{\eta}\right)}$, $s_c = \frac{f'\left(\frac{\xi}{\eta}\right)}{F\left(\frac{\xi}{\eta}\right)}$, $p_c = \frac{\left\{F\left(\frac{\xi}{\eta}\right)\right\}^{\alpha}}{1-\left\{F\left(\frac{\xi}{\eta}\right)\right\}^{\alpha}}$, $q_c = \frac{f\left(\frac{\xi}{\eta}\right)}{1-\left\{F\left(\frac{\xi}{\eta}\right)\right\}^{\alpha}}$ and $u_c = \log\left(F\left(\frac{\xi}{\eta}\right)\right)$.

The elements of the observed information matrix are given by

$$j_{\xi\xi} = \frac{\alpha n_0 p_c}{\eta^2}\left[\alpha p_c w_c^2 + (\alpha-1)w_c^2 + s_c\right] + \frac{1}{\eta^2}\sum_1\left\{(v_i^2 - t_i) + (\alpha-1)\left[w_i^2 - s_i\right]\right\}$$

$$j_{\eta\xi} = -\frac{\alpha n_0}{\eta^3}\left[w_c^2(\alpha\xi(p_c+1)-\xi) + \eta w_c + \alpha\xi s_c\right] + \frac{1}{\eta^2}\sum_1\left\{(v_i + t_i - v_i^2) + (\alpha-1)\left[\operatorname{sgn}(z_i)|z_i|w_i^2 - \operatorname{sgn}(z_i)|z_i|s_i - \operatorname{sgn}(z_i)w_i\right]\right\}$$

$$j_{\eta\eta} = \frac{\alpha\xi n_0}{\eta^4}\left[w_c^2(\alpha\xi(p_c+1)-\xi) + 2\eta w_c + \alpha\xi s_c\right] - \frac{1}{\eta^2}\sum_1\left\{1 + \frac{1}{\eta^2}\left[2z_i v_i + z_i^2 t_i - z_i^2 v_i^2\right] + (\alpha-1)\left[2|z_i|w_i + z_i^2 s_i - z_i^2 w_i^2\right]\right\}$$

$$j_{\alpha\xi} = -\frac{n_0 p_c w_c}{\eta}\left[\alpha u_c(1+p_c)+1\right] - \frac{1}{\eta}\sum_1 \operatorname{sgn}(z_i)w_i,$$

$$j_{\alpha\eta} = \frac{n_0 p_c w_c \xi}{\eta^2}\left[\alpha u_c(1+p_c)+1\right] + \frac{1}{\eta}\sum_1 |z_i|w_i$$

and

$$j_{\alpha\alpha} = n\left[\alpha^{-2} - 2^{\alpha}(2^{\alpha}-1)^{-2}(\log 2)^2\right] + n_0 p_c u_c^2(1+p_c)$$

The elements of the expected (Fisher information matrix) are computed as n^{-1} times the expectation of the corresponding elements of the observed information matrix. This matrix clearly depends on the function f, and it is important in the sense that the asymptotic distribution of the maximum likelihood estimators is asymptotically normal with the asymptotic variance as the inverse of the Fisher information matrix.

Appendix B. Observed Information Matrix for CBAsN Model

Similarly, as done before, it follows that the elements of the observed information matrix are given by

$$j_{\xi\xi} = \frac{n_0 r_c(\boldsymbol{\theta})}{\eta^2 A_c(\boldsymbol{\theta})}\left\{\frac{r_c(\boldsymbol{\theta})}{A_c(\boldsymbol{\theta})} - \frac{\xi}{\eta} + (\alpha-1)w_c\right\} - \frac{\beta}{\eta n_0 A_c(\boldsymbol{\theta})}\phi\left(\frac{\xi}{\eta}\right)\phi\left(\frac{\beta\xi}{\eta}\right)\left\{\Phi\left(\frac{\xi}{\eta}\right)\right\}^{\alpha-1} + \frac{1}{\eta^2}\sum_1\left\{1 + (\alpha-1)\left[w_i^2 - \operatorname{sgn}(z_i)z_i w_i\right] + \beta^2\left[\beta z_i w_{1i} + w_{1i}^2\right]\right\}$$

$$
\begin{aligned}
j_{\eta\xi} &= -\frac{n_0 r_c(\boldsymbol{\theta})}{\eta^2 A_c^2(\boldsymbol{\theta})}\left[A_c(\boldsymbol{\theta}) + \frac{\xi}{\eta} r_c(\boldsymbol{\theta})\right] - \frac{n_0 \xi E_c(\boldsymbol{\theta})}{\eta^2 A_c(\boldsymbol{\theta})} + \frac{1}{\eta^2}\sum_1 \beta\big[\beta^2 z_i^2 w_{1i} + \beta z_i w_{1i}^2 \\
&\quad - w_{1i}\big], + \frac{1}{\eta^2}\sum_1 \left\{2z_i + (\alpha-1)\left[-z_i w_i^2 - \operatorname{sgn}(z_i) z_i^2 w_i + \operatorname{sgn}(z_i) w_i\right]\right\} \\
j_{\beta\xi} &= -\frac{n_0 \xi}{\eta^2}\frac{\phi\left(\frac{\xi}{\eta}\right)\phi\left(\frac{\beta\xi}{\eta}\right)\left\{\Phi\left(\frac{\xi}{\eta}\right)\right\}^{\alpha-1}}{A_c(\boldsymbol{\theta})} - \frac{n_0}{\eta}\frac{r_c(\boldsymbol{\theta} B_c(\boldsymbol{\theta}))}{A_c^2(\boldsymbol{\theta})} \\
&\quad + \frac{1}{\eta^2}\sum_1 \left\{\eta w_{1i} - \beta\left[\beta z_i^2 w_{1i} + z_i w_{1i}^2\right]\right\} \\
j_{\alpha\xi} &= -\frac{n_0 r_c(\boldsymbol{\theta})}{\eta\, A_c^2(\boldsymbol{\theta})}\left[B_c(\boldsymbol{\theta}) - A_c(\boldsymbol{\theta})\log\left(\Phi\left(\frac{\xi}{\eta}\right)\right)\right] - \frac{1}{\eta}\sum_1 \operatorname{sgn}(z_i) w_i \\
j_{\eta\eta} &= \frac{n_0 r_c(\boldsymbol{\theta})}{\xi\eta^4 A_c(\boldsymbol{\theta})}\left[2\eta - \xi\left(\frac{\xi}{\eta} - (\alpha-1)w_c\right) + \xi\frac{r_c(\boldsymbol{\theta})}{A_c(\boldsymbol{\theta})}\right] \\
&\quad - \frac{n_0\beta\xi^2}{\eta^4 A_c(\boldsymbol{\theta})}\phi\left(\frac{\xi}{\eta}\right)\phi\left(\frac{\beta\xi}{\eta}\right)\left\{\Phi\left(\frac{\xi}{\eta}\right)\right\}^{\alpha-1} \\
&\quad + \frac{1}{\eta^2}\sum_1 \left\{-1 + 3z_i^2 + (\alpha-1)\left[-2|z_i| w_i + z_i^2 w_i^2 + |z_i|^3 w_i\right] - \beta\eta z_i w_{1i}\right\} \\
&\quad + \frac{\beta}{\eta^2}\sum_1 \left[\beta^2 z_i^3 w_{1i} + \beta z_i^2 w_{1i}^2 - 2 z_i w_{1i}\right] \\
j_{\beta\eta} &= \frac{n_0\xi}{\eta^3 A_c(\boldsymbol{\theta})}\left[\eta r_c(\boldsymbol{\theta}) D_c(\boldsymbol{\theta}) + \xi\phi\left(\frac{\xi}{\eta}\right)\phi\left(\frac{\beta\xi}{\eta}\right)\left\{\Phi\left(\frac{\xi}{\eta}\right)\right\}^{\alpha-1}\right] \\
&\quad + \frac{1}{\eta}\sum_1 [z_i w_{1i} - \beta^2 z_i^3 w_{1i} - \beta z_i^2 w_{1i}^2] \\
j_{\alpha\eta} &= \frac{n_0 \xi r_c(\boldsymbol{\theta})}{\eta^2 A_c^2(\boldsymbol{\theta})}\left[B_c(\boldsymbol{\theta}) - A_c(\boldsymbol{\theta})\log\left(\Phi\left(\frac{\xi}{\eta}\right)\right)\right] + \frac{1}{\eta}\sum_1 |z_i| w_i \\
j_{\beta\beta} &= \frac{n_0}{A_c^2(\boldsymbol{\theta})}\left[D_c^2(\boldsymbol{\theta}) - A_c(\boldsymbol{\theta}) M_c(\boldsymbol{\theta})\right] + \sum_1 [\beta z_i^3 w_i + z_i^2 w_{1i}^2] \\
j_{\alpha\beta} &= \frac{n_0}{A_c^2(\boldsymbol{\theta})}\left[B_c(\boldsymbol{\theta}) D_c(\boldsymbol{\theta}) - A_c(\boldsymbol{\theta}) H_c(\boldsymbol{\theta})\right] \\
j_{\alpha\alpha} &= n\left[\alpha^{-2} - 2^{\alpha}(2^{\alpha}-1)^{-2}(\log 2)^2\right] + \frac{n_0}{A_c^2(\boldsymbol{\theta})}\left[B_c^2(\boldsymbol{\theta}) - A_c(\boldsymbol{\theta}) N_c(\boldsymbol{\theta})\right]
\end{aligned}
$$

where $w_{1i} = \phi(\beta z_i)/\Phi(\beta z_i)$,

$$
E_c(\boldsymbol{\theta}) = \frac{r_c(\boldsymbol{\theta})}{\eta^2}\left[-\xi + (\alpha-1)\eta w_c\right] - \frac{\beta}{\eta}\phi\left(\frac{\xi}{\eta}\right)\phi\left(\frac{\beta\xi}{\eta}\right)\left\{\Phi\left(\frac{\xi}{\eta}\right)\right\}^{\alpha-1}
$$

$$H_c(\boldsymbol{\theta}) = -\int_{\frac{\xi}{\eta}}^{\infty} z\phi(z)\{\Phi(z)\}^{\alpha-1}\log(\Phi(z))\phi(\beta z)\,dz$$
$$M_c(\boldsymbol{\theta}) = \beta\int_{\frac{\xi}{\eta}}^{\infty} z^3\phi(z)\{\Phi(z)\}^{\alpha-1}\phi(\beta z)\,dz$$
$$N_c(\boldsymbol{\theta}) = \int_{\frac{\xi}{\eta}}^{\infty} \phi(z)\{\Phi(z)\}^{\alpha-1}\log^2(\Phi(z))\{1-\Phi(\beta z)\}\,dz$$

The elements of the expected information matrix are computed numerically and depend on the functions ϕ and Φ. The MLE distribution is asymptotically normal with the variance as the inverse of the Fisher information matrix.

Revista Colombiana de Estadística
Diciembre 2013, volumen 36, no. 2, pp. 305 a 319

On an Improved Bayesian Item Count Technique Using Different Priors

Técnica de conteo de items bayesiana mejorada usando diferentes distribuciones a priori

Zawar Hussain[1,3,a], Ejaz Ali Shah[2,b], Javid Shabbir[1,c], Muhammad Riaz[1,4,d]

[1]Department of Statistics, Faculty of Natural Sciences, Quaid-i-Azam University, Islamabad, Pakistan

[2]Department of Statistics, Faculty of Sciences, University of Hazara, Mansehra, Pakistan

[3]Department of Statistics, Faculty of Sciences, King Abdulaziz University, Jeddah, Saudi Arabia

[4]Department of Mathematics and Statistics, Faculty of Sciences, King Fahad University of Petroleum and Minerals, Dhahran, Saudi Arabia

Abstract

Item Count Technique (ICT) serves the purpose of estimating the proportion of the people with stigmatizing attributes using the indirect questioning method. An improved ICT has been recently proposed in the literature (not requiring two subsamples and hence free from finding optimum subsample sizes unlike the usual ICT) in a classical framework that performs better than the usual ICT and the Warner method of Randomized Response (RR) technique. This study extends the scope of this recently proposed ICT in a Bayesian framework using different priors in order to derive posterior distributions, posterior means and posterior variances. The posterior means and variances are compared in order to study which prior is more helpful in updating the item count technique. Moreover, we have compared the Proposed Bayesian estimation with Maximum Likelihood (ML) estimation. We have observed that simple and elicited Beta priors are superior choices (in terms of minimum variance), depending on the sample size, number of items and the sum of responses. Also, the Bayesian estimation provides relatively more precise estimators than the ML Estimation.

***Key words*:** Bayesian Estimation, Indirect Questioning, Item Count Technique, Population Proportion, Prior Information, Privacy Protection, Randomized Response Technique, Sensitive Attributes.

[a]Professor. E-mail: zhlangah@yahoo.com
[b]Professor. E-mail: alishahejaz@yahoo.com
[c]Professor. E-mail: javidshabbir@gmail.com
[d]Professor. E-mail: riaz76qau@yahoo.com

Resumen

La técnica de conteo de ítems (ICT, por sus siglas en inglés) es útil para estimar la proporción de personas que poseen atributos que pueden tener algún grado de estigmatización mediante el uso de un método de preguntas indirectas. Una ICT mejorada ha sido propuesta recientemente en la literatura bajo la inferencia clásica (la cual no requiere dos submuestras y libre de la necesidad de encontrartamaños de muestra óptimos para cada una de ellas como sucede en la ICT usual). Esta ICT mejorada se desempeña mejor que la ICT usual y que el método de Respuesta Aleatorizada (RR, por sus siglas en inglés) de Warner. Este artículo extiende su estudio bajo una visión Bayesiana usando diferentes a priori con el fin de derivar distribuciones, medias y varianzas a posteriori.Las medias y varianzas a posteriori son comparadas con el fin de estudiar cuál a priori es más útil en mejorar la técnica de conteo de ítems. Se observa que a priori simples y Beta elicitadas son las mejores escogencias (en términos dela varianza mínima) dependiendo del tamaño de muestra, el número de ítems y la suma de la respuesta. También, la estimación bayesiana proporciona estimadores relativamente más precisas que la estimación ML.

***Palabras clave*:** atributos sensitivos, estimación Bayesiana, información a priori, preguntas indirectas, proporción poblacional, protección de la privacidad, técnica de conteo de ítems, técnica de respuesta aleatorizada.

1. Introduction

Survey techniques are now being utilized in almost every branch of physical and social sciences. These branches include medical, sociology, economics, agriculture, information technology, business, marketing, quality inspection, psychology, human behavior and many others. In surveys relating to these fields, especially, sociology, psychology, economics, people do not report their true status when the study question is sensitive in nature. Collection of trustworthy (truthful) data mainly depends upon the sensitivity of the study question, survey method, privacy (confidentiality) and cooperation of the respondents. The cooperation from the respondents will be low if the study question is sensitive and direct questioning method is applied. Consequently, the inferences made through direct questioning run the risk of response bias, non response (refusal) bias or both. An ingenious method pioneered by Warner (1965) was suggested in anticipation of reducing these biases and to provide more confidentiality to respondents.

The technique proposed by Warner (1965) is known as Randomized Response Technique (RRT). A comprehensive review of developments on Randomized Response (RR) techniques is given by Tracy & Mangat (1996) and Chaudhri & Mukerjee (1998). Some of the recent developments, among others, include Gupta, Gupta & Singh (2002), Ryu, Kim, Heo & Park (2005-2006), Bar-Lev, Bobovitch & Boukai (2004), Arnab & Dorffner (2006), Huang (2010), Hussain & Shabbir (2010), Barabesi & Marcheselli (2010) and Chaudhuri (2011). A number of applications of RRT can be found in the literature, for instance, Liu & chow. L. P. (1976), Reinmuth & Guerts (1975), Guerts (1980), Larkins, Hume & Garcha (1997), etc.

Although these studies were seen to be fruitful in the sense of estimation of the parameters, there are some applied difficulties associated with RRT as reported by Guerts (1980) and Larkins et al. (1997). Guerts (1980) found that RRT could have some limitations such the requirement of increased sample sizes in order to have confidence intervals as good as obtained through the direct questioning technique. More time is needed to administer and explain the procedure to the survey respondents. He further argued that, compilation of the results in the form of tables is somewhat protracted.

Larkins et al. (1997) were of the view that RRT was not suitable in the estimation of population proportion of tax payers/non-payers. Dalton & Metzger (1992) found that RRT might not be efficient in a mailed or telephonic survey. Similarly, Hubbard, Casper & Lesser (1989) argued that the major problem for RRT is to choose a randomization device to apply as a best one in specified circumstances and the very decisive feature of an RRT is about the respondent's acceptance of the technique. More recently, Chaudhuri & Christofides (2007) criticized RRT arguing that it is burdened with the respondent's ability to understand and handling of the device and also it asks respondents to report the information which may be useless or tricky. An intelligent interviewee may fear that his/her response can be traced back to his/her true status if he/she does not understand the mathematical logic behind the randomization device. Owing to these difficulties and limitations associated with $RRTs$, alternative techniques have been suggested. Some of these include the Item Count Technique by Droitcour, Casper, Hubbard, Parsley, Visscher & Ezzati (1991), the Three Card Method by Droitcour, Larson & Scheuren (2001) and the Nominative Technique by Miller (1985). These alternatives were suggested to avoid evasive answers on sensitive questions particularly concerning private issues, communally unexpected behaviors or illegitimate acts. Chaudhuri & Christofides (2007) also supplemented such an idea.

If some prior information is available about the mean of the study variable it may be used together with sample information. One of the methods using the prior knowledge is the Bayesian method of estimation where prior knowledge is used in the form of prior distribution. It has been established through many studies that when prior information is more informative the Bayesian estimation provides the more precise estimators.

In this paper, we plan to do a a Bayesian analysis of a recent item count technique by Hussain, Shah & Shabbir (2012) and provide the Bayesian estimators assuming that prior information is available through the past studies, past experience or simply through intelligent guess. Specifically, we will consider some prior distributions and compare the Bayesian estimator in case of each prior distribution used in this study. These comparisons will be in anticipation of finding the more suitable prior. The paper is organized as: Section 2 discusses the recent technique by Hussain et al. (2012); Section 3 provides Bayesian estimation using different priors; Section 4 presents a comparative analysis, concluding remarks are furnished in Section 5.

2. A Recent Item Count Technique

Hussain et al. (2012) proposed an improved item count technique based on single sample of size n in a classical framework showing an improvement over the usual ICT and the novel method of Randomized Response (RR) technique of Warner (1965). The said technique does not require two subsamples and consequently finding optimum subsample sizes is not needed. This study extends the scope of their study in a Bayesian framework and investigates the choice of a suitable prior to update the item count technique.

In the improved ICT of Hussain et al. (2012), each respondent is provided a list of g items and asked to report the number of items applicable to him/her, where each item is a combination of an unrelated item say F_j and a sensitive characteristic say S. The i^{th} respondent is asked to count 1, if he /she possess at least one of the characteristics F_j and S, and count 0 otherwise and finally report the total count. So, for a single respondent his/her response may be 0 to g. The response 1 for a single question or item means the respondent belongs either to non sensitive characteristic, sensitive characteristic or to both. Now the probability of 1 for j^{th} item is given by:

$$P(1) = \theta_j = \theta_{F_j} + \pi - \pi\theta_{F_j} \tag{1}$$

where θ_{F_j} denotes the proportion of j^{th} innocuous characteristic andπ denotes population proportion of individuals possessing a sensitive characteristic. Let Y_i be the response of i^{th} respondent, then it can be written as: $Y_i{=}\sum_{j=1}^{g}\alpha_j$, where α_j is a Bernoulli random variable taking values 1 and 0 with probabilities θ_j and $(1-\theta_j)$ respectively. The unbiased moment (and ML) estimator for proportion of people bearing sensitive behavior is given as:

$$\hat{\pi}_M = \left(\bar{y} - \sum_{j=1}^{g}\theta_{F_j}\right)\left(g - \sum_{j=1}^{g}\theta_{F_j}\right)^{-1} \tag{2}$$

with variance given by:

$$Var(\hat{\pi}_M) = \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)}{n(g-\sum_{j=1}^{g}\theta_{F_j})^2}\{\sum_{j=1}^{g}\theta_{F_j}(1-\sum_{j=1}^{g}\theta_{F_j}) + 2\sum_{j<k}^{g}\theta_{F_j}\theta_{F_k}\} \tag{3}$$

In order to have Y_i as a binomial random variable we take $\theta_j = \theta$ (or equivalently $\theta_{F_j} = \theta_F$) for all $j = 1, 2, \ldots, g$ such that $\theta_{F_j} = \frac{1}{g}$. In this case variance of ML estimator turns out to be

$$Var(\hat{\pi}_M) = \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)}{ng(g-1)} \tag{4}$$

Now we develop Bayesian estimation of population proportion through the above mentioned item count technique of Hussain et al. (2012) by assuming that $\theta_j = \theta$ for all $j = 1, 2, \ldots, g$. We use different prior distributions for deriving

posterior distributions in order to find which posterior distribution gives high posterior probability for higher estimates of π. Prior distributions used here are Beta distribution with known hyper parameters, Non-informative Uniform distribution, Non-informative Haldane distribution, Mixture of Beta distributions and a Beta distribution with elicited hyperparameters. The posterior distribution using density kernel is defined as:

$$P(\pi|y) \propto L(y,\pi)P(\pi) \tag{5}$$

where $L\left(y,\pi\right)$ is the likelihood function and $P(\pi)$ is the prior distribution. Since α_j is the Bernoulli random variable with parameter $\theta_j = \theta$ the response variable Y_i is a binomial random variable with parameter g and θ. Thus the likelihood function becomes:

$$L\left(y,\pi\right) = \prod_{i=1}^{n} \left\{ \binom{g}{y_i} \theta^{y_i} \left(1-\theta\right)^{g-y_i} \right\} \tag{6}$$

where $\theta = \theta_F + \pi\left(1-\theta_F\right)$ Substituting $\theta = \theta_F + \pi\left(1-\theta_F\right)$ in above equation and taking $d = \frac{\theta_F}{(1-\theta_F)}$, we get

$$L\left(y,\pi\right) = \left(1-\theta_F\right)^{ng} \left\{ \prod_{i=1}^{n} \binom{g}{y_i} \right\} \left(d+\pi\right)^{n\bar{y}} \left(1-\pi\right)^{ng-n\bar{y}} \tag{7}$$

3. Bayesian Estimation using Different Priors

In this section, we derive the Bayesian estimators of π assuming different prior distributions mentioned above in Section 2.

3.1. Beta Prior

Suppose the prior distribution of $\pi$is given by:

$$P(\pi) = \frac{1}{B\left(a,b\right)} \pi^{a-1} \left(1-\pi\right)^{b-1}, \qquad 0 < \pi < 1 \tag{8}$$

where $B\left(a,b\right) = \int_0^1 \pi^{a-1}\left(1-\pi\right)^{b-1} d\pi$ is a complete Beta function.
Thus, using (7) and (8) in (5) the posterior distribution of π is derived as:

$$P\left(\pi|\, y\right) \propto \left(1-\theta_F\right)^{ng} \left\{ \prod_{i=1}^{n} \binom{g}{y_i} \right\} \left(d+\pi\right)^{n\bar{y}} \left(1-\pi\right)^{ng-n\bar{y}} \left\{ \pi^{a-1}\left(1-\pi\right)^{b-1} \right\}$$

$$P\left(\pi|\, y\right) \propto \left(1-\theta_F\right)^{ng} \left\{ \prod_{i=1}^{n} \binom{g}{y_i} \right\} \sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} \pi^{a+i-1} \left(1-\pi\right)^{b+ng-n\bar{y}-1}$$

Now we find the normalizing constant say k. As we know that for posterior distribution we must have

$$k\left(1-\theta_F\right)^{ng}\left\{\prod_{i=1}^{n}\binom{g}{y_i}\right\}\sum_{i=0}^{n\bar{y}}\binom{n\bar{y}}{i}d^{n\bar{y}-i}\int_0^1\pi^{a+i-1}\left(1-\pi\right)^{b+ng-n\bar{y}-1}d\pi=1$$

This gives

$$k=\left[\left(1-\theta_F\right)^{ng}\left\{\prod_{i=1}^{n}\binom{g}{y_i}\right\}\sum_{i=0}^{n\bar{y}}\binom{n\bar{y}}{i}d^{n\bar{y}-i}B\left(a+i,b+ng-n\bar{y}\right)\right]^{-1}$$

Thus, the posterior distribution of $\pi$is given by:

$$P\left(\pi|y\right)=\frac{\sum_{i=0}^{n\bar{y}}\binom{n\bar{y}}{i}d^{n\bar{y}-i}\pi^{a+i-1}\left(1-\pi\right)^{b+ng-n\bar{y}-1}}{\sum_{i=0}^{n\bar{y}}\binom{n\bar{y}}{i}d^{n\bar{y}-i}B\left(a+i,b+ng-n\bar{y}\right)} \tag{9}$$

Now the Bayesian estimator (posterior mean) is given by:

$$E\left(\pi|y\right)=\frac{\sum_{i=0}^{n\bar{y}}\binom{n\bar{y}}{i}d^{n\bar{y}-i}\int_0^1\pi^{a+i+1-1}\left(1-\pi\right)^{b+ng-n\bar{y}-1}d\pi}{\sum_{i=0}^{n\bar{y}}\binom{n\bar{y}}{i}d^{n\bar{y}-i}B\left(a+i,b+ng-n\bar{y}\right)}$$

$$E\left(\pi|\,y\right)=\frac{\sum_{i=0}^{n\bar{y}}\binom{n\bar{y}}{i}d^{n\bar{y}-i}B\left(a+i+1,b+ng-n\bar{y}\right)}{\sum_{i=0}^{n\bar{y}}\binom{n\bar{y}}{i}d^{n\bar{y}-i}B\left(a+i,b+ng-n\bar{y}\right)} \tag{10}$$

While, the posterior variance is given as:

$$Var\left(\pi|\,y\right)=\frac{\sum_{i=0}^{n\bar{y}}\binom{n\bar{y}}{i}d^{n\bar{y}-i}B\left(a+i+2,b+ng-n\bar{y}\right)}{\sum_{i=0}^{n\bar{y}}\binom{n\bar{y}}{i}d^{n\bar{y}-i}B\left(a+i,b+ng-n\bar{y}\right)}$$

$$-\left(\frac{\sum_{i=0}^{n\bar{y}}\binom{n\bar{y}}{i}d^{n\bar{y}-i}B\left(a+i+1,b+ng-n\bar{y}\right)}{\sum_{i=0}^{n\bar{y}}\binom{n\bar{y}}{i}d^{n\bar{y}-i}B\left(a+i,b+ng-n\bar{y}\right)}\right)^2 \tag{11}$$

3.2. Non-informative Uniform Prior

The non-informative uniform prior distribution is given as:

$$P\left(\pi\right)\propto 1. \tag{12}$$

Using (12) and (7) in (5), the posterior distribution is derived as:

$$P(\pi|y) = \frac{\sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} \pi^{i+1-1} (1-\pi)^{ng-n\bar{y}+1-1}}{\sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B(i+1, ng-n\bar{y}+1)}. \tag{13}$$

Under the non-informative prior, the posterior mean and variance are given by:

$$E(\pi|y) = \frac{\sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B(i+2, ng-n\bar{y}+1)}{\sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B(i+1, ng-n\bar{y}+1)} \tag{14}$$

$$Var(\pi|y) = \frac{\sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B(i+3, ng-n\bar{y}+1)}{\sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B(i+1, ng-n\bar{y}+1)}$$

$$- \left(\frac{\sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B(i+2, ng-n\bar{y}+1)}{\sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} Beta(i+1, ng-n\bar{y}+1)} \right)^2 \tag{15}$$

3.3. Non-informative Haldane Prior

Another non-informative prior used here is the Haldane prior (Zellner 1996) which has the probability distribution defines as:

$$P(\pi) \propto \frac{1}{p(1-p)} \tag{16}$$

It is also defined as $B(0,0)$. Thus the posterior distribution is give as:

$$P(\pi|y) = \frac{\sum_{i=1}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} \pi^{i-1} (1-\pi)^{ng-n\bar{y}-1}}{\sum_{i=1}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B(i, ng-n\bar{y})} \tag{17}$$

Posterior mean and variance are, now, given as:

$$E(\pi|y) = \frac{\sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B(i+1, ng-n\bar{y})}{\sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B(i, ng-n\bar{y})} \tag{18}$$

$$Var\left(\pi|\, y\right) = \frac{\sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} Beta\left(i+2, ng-n\bar{y}\right)}{\sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} Beta\left(i, ng-n\bar{y}\right)}$$

$$- \left(\frac{\sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B\left(i+1, ng-n\bar{y}\right)}{\sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B\left(i, ng-n\bar{y}\right)} \right)^2 \tag{19}$$

3.4. Mixture of Beta Priors

We assume that prior information come as a mixture of different Beta distributions. The mixture of Beta distributions with H components is given as:

$$P(\pi) = \sum_{h=1}^{H} \frac{W_h}{B(a_{h,} b_h)} \pi^{a_h - 1} (1-\pi)^{b_h - 1} \tag{20}$$

where W_h are the weights such that $\sum_{h=1}^{H} W_h = 1$, anda_h b_h are the hyper-parameters of h^{th}component Beta distribution.

The posterior distribution, now, is given by:

$$P\left(\pi|\, y\right) = \frac{\sum_{h=1}^{H} \frac{W_h}{B(a_h, b_h)} \sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} \pi^{a_h+i-1} \left(1-\pi\right)^{b_h+ng-n\bar{y}-1}}{\sum_{h=1}^{H} \frac{W_h}{B(a_h, b_h)} \sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B(a_h+i, b_h+ng-n\bar{y})} \tag{21}$$

Posterior mean and variance, under the assumption of a mixture of Beta distributions, are given as:

$$\left(\pi|\, y\right) = \frac{\sum_{h=1}^{H} \frac{W_h}{B(a_h, b_h)} \sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B\left(a_h+i+1, b_h+ng-n\bar{y}\right)}{\sum_{h=1}^{H} \frac{W_h}{B(a_h, b_h)} \sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B\left(a_h+i, b_h+ng-n\bar{y}\right)} \tag{22}$$

$$Var\left(\pi|\, y\right) = \frac{\sum_{h=1}^{H} \frac{W_h}{B(a_h, b_h)} \sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B\left(a_h+i+2, b_h+ng-n\bar{y}\right)}{\sum_{h=1}^{H} \frac{W_h}{B(a_h, b_h)} \sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B\left(a_h+i, b_h+ng-n\bar{y}\right)}$$

$$- \left(\frac{\sum_{h=1}^{H} \frac{W_h}{B(a_h, b_h)} \sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B\left(a_h+i+1, b_h+ng-n\bar{y}\right)}{\sum_{h=1}^{H} \frac{w_h}{B(a_h, b_h)} \sum_{i=0}^{n\bar{y}} \binom{n\bar{y}}{i} d^{n\bar{y}-i} B\left(a_h+i, b_h+ng-n\bar{y}\right)} \right)^2 \tag{23}$$

3.5. Beta Prior with Elicited Hyperparameters

There are many methods for eliciting parameters of prior distributions. The method we have used for eliciting the hyperparameters is the method of prior predictive distribution (Aslam 2003, cf.). We first derived the prior predictive distribution and then by using "SAS" we elicited the hyperparameters. Then we have derived Posterior mean and Posterior variance.

The prior predictive distribution is given as:

$$P(y) = \frac{\binom{g}{y}(1-\theta_F)^g \sum_{i=0}^{y} \binom{y}{i} B(a+i, b+g-y)}{B(a,b)} \tag{24}$$

We solved this equation further for different values of g and y and then by using "SAS" we elicited the hyperparameters a and b. For every g we have different values of a and b. Although according to our calculations, for different values of g and y, we got same value for a, but b changed accordingly. The derived expressions for posterior distribution, posterior mean, and posterior variance are same as we have derived for posterior distribution using Beta prior with known hyperparameters, but the numerical values obtained for hyperparameters are now different.

4. Comparative Analysis

In this section, we provide a comparative analysis of posterior means and posterior variances obtained through different prior distributions assumed in this study. We should mention that under the squared error loss function posterior mean is taken as Bayesian estimator while posterior variance is taken as the measure of precision. Also, under Uniform and Haldane prior distributions, posterior distributions are not defined for $ng = n\bar{y}$. If $ng < n\bar{y}$, posterior distributions under all the priors considered here are not defined. That is why, some entries in the Tables 3 and 4 are not given. For different values of sum of responses, $n\bar{y}$, number of items g and sample size n, we have computed posterior means and variances under different prior distributions and results are displayed in Tables 1-12 given below.

We compare ML estimator and proposed Bayesian estimators in terms of variablity. To compare proposed Bayesian estimators with ML estimator, we selected $g = 7$ and $\theta_F = \frac{1}{g} \simeq 0.143$ and computed variance of ML estimator for $n = 20, 30, 40$ and 50. The variances of ML estimator for the different values of π are presented in Table 13.

From Tables 1-12 it is observed that when $n\bar{y}$, n and g are small, posterior means are larger under mixture and elicited Beta prior distributions compare to posterior means under other prior distributions considered here. For a fixed g, posterior distribution using elicited Beta prior produces larger means than the others with the increase in $n\bar{y}$. As n increases posterior means under all priors

TABLE 1: Posterior means for $n\bar{y} = 30$, $\theta_F = 0.33$ and $g = 3$.

	Prior distribution				
n	Beta	Uniform	Hadlane	Mixture	Elicited Beta
20	0.2793	0.2666	0.2187	0.3083	0.2990
30	0.1403	0.0659	0.0391	0.1453	0.1023
40	0.0849	0.0264	0.0177	0.0818	0.0480
50	0.0588	0.0154	0.0114	0.0541	0.0293

TABLE 2: Posterior means for $n\bar{y} = 50$, $\theta_F = 0.33$ and $g = 3$.

	Prior distribution				
n	Beta	Uniform	Hadlane	Mixture	Elicited Beta
20	0.6188	0.7553	0.7484	0.6525	0.7608
30	0.3299	0.3439	0.3238	0.3546	0.3608
40	0.1865	0.1395	0.0970	0.1974	0.1683
50	0.1136	0.0507	0.0283	0.1140	0.0786

TABLE 3: Posterior means for $n\bar{y} = 60$, $\theta_F = 0.33$ and $g = 3$.

	Prior distribution				
n	Beta	Uniform	Hadlane	Mixture	Elicited Beta
20	0.8074	-	-	0.8613	0.9987
30	0.4549	0.5079	0.4967	0.4800	0.5179
40	0.2687	0.2599	0.2383	0.2867	0.2771
50	0.1636	0.1144	0.0755	0.1705	0.1411

TABLE 4: Posterior means for $n\bar{y} = 90$, $\theta_F = 0.33$ and $g = 3$.

	Prior distribution				
n	Beta	Uniform	Hadlane	Mixture	Elicited Beta
20	-	-	-	-	-
30	0.8609	-	-	0.9058	0.9991
40	0.5691	0.6300	0.6243	0.5880	0.6349
50	0.3864	0.4069	0.3978	0.4035	0.4152

TABLE 5: Posterior means for $n\bar{y} = 30$, $\theta_F = 0.143$ and $g = 7$.

	Prior distribution				
n	Beta	Uniform	Hadlane	Mixture	Elicited Beta
20	0.1288	0.0905	0.0657	0.1306	0.1079
30	0.0646	0.0250	0.0141	0.0612	0.0399
40	0.0387	0.0108	0.0069	0.0349	0.0198
50	0.0266	0.0064	0.0046	0.0232	0.0123

TABLE 6: Posterior means for $n\bar{y} = 50$, $\theta_F = 0.143$ and $g = 7$.

	Prior distribution				
n	Beta	Uniform	Hadlane	Mixture	Elicited Beta
20	0.2599	0.2551	0.2460	0.2701	0.2639
30	0.1381	0.1152	0.1026	0.1401	0.1254
40	0.0797	0.0108	0.0293	0.0778	0.0602
50	0.0495	0.0189	0.0101	0.0462	0.0301

TABLE 7: Posterior means for $n\bar{y} = 30$, $\theta_F = 0.33$ and $g = 3$.

	Prior distribution				
n	Beta	Uniform	Hadlane	Mixture	Elicited Beta
20	0.5519	0.5862	0.5824	0.5637	0.5899
30	0.3319	0.3364	0.3316	0.3401	0.3411
40	0.2173	0.2110	0.2058	0.2219	0.2160
50	0.1478	0.1357	0.1298	0.1495	0.1411

TABLE 8: Posterior means for $n\bar{y} = 30$, $\theta_F = 0.33$ and $g = 3$.

	Prior distribution				
n	Beta	Uniform	Haldane	Mixture	Elicited Beta
20	0.0052	0.0086	0.0106	0.0057	0.0078
30	0.0023	0.0023	0.0014	0.0027	0.0028
40	0.0011	0.0005	0.0003	0.0012	0.0009
50	0.0006	0.0002	0.0001	0.0006	0.0004

TABLE 9: Posterior means for $n\bar{y} = 50$, $\theta_F = 0.33$ and $g = 3$.

	Prior distribution				
n	Beta	Uniform	Haldane	Mixture	Elicited Beta
20	0.0052	0.0052	0.0052	0.0050	0.0047
30	0.0043	0.0059	0.0065	0.0044	0.0056
40	0.0027	0.0039	0.0042	0.0030	0.0036
50	0.0015	0.0014	0.0008	0.0017	0.0016

TABLE 10: Posterior variances for $n\bar{y} = 30$, $\theta_F = 0.143$ and $g = 7$.

	Prior distribution				
n	Beta	Uniform	Haldane	Mixture	Elicited Beta
20	0.0013	0.0015	0.0017	0.0014	0.0015
30	0.0005	0.0004	0.0002	0.0005	0.0005
40	0.0002	0.00009	0.00005	0.0002	0.0001
50	0.0001	0.00004	0.00002	0.0001	0.00007

TABLE 11: Posterior variances for $n\bar{y} = 50$, $\theta_F = 0.143$ and $g = 7$.

	Prior distribution				
n	Beta	Uniform	Haldane	Mixture	Elicited Beta
20	0.0019	0.0021	0.0022	0.0020	0.0022
30	0.0010	0.0012	0.00129	0.0010	0.0011
40	0.0005	0.0006	0.0005	0.00057	0.0006
50	0.0003	0.0002	0.0001	0.0003	0.0003

TABLE 12: Posterior variances for $n\bar{y} = 90$, $\theta_F = 0.143$ and $g = 7$.

	Prior distribution				
n	Beta	Uniform	Haldane	Mixture	Elicited Beta
20	0.0020	0.0022	0.0022	0.0020	0.0022
30	0.0014	0.0015	0.0016	0.0014	0.0015
40	0.0010	0.0010	0.0010	0.0010	0.0010
50	0.0006	0.0007	0.0007	0.0007	0.0007

TABLE 13: Variances of ML estimator for different values of π, n, $\theta_F = \frac{1}{g}$ and $g = 7$.

	π								
n	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
20	0.005	0.008	0.011	0.012	0.013	0.012	0.010	0.008	0.004
30	0.003	0.005	0.007	0.008	0.008	0.008	0.007	0.005	0.003
40	0.002	0.004	0.005	0.006	0.006	0.006	0.005	0.004	0.002
50	0.002	0.003	0.004	0.005	0.005	0.005	0.004	0.003	0.001

decrease rapidly and posterior means using mixture and simple Beta prior distributions turn out to be larger for a relatively smaller $n\bar{y}$. The reason being their dependence upon the data and hyperparameters (see Tables 1-3 and 5-7). From Tables 1-7, it is also observed that as $n\bar{y}$ increases, posterior means under all the priors considered here become larger. The reason being they mainly depend on the magnitude of $n\bar{y}$. For a given g and larger n, if we observe the maximum $n\bar{y}$, posterior distribution using elicited Beta prior yields larger means than those provided by the other prior. We also observed that as g increases, posterior means under all prior distributions decrease. Comparatively, posterior mean using a mixture of Beta priors and Beta distributions with assumed hyperparameters have larger means than the others. However, posterior mean increases under Uniform, Heldane and mixture priors, as $n\bar{y}$ increases. for larger n, they are still smaller than posterior means using mixture and simple Beta priors. It is also observed that for increased $n\bar{y}$, posterior mean using elicited Beta prior is larger but for using large value of n it is smaller than posterior mean using simple Beta and mixture priors (see Tables 5-7).

Tables 8 and 9 show that for smaller $n\bar{y}$ and g posterior variances using Beta prior with assumed hyperparameters and mixture prior are relatively smaller than the posterior variances under other priors. For fixed values of $n\bar{y}$ and g, as n increases the posterior variance with Haldane and Uniform priors remaining smaller than that obtained under other priors.The posterior variance under Haldane and Uniform priors depend only on the $n\bar{y}$. As $n\bar{y}$, increases for given n, posterior variance under elicited Beta prior remains smaller than the posterior variance obtained under other priors. As it is largely affected by $n\bar{y}$, for larger n and for fixed $n\bar{y}$ and g, posterior variance under mixture and simple Beta priors remains smaller than the posterior variances obtained under other priors.

It is also observed that for larger g, posterior distributions using Beta prior with assumed hyperparameters and mixture prior have the smaller variances as compared to the others. But, again, for larger n, posterior distributions using Haldane, Uniform and elicited Beta prior have smaller variances than other two. But as $n\bar{y}$ is increased posterior distributions under elicited Beta, Uniform and mixture prior have smaller variances than the other two (see Tables 10-12). From expression (4), it is obvious that variance of ML estimator does not depend upon $n\bar{y}$. Thus comparison of ML estimator and proposed Bayesian estimators can be made using Tables 10-13. From Tables 10-13, it is observed that when $g = 0.7$, $n\bar{y} = 30, 60, 90$ and $\theta_F = 0.143$, posterior variances under each prior are smaller than variance of ML estimator over the whole range of π. It shows a better performance of the proposed Bayesian estimation.

5. A Real Application of Proposed Methodology

A survey was designed to collect the data from the students at Quaid-i-Azam University Islamabad. Visiting websites containing adult contents was taken as the sensitive characteristic of interest. Finding unrelated characteristics with equal known proportions among the students was observed to be difficult. Alternatively, we took three boxes containing red and white cards with equal proportion ($\theta_F = 0.33$) of red cards in each box (that is, we took $g = 3$). A simple random sample of 50 students was selected from the university. Each student was asked to randomly draw a card form each box and count 1 if he/she have ever visited a website containing adult material or if the card selected from the j^{th} $(j = 1, 2, 3)$ box is a red card. Each respondent, then, was asked to report his/her total count (which may be any value from $0 - 3$). The actual data (Y_i, $i = 1, 2, \ldots, 50$) gathered from the sample students are given in table 13 below. Thus, we have $n\overline{y} = 90$. To obtain the Bayesian estimates of proportion of students who have ever visited a website containing adult material we considered five different prior distributions: (a) simple Beta prior with hyper-parameters $a = 5, b = 10$, (b) noninformative uniform prior, (c) Haldane prior, (d) a mixture prior of 4 Beta distributions with hyperparameters; (i) $a = 1, b = 2$, (ii) $a = 2, b = 4$, (iii) $a = 3, b = 6$, (iv) $a = 4, b = 8.$, (e) Beta prior with hyperparameters $(a = 2, b = 0.0540)$ elicited from the data. Findings of the survey are summarized in Table 14.

Table 14: Actual data obtained from 50 students using $\theta_F = 0.33$ and $g = 3$

Student	1	2	3	4	5	6	7	8	9	10
Response	2	2	2	3	2	1	2	2	3	3
Student	11	12	13	14	15	16	17	18	19	20
Response	3	1	0	2	2	2	2	2	1	2
Student	21	22	23	24	25	26	27	28	29	30
Response	2	2	3	1	0	0	3	2	1	1
Student	31	32	33	34	35	36	37	38	39	40
Response	2	2	3	2	1	3	1	1	2	2
Student	41	42	43	44	45	46	47	48	49	50
Response	2	0	2	3	1	1	3	1	2	2

Table 15: Summary of the survey results

Estimates	Simple Beta	Uniform	Haldane	Mixture priors	Beta prior
Proportion	0.386	0.406	0.397	0.4035	0.4152
Variance	0.0030	0.0035	0.0036	0.0030	0.0034
95% C.I	0.278-0.492	0.293-0.523	0.284-0.523	0.284-0.507	0.292-0.522

From table 15, it is observed that the simple Beta prior with assumed known hyperparametrs and mixture prior of Beta distributions yielded relatively more precise estimators with narrower 95% confidence intervals.

6. Concluding Remarks

This study investigates a recent item count technique in a Bayesian framework using different priors in order to study which prior is more helpful in updating the item count technique. We have compared the posterior means and variances in order to check which posterior performs better than other under different conditions. In case of large values of g and n, in general, we have observed that if large sum of responses, $n\bar{y}$, are observed, posterior distribution with elicited Beta prior comes up as the most suitable choice. However the sum of response, $n\bar{y}$, is not large then posterior distribution with simple beta prior a more suitable choice. Compared to ML estimator, in terms of precision, the proposed Bayesian estimators under each prior distribution (considered in this study) perform relatively better.

[Recibido: octubre de 2012 — Aceptado: agosto de 2013]

References

Arnab, R. & Dorffner, G. (2006), 'Randomized response technique for complex survey design', *Statistical Papers* (48), 131–141.

Aslam, M. (2003), 'An application of prior predictive distribution to elicit the prior density', *Journal of Statistical Theory and Application* (2), 70–83.

Bar-Lev, S. K., Bobovitch, E. & Boukai, B. (2004), 'A note on randomized response models for quantitative data', *Metrika* (60), 255–260.

Barabesi, L. & Marcheselli, M. (2010), 'Bayesian estimation of proportion and sensitivity level in randomized response procedures', *Metrika* (72), 75–88.

Chaudhri, A. & Mukerjee, R. (1998), *Randomized Response: Theory and Methods*, Marcel-Decker, New York.

Chaudhuri, A. (2011), *Randomized Response and Indirect Questioning Techniques in Surveys*, Chapman & Hall, Florida, United States.

Chaudhuri, A. & Christofides, T. C. (2007), 'Item count technique in estimating proportion of people with sensitive feature', *Journal of Statistical Planning and Inference* (137), 589–593.

Dalton, D. R. & Metzger, M. B. (1992), 'Integrity testing for personal selection: An unsparing perspective', *Journal of Business Ethics* (12), 147–156.

Droitcour, J. A., Casper, R. A., Hubbard, M. L., Parsley, T., Visscher, W. & Ezzati, T. M. (1991), The item count technique as a method of indirect questioning: a review of its development and a case study application, *in* P. P. Biemer, R. M. Groves, L. Lyberg, N. Mathiowetz & S. Sudeman, eds, 'Measurement Errors in Surveys', Willey, New york.

Droitcour, J. A., Larson, E. M. & Scheuren, F. J. (2001), The three card method: estimating sensitive survey items with permanent anonymity of response, *in* 'Proceedings of the Social Statistics Section', American Statistical Association, Alexandria, Virginia.

Guerts, M. D. (1980), 'Using a randomized response design to eliminate nonresponse and response biases in business research', *Journal of the Academy of Marketing Science* (8), 83–91.

Gupta, S., Gupta, B. & Singh, S. (2002), 'Estimation of sensitivity level of personal interview survey questions', *Journal of Statistical Planning and Inference* **100**, 239–247.

Huang, K. C. (2010), 'Unbiased estimators of mean, variance and sensitivity level for quantitative characteristics in finite population sampling', *Metrika* (71), 341–352.

Hubbard, M. L., Casper, R. A. & Lesser, J. T. (1989), Respondent's reactions to item count list and randomized response, *in* 'Proceeding of the Survey Research Section of the American Statistical Association', Washington, D. C., pp. 544–448.

Hussain, Z. & Shabbir, J. (2010), 'Three stage quantitative randomized response model', *Journal of Probability and Statistical Sciences* (8), 223–235.

Hussain, Z., Shah, E. A. & Shabbir, J. (2012), 'An alternative item count technique in sensitive surveys', *Revista Colombiana de Estadistica* (35), 39–54.

Larkins, E. R., Hume, E. C. & Garcha, B. (1997), 'The validity of randomized response method in tax ethics research', *Journal of the Applied Business Research* **13**(3), 25–32.

Liu, P. T. & chow. L. P. (1976), 'A new discrete quantitative randomized response model', *Journal of the American Statistical Association* (71), 72–73.

Reinmuth, J. E. & Guerts, M. D. (1975), 'The collection of sensitive information using a two stage randomized response model', *Journal of Marketing Research* (12), 402–407.

Ryu, J. B., Kim, J. M., Heo, T. Y. & Park, C. G. (2005-2006), 'On stratified randomized response sampling', *Model Assisted Statistics and Applications* (1), 31–36.

Tracy, D. & Mangat, N. (1996), 'Some development in randomized response sampling during the last decade-A follow up of review by Chaudhuri and Mukerjee', *Journal of Applied Statistical Science* (4), 533–544.

Warner, S. L. (1965), 'Randomized response: A survey for eliminating evasive answer bias', *Journal of the American Statistical Association* (60), 63–69.

Zellner, A. (1996), *An Introduction to Bayesian Inference in Econometrics*, Chichester, John Wiley, New York.

Revista Colombiana de Estadística
Diciembre 2013, volumen 36, no. 2, pp. 321 a 338

Bayesian Inference for Two-Parameter Gamma Distribution Assuming Different Noninformative Priors

Inferencia Bayesiana para la distribución Gamma de dos parámetros asumiendo diferentes a prioris no informativas

Fernando Antonio Moala[1,a], Pedro Luiz Ramos[1,b], Jorge Alberto Achcar[2,c]

[1]Departamento de Estadística, Facultad de Ciencia y Tecnología, Universidade Estadual Paulista, Presidente Prudente, Brasil

[2]Departamento de Medicina Social, Facultad de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brasil

Abstract

In this paper distinct prior distributions are derived in a Bayesian inference of the two-parameters Gamma distribution. Noninformative priors, such as Jeffreys, reference, MDIP, Tibshirani and an innovative prior based on the copula approach are investigated. We show that the maximal data information prior provides in an improper posterior density and that the different choices of the parameter of interest lead to different reference priors in this case. Based on the simulated data sets, the Bayesian estimates and credible intervals for the unknown parameters are computed and the performance of the prior distributions are evaluated. The Bayesian analysis is conducted using the Markov Chain Monte Carlo (MCMC) methods to generate samples from the posterior distributions under the above priors.

***Key words*:** Gamma distribution, noninformative prior, copula, conjugate, Jeffreys prior, reference, MDIP, orthogonal, MCMC.

Resumen

En este artículo diferentes distribuciones a priori son derivadas en una inferencia Bayesiana de la distribución Gamma de dos parámetros. A prioris no informativas tales como las de Jeffrey, de referencia, MDIP, Tibshirani y una priori innovativa basada en la alternativa por cópulas son investigadas. Se muestra que una a priori de información de datos maximales conlleva a una a

[a]Professor. E-mail: femoala@fct.unesp.br
[b]Student. E-mail: pedrolramos@hotmail.com
[c]Professor. E-mail: achcar@fmrp.usp.br

posteriori impropia y que las diferentes escogencias del parámetro de interés permiten diferentes a prioris de referencia en este caso. Datos simulados permiten calcular las estimaciones Bayesianas e intervalos de credibilidad para los parámetros desconocidos así como la evaluación del desempeño de las distribuciones a priori evaluadas. El análisis Bayesiano se desarrolla usando métodos MCMC (Markov Chain Monte Carlo) para generar las muestras de la distribución a posteriori bajo las a priori consideradas.

***Palabras clave*:** a prioris de Jeffrey, a prioris no informativas, conjugada, cópulas, distribución Gamma, MCMC, MDIP, ortogonal, referencia.

1. Introduction

The Gamma distribution is widely used in reliability analysis and life testing (see for example, Lawless 1982) and it is a good alternative to the popular Weibull distribution. It is a flexible distribution that commonly offers a good fit to any variable such as in environmental, meteorology, climatology, and other physical situations.

Let X be representing the lifetime of a component with a Gamma distribution, denoted by $\Gamma(\alpha,\, \beta)$ and given by

$$f(x \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\}, \text{ for all } x > 0 \tag{1}$$

where $\alpha > 0$ and $\beta > 0$ are unknown shape and scale parameters, respectively.

There are many papers considering Bayesian inference for the estimation of the Gamma parameters. Son & Oh (2006) assume vague priors for the parameters to the estimation of parameters using Gibbs sampling. Apolloni & Bassis (2099) compute the joint probability distribution of the parameters without assuming any prior. They propose a numerical algorithm based on an approximate analytical expression of the probability distribution. Pradhan & Kundu (2011) assume that the scale parameter has a Gamma prior and the shape parameter has any log-concave prior and they are independently distributed. However, most of these papers have in common the use of proper priors and the assumption of independence a priori of the parameters. Although this is not a problem and have been much used in the literature we, would like to propose a noninformative prior for the Gamma parameters which incorporates the dependence structure of parameters. Some of priors proposed in the literature are Jeffrey*s* (1967), MDIP (Zellner 1977, Zellner 1984, Zellner 1990, Tibshirani 1989), and reference prior (Bernardo 1979). Moala (2010) provides a comparison of these priors to estimate the Weibull parameters.

Therefore, the main aim of this paper is to present different noninformative priors for a Bayesian estimation of the two-parameter Gamma distribution. We also propose a bivariate prior distribution derived from copula functions (see for example, Nelsen 1999, Trivedi & Zimmer 2005*a*, Trivedi & Zimmer 2005*b*) in order to construct a prior distribution to capture the dependence structure between the parameters α and β.

We investigate the performance of the prior distributions through a simulation study using a small data set. Accurate inference for the parameters of the Gamma is obtained using MCMC (Markov Chain Monte Carlo) methods.

2. Maximum Likelihood Estimation

Let $X_1, \ldots, X_n$ be a complete sample from (1) then the likelihood function is

$$L(\alpha, \beta \mid \boldsymbol{x}) = \frac{\beta^{n\alpha}}{[\Gamma(\alpha)]^n}\Big(\prod_{i=1}^{n} x_i^{\alpha-1}\Big)\exp\Big\{-\beta\sum_{i=1}^{n} x_i\Big\} \tag{2}$$

for $\alpha > 0$ and $\beta > 0$.

Considering $\frac{\partial}{\partial\alpha}\log L$ and $\frac{\partial}{\partial\beta}\log L$ equal to 0 and after some algebric manipulations we get the likelihood equations given by

$$\widehat{\beta} = \frac{\widehat{\alpha}}{\overline{X}} \quad \text{and} \quad \log\widehat{\alpha} - \psi(\widehat{\alpha}) = log\left(\frac{\overline{X}}{\widetilde{X}}\right) \tag{3}$$

where $\psi(k) = \frac{\partial}{\partial k} log\,\Gamma(k) = \frac{\Gamma'(k)}{\Gamma(k)}$ (see Lawless 1982) is the diGamma function, $\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ and $X = \Big(\prod_{i=1}^{n} x_i\Big)^{1/n}$. The solutions for these equations provide the maximum likelihood estimators $\widehat{\alpha}$ and $\widehat{\beta}$ for the parameters of the Gamma distribution (1). As closed form solution is not possible to evaluate (3), numerical techniques must used. The Fisher information matrix is given by

$$I(\alpha,\, \beta) = \begin{bmatrix} \psi'(\alpha) & -\frac{1}{\beta} \\ -\frac{1}{\beta} & \frac{\alpha}{\beta^2} \end{bmatrix} \tag{4}$$

where $\psi'(\alpha)$ is the derivative of $\psi(\alpha)$ called as triGamma function.

For large samples, approximated confidence intervals can be constructed for the parameters α and β through normal marginal distributions given by

$$\widehat{\alpha} \sim N(\alpha,\, \sigma_1^2) \text{ and } \widehat{\beta} \sim N(0,\, \sigma_2^2), \text{ for } n \to \infty \tag{5}$$

where $\sigma_1^2 = \widehat{var}(\widehat{\alpha}) = \frac{\widehat{\alpha}}{\widehat{\alpha}\psi\,'(\widehat{\alpha})-1}$ and $\sigma_2^2 = \widehat{var}(\widehat{\beta}) = \frac{\widehat{\beta}^2\psi\,'(\widehat{\alpha})}{\widehat{\alpha}\psi\,'(\widehat{\alpha})-1}$. In this case, the approximated $100(1-\Gamma)\%$ confidence intervals for each parameter α and β are given by

$$\widehat{\alpha} - z_{\frac{\Gamma}{2}}\sigma_1 < \alpha < \widehat{\alpha} + z_{\frac{\Gamma}{2}}\sigma_1 \text{ and } \widehat{\beta} - z_{\frac{\Gamma}{2}}\sigma_2 < \beta < \widehat{\beta} + z_{\frac{\Gamma}{2}}\sigma_2 \tag{6}$$

respectively.

3. Jeffrey's Prior

A well-known weak prior to represent a situation with little information about the parameters was proposed by Jeffreys (1967). This prior denoted by $\pi_J(\alpha, \beta)$ is derived from the Fisher information matrix $I(\alpha, \lambda)$ given in (4) as

$$\pi_J(\alpha, \beta) \propto \sqrt{det\, I(\alpha, \beta)} \tag{7}$$

Jeffrey's prior is widely used due to its invariance property under one-to-one transformations of parameters although there has been an ongoing discussion about whether the multivariate form prior is appropriate.

Thus, from (4) and (7) the Jeffreys prior for (α, β) parameters is given by:

$$\pi_J(\alpha, \beta) \propto \frac{\sqrt{\alpha\psi'(\alpha) - 1}}{\beta} \tag{8}$$

4. Maximal Data Information Prior (MDIP)

It is of interest that the data gives more information about the parameter than the information on the prior density; otherwise, there would not be justification for the realization of the experiment. Thus, we wish a prior distribution $\pi(\phi)$ that provides a gain in the information supplied by data in which the largest possible relative to the prior information of the parameter, that is, which maximize the information on the data. With this idea Zellner (1977), Zellner (1984), Zellner (1990) and Min & Zellner (1993) derived a prior which maximize the average information in the data density relative to that one in the prior. Let

$$H(\phi) = \int_{R_x} f(x \mid \phi) ln f(x \mid \phi) dx, x \in R_x \tag{9}$$

be the negative entropy of $f(x \mid \phi)$, the measure of the information in $f(x \mid \phi)$ and R_x the range of density $f(x \mid \phi)$. Thus, the following functional criterion is employed in the MDIP approach:

$$G[\pi(\phi)] = \int_{\mathrm{a}}^{\mathrm{b}} H(\phi)\pi(\phi)d\phi - \int_{\mathrm{a}}^{\mathrm{b}} \pi(\phi) \ln\, \pi(\phi)d\phi \tag{10}$$

which is the prior average information in the data density minus the information in the prior density. $G[\pi(\phi)]$ is maximized by selection of $\pi(\phi)$ subject to $\int_{\mathrm{a}}^{\mathrm{b}} \pi(\phi)d\phi = 1$. The solution is then a proper prior given by

$$\pi(\phi) = k \exp\Big\{H(\phi)\Big\} \quad a \leq \phi \leq b \tag{11}$$

where $k^{-1} = \int_{\mathrm{a}}^{\mathrm{b}} \exp\Big\{H(\phi)\Big\}d\phi$ is the normalizing constant.

Therefore, the MDIP is a prior that leads to an emphasis on the information in the data density or likelihood function. That is, its information is weak in comparison with data information.

Zellner (1977), Zellner (1984), Zellner (1990) shows several interesting properties of MDIP and additional conditions that can also be imposed to the approach reflection given initial information. However, the MDIP has restrictive invariance properties.

Theorem 1. *Suppose that we do not have much prior information available about α and β. Under this condition, the prior distribution MDIP, denoted by $\pi_Z(\alpha,\, \beta)$, for the parameters $(\alpha,\, \beta)$ of the Gamma density (1) is given by:*

$$\pi_Z(\alpha,\beta) \propto \frac{\beta}{\Gamma(\alpha)} \exp\Big\{(\alpha-1)\psi(\alpha) - \alpha\Big\} \tag{12}$$

Proof. Firstly, we have to evaluate the measure information $H(\alpha,\, \beta)$ for the Gamma density which is given by

$$H(\alpha,\beta) = \int_0^\infty \ln\Big(\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\}\Big) f(x \mid \alpha,\beta)\, dx \tag{13}$$

and after some algebra, the result is

$$H(\alpha,\beta) = \alpha \ln\, \beta - \ln\, \Gamma(\alpha) + (\alpha-1) \int_0^\infty \ln(x) f(x \mid \alpha,\beta)\, dx - \beta E(X) \tag{14}$$

with $E(X) = \frac{\alpha}{\beta}$. □

Since the integral functions $\int_0^\infty u^{\alpha-1} e^{-u}\, du = \Gamma(\alpha)$ and
$\int_0^\infty u^{\alpha-1} \log(u) e^{-u}\, du = \Gamma'(\alpha)$, the function (14) envolving these integrals can be expressed as

$$H(\alpha,\beta) = -\ln\, \Gamma(\alpha) + \ln\, \beta + (\alpha-1)\psi(\alpha) - \alpha \tag{15}$$

Therefore, the MDIP prior for the parameters α and β is given by

$$\pi_Z(\alpha,\beta) \propto \frac{\beta}{\Gamma(\alpha)} \exp\Big\{(\alpha-1)\psi(\alpha) - \alpha\Big\} \tag{16}$$

However, the corresponding joint posterior density is not proper, but surprisingly, the prior density given by

$$\pi_Z(\alpha,\beta) \propto \frac{\beta}{\Gamma(\alpha)} \exp\Big\{(\alpha-1)\frac{\psi(\alpha)}{\Gamma(\alpha)} - \alpha\Big\} \tag{17}$$

yields a proper posteriory density. Thus, we will use (17) as MDIP prior in the numerical illustration in Section 8.

5. Reference Prior

Another well-known class of noninformative priors is the reference prior, first described by Bernardo (1979) and further developed by Berger & Bernardo (1992).

The idea is to derive a prior $\pi(\phi)$ that maximizes the expected posterior information about the parameters provided by independent replications of an experiment relative to the information in the prior. A natural measure of the expected information about ϕ provided by data $\boldsymbol{x}$ is given by

$$I(\phi) = E_{\boldsymbol{x}}[K(p(\phi \mid \boldsymbol{x}), \pi(\phi))] \tag{18}$$

where

$$K(p(\phi \mid \boldsymbol{x}), \pi(\phi)) = \int_{\Phi} p(\phi \mid \boldsymbol{x}) \log \frac{p(\phi \mid \boldsymbol{x})}{\pi(\phi)} d\phi \tag{19}$$

is the Kullback-Leibler distance. So, the reference prior is defined as the prior $\pi(\phi)$ that maximizes the expected Kullback-Leibler distance between the posterior distribution $p(\phi \mid \boldsymbol{x})$ and the prior distribution $\pi(\phi)$, taken over the experimental data.

The prior density $\pi(\phi)$ which maximizes the functional (19) is found through calculus of variation and, the solution is not explicit. However, when the posterior $p(\phi \mid \boldsymbol{x})$ is asymptotically normal, this approach leads to Jeffreys prior for a single parameter situation. If on the other hand, we are interested in one of the parameters, being the remaining parameters nuisances, the situation is quite different, and the appropriated reference prior is not a multivariate Jeffrey prior. Bernardo (1979) argues that when nuisance parameters are present the reference prior should depend on which parameter(s) are considered to be of primary interest. The reference prior in this case is derived as follows. We will present here the two-parameters case in details. For the multiparameter case, see Berger & Bernardo (1992).

Let $\boldsymbol{\theta} = (\theta_1, \theta_2)$ be the whole parameter, θ_1 being the parameter of interest and θ_2 the nuisance parameter. The algorithm is as follows:

Step 1: Determine $\pi_2(\theta_2 \mid \theta_1)$, the conditional reference prior for θ_2 assuming that θ_1 is known, is given by,

$$\pi_2(\theta_2 \mid \theta_1) = \sqrt{\mathrm{I}_{22}(\theta_1, \theta_2)} \tag{20}$$

where $\mathrm{I}_{22}(\theta_1, \theta_2)$ is the (2,2)-entry of the Fisher Information Matrix.

Step 2: Normalize $\pi_2(\theta_2 \mid \theta_1)$.

If $\pi_2(\theta_2 \mid \theta_1)$ is improper, choose a sequence of subsets $\Omega_1 \subseteq \Omega_2 \subseteq \ldots \to \Omega$ on which $\pi_2(\theta_2 \mid \theta_1)$ is proper. Define the normalizing constant and the proper prior $p_m(\theta_2 \mid \theta_1)$ respectively as

$$c_{\mathrm{m}}(\theta_1) = \frac{1}{\int_{\Omega_m} \pi_2(\theta_2 \mid \theta_1) d\theta_2} \tag{21}$$

and

$$p_m(\theta_2 \mid \theta_1) = c_m(\theta_1) \pi_2(\theta_2 \mid \theta_1) 1_{\Omega_m}(\theta_2), \tag{22}$$

Step 3: Find the marginal reference prior $\pi_m(\theta_1)$ for $\theta_1$the reference prior for the experiment found by marginalizing out with respect to $p_m(\theta_2 \mid \theta_1)$. We obtain

$$\pi_m(\theta_1) \propto exp\Big\{ \frac{1}{2} \int_{\Omega_m} p_m(\theta_2 \mid \theta_1) log \left\| \frac{det\ I(\theta_1, \theta_2)}{I_{22}(\theta_1, \theta_2)} \right\| d\theta_2 \Big\} \tag{23}$$

Step 4: Compute the reference prior $\pi_{\theta_1}(\theta_1,\, \theta_2)$ when θ_1 is the parameter of interest

$$\pi_{\theta_1}(\theta_1, \theta_2) = \lim_{m\to\infty} \left(\frac{c_m(\theta_1)\pi_m(\theta_1)}{c_m(\theta_1^*)\ \pi_m(\theta_1^*)} \right) \pi(\theta_2 \mid \theta_1) \tag{24}$$

where θ_1^* is any fixed point with positive density for all π_m.

We will derive the reference prior for the parameters of the Gamma distribution given in (1), where α will be considered as the parameter of interest and β the nuisance parameter.

Theorem 2. *The reference prior for the parameters of the Gamma distribution given in (1), where α will be considered as the parameter of interest and β the nuisance parameter, is given by:*

$$\pi_\alpha(\alpha, \beta) = \frac{1}{\beta}\sqrt{\frac{\alpha\psi'(\alpha) - 1}{\alpha}} \tag{25}$$

If β is the parameter of interest and α the nuisance, thus the prior is

$$\pi_\beta(\alpha, \beta) \propto \frac{\sqrt{\psi'(\alpha)}}{\beta} \tag{26}$$

Proof. By the approach proposed by Berger & Bernardo (1992), we find the reference prior for the nuisance parameter β, conditionally on the parameter of interest α, given by

$$\pi(\beta \mid \alpha) = \sqrt{\mathrm{I}_{\beta\beta}(\alpha,\, \beta)} \propto \frac{1}{\beta} \tag{27}$$

where $\mathrm{I}_{\beta\beta}(\alpha, \beta)$ is the (2,2)-entry of the Fisher Information Matrix given in (4). □

As in Moala (2010), a natural sequence of compact sets for $(\alpha,\, \beta)$ is $(l_{1n}, l_{2n})\times (q_{1n}, q_{2n})$, so that $l_{1i},\ q_{1i} \to 0$ and $l_{2i},\ q_{2i} \to \infty$ when $i \to \infty$. Therefore, the normalizing constant is given by,

$$c_i(\alpha) = \frac{1}{\int_{q_{1i}}^{q_{2i}} \frac{1}{\beta} d\beta} = \frac{1}{\log\ q_{2i} - \log\ q_{1i}}. \tag{28}$$

Now from (23), the marginal reference prior for α is given by

$$\pi_i(\alpha) = \exp\Big\{ \frac{1}{2} \int_{q_{1i}}^{q_{2i}} c_i(\alpha) \frac{1}{\beta} \log \left\| \frac{\frac{\alpha\psi'(\alpha)-1}{\beta^2}}{\frac{\alpha}{\beta^2}} \right\| d\beta \Big\} \tag{29}$$

which after some mathematical arrangement, we have

$$\pi_i(\alpha) = \sqrt{\frac{\alpha\psi'(\alpha) - 1}{\alpha}} \exp\Big\{ \frac{1}{2} c_i(\alpha) \int_{q_{1i}}^{q_{2i}} \frac{1}{\beta} d\beta \Big\} \tag{30}$$

Therefore, the resulting marginal reference prior for α is given by

$$\pi_i(\alpha) \propto \sqrt{\frac{\alpha\psi'(\alpha) - 1}{\alpha}} \tag{31}$$

and the global reference prior for $(\alpha,\ \beta)$ with parameter of interest α is given by,

$$\pi_{\alpha}(\alpha,\beta)=\lim_{i\to\infty}\left(\frac{c_i(\alpha)\pi_i(\alpha)}{c_i(\alpha^*)\ \pi_i(\alpha^*)}\right)\ \pi(\beta\,|\,\alpha)\propto\frac{1}{\beta}\sqrt{\frac{\alpha\psi'(\alpha)-1}{\alpha}} \tag{32}$$

considering $\alpha^*=1$

Similarly we obtain the reference prior considering β as the parameter of interest and α as nuisance. In this case, the prior is

$$\pi_{\beta}(\alpha,\beta)\propto\frac{\sqrt{\psi'(\alpha)}}{\beta} \tag{33}$$

6. Tibishirani's Prior

Given a vector parameter $\boldsymbol{\phi}$, Tibshirani (1989) developed an alternative method to derive a noninformative prior $\pi(\delta)$ for the parameter of interest $\delta=t(\boldsymbol{\phi})$ so that the credible interval for δ has coverage error $O(n^{-1})$ in the frequentist sense. This means that the difference between the posterior and frequentist confidence interval should be small. To achieve that, Tibshirani (1989) proposed to reparametrize the model in terms of the orthogonal parameters $(\delta,\ \lambda)$ (see Cox & Reid 1987) where δ is the parameter of interest and λ is the orthogonal nuisance parameter. In this way, the approach specifies the weak prior to be any prior of the form

$$\pi(\delta,\lambda)=g(\lambda)\sqrt{I_{\delta\delta}(\delta,\ \lambda)} \tag{34}$$

where $g(\lambda)>0$ is an arbitrary function and $I_{\delta\delta}(\delta,\ \lambda)$ is the "delta" entry of the Fisher Information Matrix.

Theorem 3. *The Tibishirani's prior distribution $\pi_T(\alpha,\ \beta)$ for the parameters $(\alpha,\ \beta)$ of the Gamma distribution given in (1) by considering α as the parameter of interest and β the nuisance parameter is given by:*

$$\pi_T(\alpha,\beta)\propto\frac{1}{\beta}\sqrt{\frac{\alpha\psi'(\alpha)-1}{\alpha}} \tag{35}$$

Proof. For the Gamma model (1), we will propose an orthogonal reparametrization $(\delta,\ \lambda)$ where $\delta=\alpha$ is the parameter of interest and λ is the nuisance parameter to be evaluated. The orthogonal parameter λ is obtained by solving the differential equation:

$$I_{\beta\beta}\frac{\partial\beta}{\partial\alpha}=-I_{\alpha\beta} \tag{36}$$

□

From (4) and (36) we have

$$\frac{\alpha}{\beta^2}\frac{\partial\beta}{\partial\alpha}=\frac{1}{\beta} \tag{37}$$

Separating the variables, (37) becomes the following,

$$\frac{1}{\beta}\partial\beta = \frac{1}{\alpha}\partial\alpha \tag{38}$$

Integrating both sides we get,

$$\log\,\beta = log\,\alpha + h(\lambda) \tag{39}$$

where $h(\lambda)$ is an arbitrary function of λ.

By chosing $h(\lambda) = \log\lambda$, we obtained the solution to (36), the nuisance parameter λ orthogonal to δ,

$$\lambda = \frac{\beta}{\alpha} \tag{40}$$

Thus, the information matrix for the orthogonal parameters is given by

$$I(\boldsymbol{\delta}, \boldsymbol{\lambda}) = \begin{bmatrix} \psi'(\delta) - \frac{1}{\delta} & 0 \\ 0 & \frac{\delta}{\lambda^2} \end{bmatrix} \tag{41}$$

From (34) and (41), the corresponding prior for (δ, λ) is given by

$$\pi_{\boldsymbol{\delta}}(\delta, \lambda) \propto g(\lambda)\sqrt{\frac{\delta\psi'(\delta) - 1}{\delta}} \tag{42}$$

where $g(\lambda)$ is an arbitrary function.

Due to a lack of uniqueness in the choice of the orthogonal parametrization, then the class of orthogonal parameters is of the form $g(\lambda)$, where $g(\cdot)$ is any reparametrization. This non-uniqueness is reflected by the function $g(\cdot)$ corresponding to (26). One possibility, in the single nuisance parameter case, is to require that (δ, λ) also satisfies Stein's condition (see Tibshirani 1989) for λ with p taken as the nuisance parameter. Under this condition we obtain

$$\pi_{\lambda}(\delta, \lambda) \propto g^*(\delta)\frac{\sqrt{\delta}}{\lambda} \tag{43}$$

Now, requiring $g(\lambda)\sqrt{\frac{\delta\psi'(\delta)-1}{\delta}} = g^*(\delta)\frac{\sqrt{\delta}}{\lambda}$ we have that

$$\pi_T(\delta, \lambda) \propto \frac{1}{\lambda}\sqrt{\frac{\delta\psi'(\delta) - 1}{\delta}} \tag{44}$$

Thus, from (40), the prior expressed in terms of the (α, β) parametrization is given by

$$\pi_T(\alpha, \beta) \propto \frac{1}{\beta}\sqrt{\frac{\alpha\psi'(\alpha) - 1}{\alpha}} \tag{45}$$

Note that this prior coincides with reference prior (25) considering α as the parameter of interest.

7. Copula Prior

In this section we derive a bivariate prior distribution from copula functions (see for example, Nelsen 1999, Trivedi & Zimmer 2005*a*, Trivedi & Zimmer 2005*b*) in order to construct a prior distribution to capture the dependence structure between the parameters α and β. Copulas can be used to correlate two or more random variables and they provide great flexibility to fit known marginal densities.

A special case is given by the Farlie-Gumbel-Morgenstern copula which is suitable to model weak dependences (see Morgenstern 1956) with corresponding bivariate prior distribution for α and β given ρ,

$$\pi(\alpha, \beta \mid \rho) = f_1(\alpha) f_2(\beta) + \rho f_1(\alpha) f_2(\beta)[1 - 2F_1(\alpha)][1 - 2F_2(\beta)], \qquad (46)$$

where $f_1(\alpha)$ and $f_2(\beta)$ are the marginal densities for the random quantities α and β; $F_1(\alpha)$ and $F_2(\beta)$ are the corresponding marginal distribution functions for α and β, and $-1 \leq \rho \leq 1$.

Observe that if $\rho = 0$, we have independence between α and β.

Different choices could be considered as marginal distributions for α and β as Gamma, exponential, Weibull or uniform distributions.

In this paper, we will assume Gamma marginal distribution $\Gamma(\mathrm{a}_1, b_1)$ and $\Gamma(\mathrm{a}_2, b_2)$ for α and β, respectively, with known hyperparameters a_1 , a_2, b_1 and b_2. Thus,

$$\begin{aligned} \pi(\alpha, \beta \mid a_1\,, a_2, b_1, b_2, \rho) \propto \alpha^{\mathrm{a}_1 - 1} \beta^{\mathrm{a}_2 - 1} \exp\Big\{-b_1\alpha - b_2\beta\Big\} \times \\ \Big[1 + \rho\Big(1 - 2I(a_1, b_1\alpha)\Big)\Big(1 - 2I(a_1, b_1\alpha)\Big)\Big] \end{aligned} \qquad (47)$$

where $I(k, x) = \frac{1}{\Gamma(\mathrm{k})} \int_0^x u^{k-1} e^{-u} du$ is the incomplete Gamma function.

Assuming the prior (47), the joint posterior distribution for α, β and ρ is given by,

$$\begin{aligned} p(\alpha, \beta, \rho \mid x) \propto \frac{\beta^{n\alpha}}{[\Gamma(\alpha)]^n} \Big(\prod_{i=1}^{n} x_i^{\alpha - 1}\Big) \\ \exp\Big\{-\beta \sum_{i=1}^{n} x_i\Big\} \pi(\alpha, \beta \mid a_1\,, a_2, b_1, b_2, \rho) \pi(\rho) \end{aligned} \qquad (48)$$

where $\pi(\rho)$ is a prior distribution for ρ.

In general, many different priors can be used for ρ; one possibility is to consider an uniform prior distribution for ρ over the interval $[-1, 1]$.

8. Numerical Illustration

8.1. Simulation Study

In this section, we investigate the performance of the proposed prior distributions through a simulation study with samples of size $n = 5$, $n = 10$ and $n = 30$ generated from the Gamma distribution with parameters $\alpha = 2$ and $\beta = 3$.

As we do not have an analytic form for marginal posterior distributions we need to appeal to the MCMC algorithm to obtain the marginal posterior distributions and hence to extract characteristics of parameters such as Bayes estimators and credible intervals. The chain is run for 10,000 iterations with a burn-in period of 1,000. Details of the implementation of the MCMC algorithm used in this paper are given below.

i) choose starting values α_0 and β_0.

ii) at step $i + 1$, we draw a new value α_{i+1} conditional on the current α_i from the Gamma distribution $\Gamma(\alpha_i/\mathrm{c}, \mathrm{c})$;

iii) the candidate α_{i+1} will be accepted with a probability given by the Metropolis ratio

$$u(\alpha_i\,,\,\alpha_{i+1}) = \min\left\{1\,,\,\frac{\Gamma(\alpha_i/c,c)p\big(\alpha_{i+1},\,\beta_i\mid\,\boldsymbol{x}\big)}{\Gamma(\alpha_{i+1}/c,c)p\big(\alpha_i,\beta_i\mid\,\boldsymbol{x}\big)}\right\}$$

iv) sample the new value $\boldsymbol{\beta}_{i+1}$ from the Gamma distribution $\Gamma(\beta_i/d, d)$;

v) the candidate β_{i+1} will be accepted with a probability given by the Metropolis ratio

$$u(\beta_i\,,\,\beta_{i+1}) = \min\left\{1\,,\,\frac{\Gamma(\beta_i/d,d)p\big(\alpha_{i+1},\,\beta_{i+1}\mid\,\boldsymbol{x}\big)}{\Gamma(\beta_{i+1}/d,d)p\big(\alpha_{i+1},\beta_i\mid\,\boldsymbol{x}\big)}\right\}$$

The proposal distribution parameters c and d were chosen to obtain a good mixing of the chains and the convergence of the MCMC samples of parameters are assessed using the criteria given by Raftery and Lewis (1992). More details of MCMC in oder to construct these chains see, for example, Smith & Roberts (1993), Gelfand & Smith (1990), Gilks, Clayton, Spiegelhalter, Best, McNiel, Sharples & Kirby (1993).

We examine the performance of the priors by computing point estimates for parameters α and β based on 1,000 simulated samples and then we averaged the estimates of the parameters, obtain the variances and the coverage probability of 95% confidence intervals. Table 1 shows the point estimates for α and its respective variances given between parenthesis. Table 2 shows the same summaries for β.

The results of our numerical studies show that there is little difference between the point estimates for both parameters α and β. However, the MDIP prior produces a much smaller variance than using the other assumed priors. The uniform prior and MLE estimate produce bad estimations with large variances showing

Table 1: Summaries for parameter α.

$\boldsymbol{\alpha = 2}$	Jeffreys	MDIP	Tibshirani	Reference	Copula	Uniform	MLE
$n = 5$	2.2529	2.1894	2.3666	2.3602	2.0909	3.3191	3.2850
	(2.1640)	(0.5112)	(3.0297)	(2.4756)	(2.1987)	(3.1577)	(7.5372)
$n = 10$	2.5227	2.2253	2.4138	2.4761	2.3068	2.9769	2.7013
	(1.3638)	(0.3855)	(1.3849)	(1.3301)	(1.2052)	(1.4658)	(1.8308)
$n = 30$	2.1259	2.1744	2.0606	2.1079	2.0369	2.2504	2.2253
	(0.2712)	(0.1910)	(0.2651)	(0.2728)	(0.2571)	(0.2829)	(0.3138)

Table 2: Summaries for parameter β.

$\boldsymbol{\beta = 3}$	Jeffreys	MDIP	Tibshirani	Reference	Copula	Uniform	MLE
$n = 5$	2.9136	3.2680	3.2161	3.1058	2.7486	4.3419	5.2673
	(4.2292)	(1.6890)	(6.1384)	(4.7787)	(4.2447)	(5.6351)	(23.1881)
$n = 10$	3.8577	3.7727	3.6705	3.7649	3.6112	4.5186	4.2960
	(3.8086)	(1.5872)	(3.8110)	(3.6667)	(3.5599)	(3.7803)	(5.4296)
$n = 30$	3.2328	3.4255	3.1475	3.1798	3.0805	3.4633	3.4005
	(0.7950)	(0.6292)	(0.7861)	(0.7856)	(0.7453)	(0.8335)	(0.9163)

that, despite being widely used in the literature, they are not suitable for the Gamma distribution. As expected, the performance of these priors improves when the sample size increases.

Frequentist property of coverage probabilities for the parameters α and β have also been studied to compare the priors and MLE. Table 3 summarizes the simulated coverage probabilities of 95% confidence intervals. For the three sample sizes considered here, the intervals of MDIP prior produce an over-coverage for small sample sizes while, the intervals of uniform prior and MLE seem to have an under-coverage for some cases. Coverage probabilities are very close to the nominal value when n increases.

Table 3: Frequentist coverage probability of the 95% confidence intervals for α and β.

$\boldsymbol{\alpha = 2}$	Jeffreys	MDIP	Tibshirani	Reference	Copula	Uniform	MLE
$n = 5$	96.30%	99.60%	97.20%	96.10%	95.30%	95.70%	95.60%
$n = 10$	96.40%	99.50%	94.90%	95.00%	95.80%	90.60%	95.30%
$n = 30$	96.10%	96.20%	98.10%	95.80%	96.80%	95.50%	95.00%
$\boldsymbol{\beta = 3}$	Jeffreys	MDIP	Tibshirani	Reference	Copula	Uniform	MLE
$n = 5$	96.60%	99.60%	97.50%	97.00%	96.30%	99.90%	94.30%
$n = 10$	98.10%	98.00%	96.00%	96.70%	97.80%	93.90%	95.70%
$n = 30$	97.30%	95.80%	96.90%	96.10%	97.40%	94.90%	96.80%

8.2. Rainfall Data Example

Data in Table 4 represent the average monthly rainfall obtained from the Information System for Management of Water Resources of the State of São Paulo, including a period of 56 years from 1947 to 2003, by considering the month of November.

Let us assume a Gamma distribution with density (1) to analyse the data.

TABLE 4: Historical rainfall averages over last 56 years in State of São Paulo.

0.2,3.5,2.8,3.7,8.7,6.9,7.4,0.8,4.8,2.5,2.9,3.1,4.0,5.0,3.8,3.5,5.4,3.3,2.9,
1.7,7.3,2.9,4.6,1.1,1.4,3.9,6.2,4.1,10.8,3.8,7.3,1.8,6.7,3.5,3.2,5.2,2.8,5.2,
5.4,2.2,9.9,2.1,4.7,5.5,2.6,4.1,5.4,5.5,2.1,1.9,8.8,1.3,24.1,5.4,6.2,2.9

Table 5 presents the posterior means assuming the different prior distributions and maximum likelihood estimates (MLE) for the parameters α and β.

TABLE 5: Posterior means for parameters α and β of rainfall data.

	Uniform	Jeffreys	Ref-$\boldsymbol{\beta}$	MDIP	Tibshirani	Copula	MLE
α	2.493	2.387	2.393	2.659	2.357	2.380	2.395
β	0.543	0.516	0.517	0.641	0.510	0.515	0.518

From Table 5, we observe similar inference results assuming the different prior distributions for α and β, except for MDIP prior as observed in the simulation study introduced in the example presented in section 8.1.

The 95% posterior credible intervals obtained using the different priors for the parameters are displayed in Table 6. The MLE intervals for the parameters α and β are given respectively by (1.56; 3.22) and (0.31; 0.72).

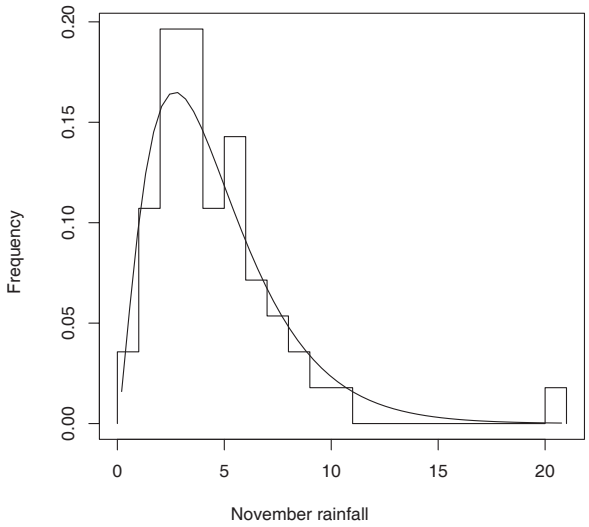


FIGURE 1: Histogram and fitted Gamma distribution for rainfall data.

TABLE 6: 95% posterior intervals for the parameters α and β of rainfall data.

	Uniform	Jeffreys	Ref-$\boldsymbol{\beta}$	MDIP	Tibshirani	Copula
α	(1.71; 3.43)	(1.63; 3.29)	(1.64; 3.28)	(1.91; 3.52)	(1.60; 3.25)	(1.60; 3.34)
β	(0.35; 0.76)	(0.33; 0.73)	(0.34; 0.73)	(0.44; 0.87)	(0.33; 0.73)	(0.32; 0.75)

Figure 2 shows the marginal posterior densities for both parameters α and β. We can see that the MDIP prior leads to a posterior slightly more sharply peaked for both parameters, while the other priors are quite similar, agreeing with simulated data with sample size $n = 30$.

To determine the appropriate prior distribution to be used with the rainfall data fitted by the Gamma distribution, some selection criteria can be examined. These include information-based criteria (AIC, BIC and DIC) given in the Table 7 for each prior distribution.

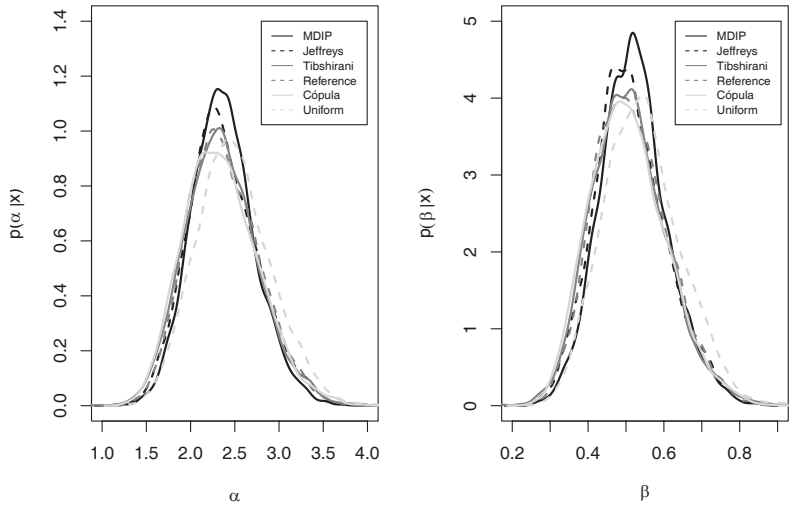


FIGURE 2: Plot of marginal posterior densities for the parameters α and β of rainfall data.

TABLE 7: Information-based model selection criteria (AIC, BIC and DIC) for rainfall data.

Prior	AIC	BIC	DIC
Jeffreys	272.213	268.162	267.827
MDIP	272.247	268.196	267.502
Ref-β	272.212	268.162	267.922
Tibshirani	272.219	268.169	268.068
Copula	272.222	268.171	268.197
Uniform	272.266	268.215	267.935

From the results of Table 7 and Figure 2 we observe that the choice of the prior distributions for parameters α and β has a negligible effect on the posterior distribution, surely due to the large amount of data in this study.

8.3. Reliability Data Example

In this example, we consider a lifetime data set related to an electrical insulator subjected to constant stress and strain introduced by Lawless (1982). The dataset does not have censored values and represent the lifetime (in minutes) to failure: 0.96, 4.15, 0.19, 0.78, 8.01, 31.75, 7.35, 6.50, 8.27, 33.91, 32.52, 16.03, 4.85, 2.78, 4.67, 1.31, 12.06, 36.71 and 72.89. Let us denote this data as "Lawless data". We assume a Gamma distribution with density (1) to analyse the data.

The maximum likelihood estimators and the Bayesian summaries for α and β, considering the different prior distributions are given in Table 8. Table 9 shows the 95% posterior intervals for α and β. The estimated marginal posterior distributions for the parameters are shown in Figure 3.

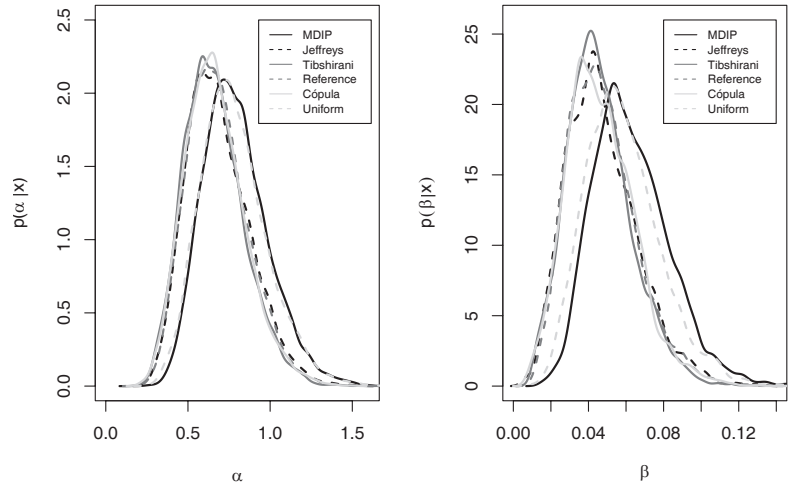


FIGURE 3: Plot of marginal posterior densities for the parameters α and β for Lawless data.

Tables 8 and 9 present the posterior statistics and 95% confidence intervals for both parameters resulting from the proposed priors. Again the performance of the MDIP prior clashes from the others.

TABLE 8: Posterior means for parameters α and β (Lawless data).

	Uniform	Jeffreys	Ref-$\boldsymbol{\beta}$	MDIP	Tibshirani	Copula	MLE
α	0.779	0.686	0.681	0.789	0.660	0.666	0.690
β	0.058	0.047	0.047	0.063	0.046	0.047	0.048

TABLE 9: 95% posterior intervals for the parameters α and β (Lawless data).

	Uniform	Jeffreys	Ref-$\boldsymbol{\beta}$	MDIP	Tibshirani	Copula
α	(0.433, 1.229)	(0.371, 1.111)	(0.374, 1.087)	(0.459, 1.232)	(0.350, 1.061)	(0.353, 1.080)
β	(0.025, 0.105)	(0.017, 0.091)	(0.018, 0.090)	(0.031, 0.108)	(0.017, 0.085)	(0.017, 0.087)

Table 10 shows the AIC, BIC and DIC values for all priors under investigation, with similar results as presented in Table 7 are obtained in this comparison which shows no differences using the different assumed priors.

Table 10: Information-based model selection criteria (AIC, BIC and DIC) for(Lawless data.

Prior	AIC	BIC	DIC
Jeffreys	143.125	141.236	141.409
MDIP	143.642	141.753	141.082
Ref-β	143.126	141.237	141.168
Tibshirani	143.148	141.259	141.247
Copula	143.138	141.249	141.471
Uniform	143.401	141.512	141.119

9. Conclusion and Discussion

The large number of noninformative priors can cause difficulties in the choosing one, especially when these priors does not produce similar results. Thus, in this paper, we presented a Bayesian analysis using a variety of prior distributions for the estimation of the parameters of the Gamma distribution.

We have shown that the use of the maximal data information process proposed by Zellner (1977), Zellner (1984), Zellner (1990) yields an improper posterior distribution for the parameters α and β. In this way, we proposed a "modified" MDIP prior analytically similar to the original one but with proper posterior. We also shown that the reference prior provides nonuniqueness of prior due to the choice of the parameter of interest, although the simulation shows the same performance. We have shown that the Tibshirani prior applied to the parameters of the Gamma distribution is equal to the reference prior when α is the parameter of interest.

Besides, a simulation study to check the impact of the use of different noninformative priors in the posterior distributions was also carried out. From this study we can conclude that it is necessary to carefully choose a prior for the parameters of the Gamma distribution when there is not enough data.

As expected, a moderated large sample size is need to achieve the desirable accuracy. In this case, the choice of the priors become irrelevant. However, the disagreement is substantial for small sample sizes.

Our simulation study indicates that the class of priors: Jeffreys, Reference, Tibshirani and Copula, had the same performance while the Uniform prior had worse performance. On the other hand , the "modified" MDIP prior produced the best estimations for α and β. Thus, the simulation study showed that the effect of the prior distributions can be substantial in the estimation of parameters and therefore the modified MDIP prior should be the recommended noninformative prior for the estimation of parameters of the Gamma distribution.

[Recibido: enero de 2013 — Aceptado: septiembre de 2013]

References

Apolloni, B. & Bassis, S. (2099), 'Algorithmic inference of two-parameter gamma distribution', *Communications in Statistics - Simulation and Computation* **38**(9), 1950–1968.

Berger, J. & Bernardo, J. M. (1992), On the development of the reference prior method, Fourth Valencia International Meeting on Bayesian Statistics, Spain.

Bernardo, J. M. (1979), 'Reference posterior distributions for Bayesian inference', *Journal of the Royal Statistical Society* **41**(2), 113–147.

Cox, D. R. & Reid, N. (1987), 'Parameter orthogonality and approximate conditional inference (with discussion)', *Journal of the Royal Statistical Society, Series B* **49**, 1–39.

Gelfand, A. E. & Smith, F. M. (1990), 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association* **85**, 398–409.

Gilks, W., Clayton, D., Spiegelhalter, D., Best, N., McNiel, A., Sharples, L. & Kirby, A. (1993), 'Modeling complexity: Applications of Gibbs sampling in medicine', *Journal of the Royal Statistical Society, Series B* **55**(1), 39–52.

Jeffreys, S. H. (1967), *Theory of Probability*, 3 edn, Oxford University Press, London.

Lawless, J. (1982), *Statistical Models and Methods for Lifetime Data*, John Wiley, New York.

Min, C.-k. & Zellner, A. (1993), Bayesian Analysis, Model Selection and Prediction, *in* 'Physics and Probability: Essays in honor of Edwin T Jaynes', Cambridge University Press, pp. 195–206.

Moala, F. (2010), 'Bayesian analysis for the Weibull parameters by using noninformative prior distributions', *Advances and Applications in Statistics* (14), 117–143.

Morgenstern, D. (1956), 'Einfache beispiele sw edimensionaler vertielung', *Mit Mathematics Statistics* **8**, 234–235.

Nelsen, R. B. (1999), *An Introduction to Copulas*, Springer Verlag, New York.

Pradhan, B. & Kundu, D. (2011), 'Bayes estimation and prediction of the two-parameter Gamma distribution', *Journal of Statistical Computation and Simulation* **81**(9), 1187–1198.

Smith, A. & Roberts, G. (1993), 'Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo Methods', *Journal of the Royal Statistical Society: Series B* **55**, 3–24.

Son, Y. & Oh, M. (2006), 'Bayesian estimation of the two-parameter Gamma distribution', *Communications in Statistics – Simulation and Computation* **35**, 285–293.

Tibshirani, R. (1989), 'Noninformative prioris for one parameters of many', *Biometrika* **76**, 604–608.

Trivedi, P. K. & Zimmer, D. M. (2005*a*), *Copula Modelling*, New Publishers, New York.

Trivedi, P. K. & Zimmer, D. M. (2005*b*), 'Copula modelling: An introduction to practicioners', *Foundations and Trends in Econometrics* .

Zellner, A. (1977), Maximal data information prior distributions, *in* A. Aykac & C. Brumat, eds, 'In New Methods in the Applications of Bayesian Methods', North-Holland, Amsterdam.

Zellner, A. (1984), *Maximal Data Information Prior Distributions, Basic Issues in Econometrics*, The University of Chicago Press, Chicago, USA.

Zellner, A. (1990), Bayesian methods and entropy in economics and econometrics, *in* W. J. Grandy & L. Schick, eds, 'Maximum Entropy and Bayesian Methods', Dordrecht, Netherlands: Kluwer Academic Publishers, pp. 17–31.

Revista Colombiana de Estadística
Diciembre 2013, volumen 36, no. 2, pp. 339 a 358

Inference for the Weibull Distribution Based on Fuzzy Data

Inferencia para la istribución Weibull basada en datos difusos

Abbas Pak[1,a], Gholam Ali Parham[1,b], Mansour Saraj[2,c]

[1]Department of Statistics, Faculty of Mathematical Sciences and Computer, Shahid Chamran University of Ahvaz, Ahvaz, Iran

[2]Department of Mathematics, Faculty of Mathematical Sciences and Computer, Shahid Chamran University of Ahvaz, Ahvaz, Iran

Resumen

Classical estimation procedures for the parameters of Weibull distribution are based on precise data. It is usually assumed that observed data are precise real numbers. However, some collected data might be imprecise and are represented in the form of fuzzy numbers. Thus, it is necessary to generalize classical statistical estimation methods for real numbers to fuzzy numbers. In this paper, different methods of estimation are discussed for the parameters of Weibull distribution when the available data are in the form of fuzzy numbers. They include the maximum likelihood estimation, Bayesian estimation and method of moments. The estimation procedures are discussed in details and compared via Monte Carlo simulations in terms of their average biases and mean squared errors. Finally, a real data set taken from a light emitting diodes manufacturing process is investigated to illustrate the applicability of the proposed methods.

***Palabras clave*:** Bayesian estimation, EM algorithm, Fuzzy data analysis, Maximum likelihood principle.

Abstract

Los procedimientos clásicos de estimación para los parámetros de la distribución Weibull se encuentran basados en datos precisos. Se asume usualmente que los datos observados son números reales precisos. Sin embargo, algunos datos recolectados podrían ser imprecisos y ser representados en la forma de números difusos. Por lo tanto, es necesario generalizar los métodos de estimación estadísticos clásicos de números reales a números difusos. En este artículo, diferentes métodos de estimación son discutidos para los

[a]PhD Student. E-mail: a-pak@scu.ac.ir
[b]Associate professor. E-mail: parham-g@scu.ac.ir
[c]Associate professor. E-mail: seraj.a@scu.ac.ir

parámetros de la distribución Weibull cuando los datos disponibles están en la forma de números difusos. Estos incluyen la estimación por máxima verosimilitud, la estimación Bayesiana y el método de momentos. Los procedimientos de estimación se discuten en detalle y se comparan vía simulaciones de Monte Carlo en términos de sesgos promedios y errores cuadráticos medios.

***Key words*:** algoritmo EM, análisis de datos difusos, estimación Bayesiana, principio de máxima verosimilitud.

1. Introduction

The Weibull distribution was originally proposed by Waloddi Weibull back in 1937 for estimating machinery lifetime. Nowadays, the Weibull distribution is a broadly used in statistical model in engineering and life-time data analysis. The probability density function (pdf) and the cumulative distribution function (cdf) of a two-parameter Weibull random variable X can be written as

$$f(x;\alpha,\lambda) = \alpha\lambda x^{\alpha-1}\exp(-\lambda x^{\alpha}), \quad x > 0 \tag{1}$$

and

$$F(x;\alpha,\lambda) = 1 - \exp(-\lambda x^{\alpha}), \quad x > 0 \tag{2}$$

respectively, where $\lambda > 0$ is the scale and $\alpha > 0$ is the shape parameter. Several authors have addressed inferential issues for the parameters of a Weibull distribution; among others, Al-Baidhani & Sinclair (1987) compared the generalized least squares, maximum likelihood, and the two mixed method of estimating the parameters of a Weibull distribution. Qiao & Tsokos (1994) introduced an effective iterative procedure for the estimation. Watkins (1994) discussed maximum likelihood estimation for the two parameter Weibull distribution when the data for analysis contains both times to failure and censored times in operation. Marks (2005) considered the estimation of Weibull distribution parameters using the symmetrically located percentiles from a sample. Helu, Abu-Salih & Alkam (2010) proposed different methods of estimation for the parameters of Weibull distribution based on different sampling schemes-namely, simple random sample, ranked set sample, and modified ranked set sample.

The above inference techniques are limited to precise data. In real world situations, the data sometimes can not be measured and recorded precisely due to machine errors, human errors or some unexpected situations. The two types of such data namely, censored data and truncated data are widely used in practice. Censored data typically arise when an event of interest, such as a disease or a failure, is only partially observed, because information is gathered at certain examination times. Two usual models are random right-censorship and random interval-censorship. In the first case, the observations are assumed to be of the form $Y_i = \min(X_i, W_i)$, $i = 1, \ldots, n$, where the X_i are the (partially observed) survival times, and the W_i are the censoring times. In this model, both survival and censoring times are assumed to be random, and mutually independent. Estimating the parameters of Weibull distribution from such data have been considered

by several authors. See, for example Ageel (2002), Balakrishnan & Kateri (2008), Nandi & Dewan (2010), Joarder, Krishna & Kundu (2011), Banerjee & Kundu (2012), and Lin, Chou & Huang (2012). In the case of so-called random interval censored data, the event is only known to happen between two random examination times. The observations are thus of the form (U_i, V_i), $i = 1, \ldots, n$, and it is only known that $U_i \leq X_i \leq V_i$ for all i. Here again it is customary to assume independence between survival times X_i and censoring interval endpoints. Statistical analysis of Weibull distribution based on interval censored data has been discussed by Ng & Wang (2009) and Tan (2009), among others. Truncation is similar to but distinct from the concept of censoring. When the existence of the unseen "observation " is not known for observations that fall outside the particular range, the data that are observed are said to be truncated. Recently, Balakrishnan & Mitra (2012) developed the EM algorithm for the estimation of the parameters of the Weibull distribution based on left truncated and right censored data.

The problem addressed in this paper, is different from censoring and trancation. We are not concerned with imprecision arising from random inspection times, but with the situation in which the result of a random experiment is reported from the observer to the statistician with some imprecision, arising from its limited perception or recollection of the precise numerical values. For instance, the lifetime of some shaft may be reported as imprecise quantities such as: "about $1,000h$", "approximately $1,400h$", "almost between $1,000h$ and $1,200h$", "essentially less than $1,200h$", and so on. The lack of precision of such data can be described using fuzzy sets. The classical statistical estimation methods are not appropriate to deal with fuzzy sets. Therefore, the conventional procedures used for estimating the parameters of Weibull distribution will have to be adapted to the new situation. The main aim of this paper is to develop the inferential procedures for the two-parameter Weibull distribution when the available data are in the form of fuzzy numbers. In Section 2, we review the fundamental notation and basic definitions of fuzzy set theory. In Section 3, we first introduce a generalized likelihood function based on fuzzy data. We then discuss the computation of maximum likelihood estimates (MLEs) of the parameters α and λ by using the Newton-Raphson (NR) and Expectation Maximization (EM) algorithms, in Section 4. In Section 5, the Bayes estimates of the unknown parameters are obtained by using the approximation form of Tierney & Kadane (1986) under the assumption of Gamma priors. The estimation via method of moments is provided in Section 6. A Monte Carlo simulation study is presented in Section 7, which provides a comparison of all estimation procedures developed in this paper and one real data set is analyzed for illustrative purposes.

2. Basic Definition of Fuzzy Sets

To appreciate the nature of a fuzzy set, let us consider the following hypothetical example taken from Gertner & Zhu (1996). Consider an experiment characterized by a probability space $\mathcal{S} = (\mathcal{X}, \mathcal{B}_\mathcal{X}, P_\theta)$, where $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ is a measurable space and P_θ belongs to a specified family of probability measures $\{P_\theta, \theta \in \Theta\}$ on $(\mathcal{X}, \mathcal{B}_\mathcal{X})$.

Any indicator function $I_A : \mathcal{X} \to \{0, 1\}$, defined by

$$I_A(x) = \begin{cases} 1 & x \in A, \\ 0 & x \notin A, \end{cases}$$

characterizes a crisp subset A in $\mathcal{X}$. For example, if $\mathcal{X} = \{x_i, i = 1, \ldots, n\}$, represents all trees in a forest stand, then $A = \{x,\ x\text{'s age} \leq 40 \text{ yr}\}$ is its subset. So if tree x_3 is 27 yr old, $x_3 \in A$ and $I_A(x_3) = 1$; and if x_{239}'s age equals 56, $x_{239} \notin A$ and $I_A(x_{239}) = 0$. However, when referring to a "young tree", the set above described becomes a fuzzy set. Now relate each tree to its youthfulness by assigning a value between 1, representing absolutely young, and 0, representing absolutely not young, as the membership degree describing the subjective uncertainty of a tree being considered young. For instance, $\mu_{\text{young}}(x_3) = 0.9$, since x_3 will most likely be allocated into a younger class, whereas $\mu_{\text{young}}(x_{239}) = 0.49$ for x_{239} seems neither very young nor very old compared to other older trees in that stand. Thus, similar to crisp sets, a fuzzy subset $\widetilde{A}$ in $\mathcal{X}$ is characterized by a membership function $\mu_{\tilde{A}}(x)$ which associates with each point x in $\mathcal{X}$ a real number in the interval $[0, 1]$, with the value of $\mu_{\tilde{A}}(x)$ at x representing the "grade of membership"of x in $\widetilde{A}$. We hereafter assume that the sample space $\mathcal{X}$ is a set in a Euclidean space and $\mathcal{B}_{\mathcal{X}}$ is the smallest Borel $\sigma-$field on $\mathcal{X}$. A fuzzy event in $\mathcal{X}$ is a fuzzy subset $\tilde{A}$ of $\mathcal{X}$, whose membership function $\mu_{\tilde{A}}$ is Borel measurable. Many examples of fuzzy samples and observations appear in social and natural sciences. These occur when the linguistic concepts or propositions cannot be precisely defined, or accurate measurements of variables are not possible or necessary.

Example 1. An investigator is interested in analyzing the amount of an adverse substance extracted from a special brand of cigarettes. Assume that the investigator has not a mechanism of measurement which is sufficiently precise to determine exactly the amount of adverse substance of cigarettes, but rather he can only approximate them by means of imprecise observations, for instance, "The amount of adverse substance of cigarette is approximately 30 to 40 milligrams". A fuzzy approach lies in expressing the preceding observation as a fuzzy event $\widetilde{A}$ such as that defined, for instance, by the membership function (Figure 1).

$$\mu_{\widetilde{A}}(x) = \begin{cases} \frac{x-20}{10} & 20 \leq x \leq 30, \\ 1 & 30 \leq x \leq 40, \\ \frac{50-x}{10} & 40 \leq x \leq 50, \\ 0 & otherwise, \end{cases}$$

The notion of probability was extended to fuzzy events by Zadeh (1968) as follows.

Definition 1. Let $(\mathbb{R}^n, \mathcal{A}, P)$ be a probability space in which $\mathcal{A}$ is the $\sigma-$field of Borel sets in $\mathbb{R}^n$ and P is a probability measure over $\mathbb{R}^n$. Then, the probability of a fuzzy event $\widetilde{A}$ in $\mathbb{R}^n$ is defined by:

$$P(\widetilde{A}) = \int \mu_{\widetilde{A}}(\boldsymbol{x}) dP. \tag{3}$$

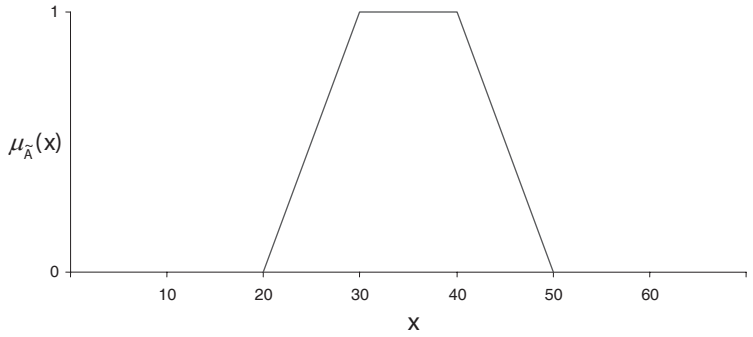


FIGURE 1: Fuzzy approach of the imprecise observation "approximately 30 to 40 ".

In particular, assume that P is the probability distribution of a continuous random variable $\boldsymbol{Y}$ with p.d.f. $g(\boldsymbol{Y})$. The conditional density of $\boldsymbol{Y}$ given $\widetilde{A}$ is given by

$$g(\boldsymbol{y} \mid \widetilde{A}) = \frac{\mu_{\widetilde{A}}(\boldsymbol{y})g(\boldsymbol{y})}{\int \mu_{\widetilde{A}}(\boldsymbol{u})g(\boldsymbol{u})d\boldsymbol{u}}. \tag{4}$$

The set consisting of all observable events from the experiment $\mathcal{S}$ determines a fuzzy information system (f.i.s.) associated with it, which is defined as follows.

Definition 2. (Tanaka **?**). A fuzzy information system $\widetilde{\mathcal{S}}$ associated with the experiment $\mathcal{S}$ is a fuzzy partition $\mathcal{F} = \{\widetilde{x}_1, \ldots, \widetilde{x}_K\}$ of $\mathcal{X}$, i.e., a set of K fuzzy events on $\mathcal{X}$ satisfying the orthogonality condition

$$\sum_{k=1}^{K} \mu_{\widetilde{x}_k}(x) = 1,$$

where $\mu_{\widetilde{x}_k}$ denotes the membership function of $\widetilde{x}_k$.

We now examine a brief example illustrating the preceding concept:

Example 2. To evaluate the problem of psychological depression in a population, there is no exact method that can measure and express the exact value for the severity of the disease in each person and, so measurement results may be reported by means of the following fuzzy observations: $\widetilde{x}_1 =$ "approximately lower than 20", $\widetilde{x}_2 =$"approximately 25 to 30", $\widetilde{x}_3 =$ "approximately 35 ", $\widetilde{x}_4 =$"approximately 40 to 45", $\widetilde{x}_5 =$ "approximately 50", $\widetilde{x}_6 =$ "approximately higher than 55", which are characterized by the membership functions

$$\mu_{\widetilde{x}_1}(x) = \begin{cases} 1 & x \leq 20, \\ \frac{25-x}{5} & 20 \leq x \leq 25, \\ 0 & otherwise, \end{cases} \qquad \mu_{\widetilde{x}_2}(x) = \begin{cases} \frac{x-20}{5} & 20 \leq x \leq 25, \\ 1 & 25 \leq x \leq 30, \\ \frac{35-x}{5} & 30 \leq x \leq 35, \\ 0 & otherwise, \end{cases}$$

$$\mu_{\widetilde{x}_3}(x) = \begin{cases} \frac{x-30}{5} & 30 \le x \le 35, \\ \frac{40-x}{5} & 35 \le x \le 40, \\ 0 & otherwise, \end{cases} \qquad \mu_{\widetilde{x}_4}(x) = \begin{cases} \frac{x-35}{5} & 35 \le x \le 40, \\ 1 & 40 \le x \le 45, \\ \frac{50-x}{5} & 45 \le x \le 50, \\ 0 & otherwise, \end{cases}$$

$$\mu_{\widetilde{x}_5}(x) = \begin{cases} \frac{x-45}{5} & 45 \le x \le 50, \\ \frac{55-x}{5} & 50 \le x \le 55, \\ 0 & otherwise, \end{cases} \qquad \mu_{\widetilde{x}_6}(x) = \begin{cases} \frac{x-50}{5} & 50 \le x \le 55, \\ 1 & x \ge 55, \\ 0 & otherwise, \end{cases}$$

respectively, (see Fig.2). Clearly, a f.i.s. $\widetilde{\mathcal{S}} = \{\widetilde{x}_1, ..., \widetilde{x}_7\}$ can be immediately constructed by defining $\mu_{\widetilde{x}_7} = 1 - \sum_{i=1}^{6} \mu_{\widetilde{x}_i}$)

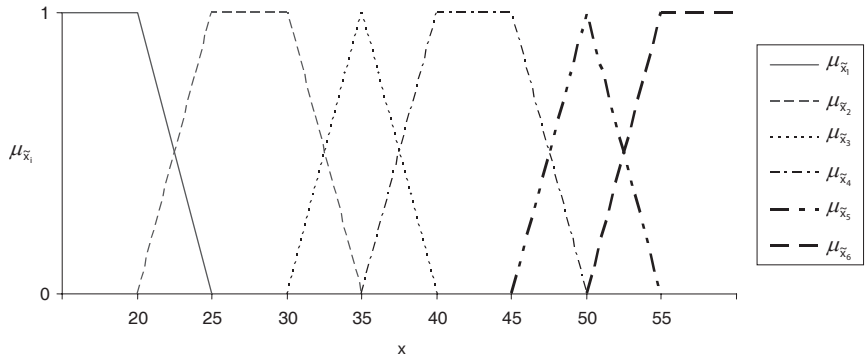


FIGURE 2: Membership functions of the fuzzy observations $\widetilde{x}_1$, $\widetilde{x}_2$, $\widetilde{x}_3$, $\widetilde{x}_4$, $\widetilde{x}_5$ and $\widetilde{x}_6$.

For more details about the membership functions and probability measures of fuzzy sets, one can refer to Singpurwalla & Booker (2004).

In order to model imprecise data, a generalization of real numbers is necessary. These data can be represented by fuzzy numbers. A fuzzy number is a subset, denoted by $\widetilde{x}$, of the set of real numbers (denoted by $\mathbb{R}$) and is characterized by the so called membership function $\mu_{\widetilde{x}}(.)$. Fuzzy numbers satisfy the following constraints (see Dubois & Prade (1980)):

(1) $\mu_{\widetilde{x}} : \mathbb{R} \longrightarrow [0, 1]$ is Borel-measurable;

(2) $\exists x_0 \in \mathbb{R} : \mu_{\widetilde{x}}(x_0) = 1$;

(3) The so-called $\lambda-$cuts $(0 < \lambda \le 1)$, defined as $B_\lambda(\widetilde{x}) = \{x \in \mathbb{R} : \mu_{\widetilde{x}}(x) \ge \lambda\}$, are all closed intervals, i.e., $B_\lambda(\widetilde{x}) = [a_\lambda, b_\lambda], \forall \lambda \in (0, 1]$.

With the definition of a fuzzy number given above, an exact (non-fuzzy) number can be treated as a special case of a fuzzy number. For a non-fuzzy real observation $x_0 \in \mathbb{R}$, its corresponding membership function is $\mu_{x_0}(x_0) = 1$.

Among the various types of fuzzy numbers, the triangular and trapezoidal fuzzy numbers are most convenient and useful in describing fuzzy data. For triangular membership functions, the triangular fuzzy number can be defined as $\widetilde{x} = (a, b, c)$ and its membership function is defined by the following expression:

$$\mu_{\widetilde{x}}(x) = \begin{cases} \frac{x-a}{b-a} & a \leq x \leq b, \\ \frac{c-x}{c-b} & b \leq x \leq c, \\ 0 & otherwise. \end{cases}$$

The trapezoidal fuzzy number can be defined as $\widetilde{x} = (a, b, c, d)$ with membership function

$$\mu_{\widetilde{x}}(x) = \begin{cases} \frac{x-a}{b-a} & a \leq x \leq b, \\ 1 & b \leq x \leq c, \\ \frac{d-x}{d-c} & c \leq x \leq d, \\ 0 & otherwise. \end{cases}$$

3. Fuzzy Data and the Likelihood Function

Suppose that $X_1, \ldots, X_n$ is a random sample of size n from Weibull distribution with pdf given by (1). Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ denotes the corresponding random vector. If a realization $\boldsymbol{x} = (x_1, \ldots, x_n)$ of $\boldsymbol{X}$ was known exactly, we could obtain the complete-data likelihood function as

$$L(\alpha, \lambda; \boldsymbol{x}) = \alpha^n \lambda^n \exp(-\lambda \sum_{i=1}^{n} x_i^{\alpha}) \prod_{i=1}^{n} x_i^{\alpha-1} \tag{5}$$

Now consider the problem where $\boldsymbol{x}$ is not observed precisely and only partial information about $\boldsymbol{x}$ is available in the form of a fuzzy subset $\widetilde{\boldsymbol{x}}$ with the Borel measurable membership function $\mu_{\widetilde{\boldsymbol{x}}}(\boldsymbol{x})$. In this setting, the fuzzy observation $\widetilde{\boldsymbol{x}}$ can be understood as encoding the observer's partial knowledge about the realization $\boldsymbol{x}$ of random vector $\boldsymbol{X}$, and the membership function $\mu_{\widetilde{\boldsymbol{x}}}$ is seen as a possibility distribution interpreted as a soft constraint on the unknown quantity $\boldsymbol{x}$. The fuzzy set $\widetilde{\boldsymbol{x}}$ can be considered to be generated by a two-step process:

1. A realization $\boldsymbol{x}$ is drawn from $\boldsymbol{X}$;
2. The observer encodes his/her partial knowledge of $\boldsymbol{x}$ in the form of a possibility distribution $\mu_{\widetilde{\boldsymbol{x}}}$.

It must be noted that, in this model, only step 1 is considered to be a random experiment. Step 2 implies gathering information about $\boldsymbol{x}$ and modeling this information as a possibility distribution.

Example 3. Consider a life-testing experiment in which n identical ball bearings are placed on test, and we are interested in the lifetime of these ball bearings. The unknown lifetime x_i of ball bearing i may be regarded as a realization of a

random variable X_i induced by random sampling from a total population of ball bearings. In practice, however, measuring the lifetime of a ball bearing may not yield an exact result. A ball bearing may work perfectly over a certain period but be braking for some time, and finally be unusable at a certain time. Assume that two intervals are determined for the lifetime of each ball bearing i as follows:

- an interval $[a_i, d_i]$ certainly containing x_i;
- an interval $[b_i, c_i]$ containing highly plausible values for x_i.

This information may be encoded as a trapeozoidal fuzzy number $\widetilde{x}_i = (a_i, b_i, c_i, d_i)$ with support $[a_i, d_i]$ and core $[b_i, c_i]$, interpreted as a possibility distribution constraining the unknown value x_i. Information about $\boldsymbol{x}$ may be represented by the joint possibility distribution

$$\mu_{\widetilde{\boldsymbol{x}}}(\boldsymbol{x}) = \mu_{\widetilde{x}_1}(x_1) \times ... \times \mu_{\widetilde{x}_n}(x_n). \tag{6}$$

Once $\widetilde{\boldsymbol{x}}$ is given, and assuming its membership function to be the Borel measurable, we can compute its probability according to Zadeh's definition of the probability of a fuzzy event. By using the expression (3), the observed-data likelihood function can then be obtained as

$$L_O(\alpha, \lambda; \widetilde{\boldsymbol{x}}) = P(\widetilde{\boldsymbol{x}}; \alpha, \lambda) = \int f(\boldsymbol{x}; \alpha, \lambda) \mu_{\widetilde{\boldsymbol{x}}}(\boldsymbol{x}) d\boldsymbol{x}. \tag{7}$$

Since the data vector $\boldsymbol{x}$ is a realization of an independent identically distributed (i.i.d.) random vector $\boldsymbol{X}$, and assuming the joint membership function $\mu_{\widetilde{\boldsymbol{x}}}(\boldsymbol{x})$ to be decomposable as in (6), the likelihood function (7) can be written as:

$$L_O(\alpha, \lambda; \widetilde{\boldsymbol{x}}) = \prod_{i=1}^{n} \int \alpha\lambda x^{\alpha-1} \exp(-\lambda x^{\alpha}) \mu_{\widetilde{x}_i}(x) dx, \tag{8}$$

and the observed-data log likelihood is

$$\begin{aligned} L^*(\alpha, \lambda; \widetilde{\boldsymbol{x}}) &= \log L_O(\alpha, \lambda; \widetilde{\boldsymbol{x}}) \\ &= n(\log \alpha + \log \lambda) + \sum_{i=1}^{n} \log \int x^{\alpha-1} \exp(-\lambda x^{\alpha}) \mu_{\widetilde{x}_i}(x) dx. \end{aligned} \tag{9}$$

4. Maximum Likelihood Estimation

The idea behind maximum likelihood parameter estimation is to determine the parameters that maximize the probability (likelihood) of the sample data. From a statistical point of view, the method of maximum likelihood is considered to be more robust and yields estimators with good statistical properties. In other words, maximum likelihood methods are versatile and apply to most models and to different types of data. The maximum likelihood estimate of the parameters α

and λ can be obtained by maximizing the log-likelihood $L^*(\alpha, \lambda; \widetilde{\boldsymbol{x}})$. Equating the partial derivatives of the log-likelihood (9) with respect to α and λ to zero, the resulting two equations are:

$$\frac{\partial}{\partial\alpha}L^*(\alpha,\lambda;\widetilde{\boldsymbol{x}}) = \frac{n}{\alpha} + \sum_{i=1}^{n} \frac{\int (x^{\alpha-1} - \lambda x^{2\alpha-1}) \log x \exp\left(-\lambda x^{\alpha}\right) \mu_{\widetilde{x}_i}(x)dx}{\int x^{\alpha-1} \exp\left(-\lambda x^{\alpha}\right) \mu_{\widetilde{x}_i}(x)dx} = 0 \quad (10)$$

and

$$\frac{\partial}{\partial\lambda}L^*(\alpha,\lambda;\widetilde{\boldsymbol{x}}) = \frac{n}{\lambda} - \sum_{i=1}^{n} \frac{\int x^{2\alpha-1} \exp\left(-\lambda x^{\alpha}\right) \mu_{\widetilde{x}_i}(x)dx}{\int x^{\alpha-1} \exp\left(-\lambda x^{\alpha}\right) \mu_{\widetilde{x}_i}(x)dx} = 0. \quad (11)$$

Since there are no closed form of the solutions to the likelihood equations (10) and (11), an iterative numerical search can be used to obtain the MLEs. In the following, we describe the NR method and the EM algorithm to determine the MLEs of the parameters α and λ.

4.1. NR Algorithm

NR algorithm is a direct approach for estimating the relevant parameters in a likelihood function. In this algorithm, the solution of the likelihood equation is obtained through an iterative procedure. Let $\boldsymbol{\theta} = (\alpha, \lambda)^T$ be the parameter vector. Then, at the $(h+1)$th step of iteration process, the updated parameter is obtained as

$$\boldsymbol{\theta}^{(h+1)} = \boldsymbol{\theta}^{(h)} - \left[\frac{\partial^2 L^*(\boldsymbol{\theta};\widetilde{\boldsymbol{x}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} \mid_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(h)}}\right]^{-1} \left[\frac{\partial L^*(\boldsymbol{\theta};\widetilde{\boldsymbol{x}})}{\partial\boldsymbol{\theta}} \mid_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(h)}}\right] \quad (12)$$

where

$$\frac{\partial L^*(\boldsymbol{\theta};\widetilde{\boldsymbol{x}})}{\partial\boldsymbol{\theta}} = \begin{pmatrix} \frac{\partial L^*(\alpha,\lambda;\widetilde{\boldsymbol{x}})}{\partial\alpha} \\ \frac{\partial L^*(\alpha,\lambda;\widetilde{\boldsymbol{x}})}{\partial\lambda} \end{pmatrix}$$

and

$$\frac{\partial^2 L^*(\boldsymbol{\theta};\widetilde{\boldsymbol{x}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} = \begin{pmatrix} \frac{\partial^2 L^*(\alpha,\lambda;\widetilde{\boldsymbol{x}})}{\partial\alpha^2} & \frac{\partial^2 L^*(\alpha,\lambda;\widetilde{\boldsymbol{x}})}{\partial\lambda\partial\alpha} \\ \frac{\partial^2 L^*(\alpha,\lambda;\widetilde{\boldsymbol{x}})}{\partial\lambda\partial\alpha} & \frac{\partial^2 L^*(\alpha,\lambda;\widetilde{\boldsymbol{x}})}{\partial\lambda^2} \end{pmatrix}$$

The second-order derivatives of the log-likelihood with respect to the parameters, required for proceeding with the NR method, are obtained as follows.

$$\begin{aligned} \frac{\partial^2}{\partial\alpha^2}L^*(\alpha,\lambda;\widetilde{\boldsymbol{x}}) &= -\frac{n}{\alpha^2} \\ &\quad + \sum_{i=1}^{n}\Big\{\frac{\int(\lambda^2 x^{3\alpha-1} - \lambda x^{2\alpha-1})(\log x)^2 \exp\left(-\lambda x^{\alpha}\right) \mu_{\widetilde{x}_i}(x)dx}{\int x^{\alpha-1} \exp\left(-\lambda x^{\alpha}\right) \mu_{\widetilde{x}_i}(x)dx} \\ &\qquad + \frac{\int(x^{\alpha-1} - 2\lambda x^{2\alpha-1})(\log x)^2 \exp\left(-\lambda x^{\alpha}\right) \mu_{\widetilde{x}_i}(x)dx}{\int x^{\alpha-1} \exp\left(-\lambda x^{\alpha}\right) \mu_{\widetilde{x}_i}(x)dx}\Big\} \\ &\quad - \sum_{i=1}^{n}\left[\frac{\int(x^{\alpha-1} - \lambda x^{2\alpha-1}) \log x \exp\left(-\lambda x^{\alpha}\right) \mu_{\widetilde{x}_i}(x)dx}{\int x^{\alpha-1} \exp\left(-\lambda x^{\alpha}\right) \mu_{\widetilde{x}_i}(x)dx}\right]^2 \end{aligned}$$

$$\frac{\partial^2}{\partial\lambda^2}L^*(\alpha,\lambda;\widetilde{\boldsymbol{x}}) = -\frac{n}{\lambda^2}+\sum_{i=1}^{n}\{\frac{\int x^{3\alpha-1}\exp(-\lambda x^\alpha)\,\mu_{\widetilde{x}_i}(x)dx}{\int x^{\alpha-1}\exp(-\lambda x^\alpha)\,\mu_{\widetilde{x}_i}(x)dx} -[\frac{\int x^{2\alpha-1}\exp(-\lambda x^\alpha)\,\mu_{\widetilde{x}_i}(x)dx}{\int x^{\alpha-1}\exp(-\lambda x^\alpha)\,\mu_{\widetilde{x}_i}(x)dx}]^2\},$$

$$\begin{aligned}\frac{\partial^2}{\partial\lambda\partial\alpha}L^*(\alpha,\lambda;\widetilde{\boldsymbol{x}}) &= -\sum_{i=1}^{n}\frac{\int(2x^{2\alpha-1}-\lambda x^{3\alpha-1})\log x\exp(-\lambda x^\alpha)\,\mu_{\widetilde{x}_i}(x)dx}{\int x^{\alpha-1}\exp(-\lambda x^\alpha)\,\mu_{\widetilde{x}_i}(x)dx}\\ &+\sum_{i=1}^{n}\{\frac{\int(1-\lambda x^\alpha)x^{\alpha-1}\log x\exp(-\lambda x^\alpha)\,\mu_{\widetilde{x}_i}(x)dx}{\int x^{\alpha-1}\exp(-\lambda x^\alpha)\,\mu_{\widetilde{x}_i}(x)dx}\\ &\times\frac{\int x^{2\alpha-1}\exp(-\lambda x^\alpha)\,\mu_{\widetilde{x}_i}(x)dx}{\int x^{\alpha-1}\exp(-\lambda x^\alpha)\,\mu_{\widetilde{x}_i}(x)dx}\}\end{aligned}$$

The iteration process then continues until convergence, i.e., until $\| \boldsymbol{\theta}^{(h+1)}-\boldsymbol{\theta}^{(h)} \|<\varepsilon$, for some pre-fixed $\varepsilon>0$. The maximum likelihood estimate of (α,λ) via NR algorithm is thereafter refereed as "$(\hat{\alpha}_{NR},\hat{\lambda}_{NR})$" in this paper.

It should be pointed out that the second-order derivatives of the log-likelihood are required at every iteration in the NR method. Sometimes the calculation of the derivatives based on fuzzy data can be rather tedious. Another viable alternative to the NR algorithm is the well-known EM algorithm. In the following, we discuss how that can be used to determine the MLEs in this case.

4.2. EM Algorithm

The EM algorithm is a broadly applicable approach to the iterative computation of maximum likelihood estimates and useful in a variety of incomplete-data problems. Since the observed fuzzy data $\widetilde{\boldsymbol{x}}$ can be seen as an incomplete specification of a complete data vector $\boldsymbol{x}$, the EM algorithm is applicable to obtain the maximum likelihood estimates of the unknown parameters. In the following, we use the fuzzy EM algorithm (see Denoeux (2011)) to determine the MLEs of α and λ.

From the Eq. (5), the log-likelihood function for the complete data vector $\boldsymbol{x}$ becomes:

$$\log L(\alpha,\lambda;\boldsymbol{x}) = n\log\alpha+n\log\lambda+(\alpha-1)\sum_{i=1}^{n}\log x_i-\lambda\sum_{i=1}^{n}x_i^\alpha \tag{13}$$

Taking the derivative with respect to α and λ, respectively, on (13), the following likelihood equations are obtained:

$$\frac{n}{\lambda}=\sum_{i=1}^{n}x_i^\alpha \tag{14}$$

and

$$\frac{n}{\alpha}=\lambda\sum_{i=1}^{n}x_i^\alpha\log x_i-\sum_{i=1}^{n}\log x_i \tag{15}$$

Therefore the EM algorithm is given by the following iterative process:

1. Given starting values of α and λ, say $\alpha^{(0)}$ and $\lambda^{(0)}$ and set $h = 0$.
2. In the $(h+1)$th iteration,

- The E-step requires to compute the following conditional expectations using the expression (4):

$$E_{1i} = E_{\alpha^{(h)},\lambda^{(h)}}(X^{\alpha} \mid \widetilde{x}_i) = \frac{\int x^{2\alpha^{(h)}-1} \exp\left(-\lambda^{(h)} x^{\alpha^{(h)}}\right) \mu_{\widetilde{x}_i}(x)dx}{\int x^{\alpha^{(h)}-1} \exp\left(-\lambda^{(h)} x^{\alpha^{(h)}}\right) \mu_{\widetilde{x}_i}(x)dx}$$

$$E_{2i} = E_{\alpha^{(h)},\lambda^{(h)}}(\log X \mid \widetilde{x}_i) = \frac{\int x^{\alpha^{(h)}-1} \log x \exp\left(-\lambda^{(h)} x^{\alpha^{(h)}}\right) \mu_{\widetilde{x}_i}(x)dx}{\int x^{\alpha^{(h)}-1} \exp\left(-\lambda^{(h)} x^{\alpha^{(h)}}\right) \mu_{\widetilde{x}_i}(x)dx}$$

$$\begin{aligned} E_{3i} &= E_{\alpha^{(h)},\lambda^{(h)}}(X^{\alpha} \log X \mid \widetilde{x}_i) \\ &= \frac{\int x^{2\alpha^{(h)}-1} \log x \exp\left(-\lambda^{(h)} x^{\alpha^{(h)}}\right) \mu_{\widetilde{x}_i}(x)dx}{\int x^{\alpha^{(h)}-1} \exp\left(-\lambda^{(h)} x^{\alpha^{(h)}}\right) \mu_{\widetilde{x}_i}(x)dx} \end{aligned}$$

and the likelihood equations (14) and (15) are replaced by

$$\frac{n}{\lambda} = \sum_{i=1}^{n} E_{1i}, \tag{16}$$

and

$$\frac{n}{\alpha} = \lambda \sum_{i=1}^{n} \left[E_{3i} - E_{2i}\right]. \tag{17}$$

- The M-step requires to solve the Eqs. (16) and (17), and obtain the next values, $\lambda^{(h+1)}$ and $\alpha^{(h+1)}$, of λ and α, respectively, as follows:

$$\lambda^{(h+1)} = \frac{n}{\sum\limits_{i=1}^{n} E_{1i}}$$

$$\alpha^{(h+1)} = \left\{ \frac{1}{n} \lambda^{(h+1)} \sum_{i=1}^{n} \left[E_{3i} - E_{2i}\right] \right\}^{-1}$$

3. Checking convergence, if the convergence occurs then the current $\alpha^{(h+1)}$ and $\lambda^{(h+1)}$ are the maximum likelihood estimates of α and λ via EM algorithm; otherwise, set $h = h + 1$ and go to Step 2.

The maximum likelihood estimate of (α, λ) via EM algorithm is thereafter refereed as "$(\hat{\alpha}_{EM}, \hat{\lambda}_{EM})$" in this paper.

5. Bayesian Estimation

In recent decades, the Bayes viewpoint, as a powerful and valid alternative to traditional statistical perspectives, has received frequent attention for statistical inference. In this section, we consider the Bayesian estimation under the assumptions that α and λ have independent gamma priors with the pdfs

$$\pi_1(\alpha) = \frac{d^c}{\Gamma(c)}\alpha^{c-1}\exp(-\alpha d), \qquad \alpha > 0 \tag{18}$$

and

$$\pi_2(\lambda) = \frac{b^a}{\Gamma(a)}\lambda^{a-1}\exp(-\lambda b), \qquad \lambda > 0 \tag{19}$$

with the parameters $\alpha \sim Gamma(c, d)$ and $\lambda \sim Gamma(a, b)$. Based on the above priors, the joint posterior density function of α and λ given the data can be written as follows:

$$\pi(\alpha, \lambda \mid \widetilde{\boldsymbol{x}}) = \frac{\pi_1(\alpha)\pi_2(\lambda)\ell(\alpha, \lambda; \widetilde{\boldsymbol{x}})}{\int_0^\infty \int_0^\infty \pi_1(\alpha)\pi_2(\lambda)\ell(\alpha, \lambda; \widetilde{\boldsymbol{x}})d\alpha d\lambda} \tag{20}$$

where

$$\ell(\alpha, \lambda; \widetilde{\boldsymbol{x}}) = \alpha^{(n+c-1)}\lambda^{(n+a-1)}\exp(-\alpha d)\exp(-\lambda b)\prod_{i=1}^{n}\int x^{\alpha-1}\exp\left(-\lambda x^\alpha\right)\mu_{\widetilde{x}_i}(x)dx$$

is the likelihood function based on the fuzzy sample $\widetilde{\boldsymbol{x}}$. Then, under a squared error loss function, the Bayes estimate of any function of α and λ, say $g(\alpha, \lambda)$, is

$$\begin{aligned} E(g(\alpha, \lambda) \mid \widetilde{\boldsymbol{x}}) &= \frac{\int_0^\infty \int_0^\infty g(\alpha, \lambda)\pi_1(\alpha)\pi_2(\lambda)\ell(\alpha, \lambda; \widetilde{\boldsymbol{x}})d\alpha d\lambda}{\int_0^\infty \int_0^\infty \pi_1(\alpha)\pi_2(\lambda)\ell(\alpha, \lambda; \widetilde{\boldsymbol{x}})d\alpha d\lambda} \\ &= \frac{\int_0^\infty \int_0^\infty g(\alpha, \lambda)e^{Q(\alpha,\lambda)}d\alpha d\lambda}{\int_0^\infty \int_0^\infty e^{Q(\alpha,\lambda)}d\alpha d\lambda} \end{aligned} \tag{21}$$

where $Q(\alpha, \lambda) = \ln[\pi_1(\alpha)\pi_2(\lambda)] + \ln \ell(\alpha, \lambda; \widetilde{\boldsymbol{x}}) \equiv \rho(\alpha, \lambda) + L(\alpha, \lambda)$. Note that Eq. (21) cannot be obtained analytically; therefore, in the following we adopt Tierney and Kadane's approximation for computing the Bayes estimates.

Setting $H(\alpha, \lambda) = Q(\alpha, \lambda)/n$ and $H^*(\alpha, \lambda) = [\ln g(\alpha, \lambda) + Q(\alpha, \lambda)]/n$, the expression in (21) can be reexpressed as

$$E(g(\alpha, \lambda) \mid \widetilde{\boldsymbol{x}}) = \frac{\int_0^\infty \int_0^\infty e^{nH^*(\alpha,\lambda)}d\alpha d\lambda}{\int_0^\infty \int_0^\infty e^{nH(\alpha,\lambda)}d\alpha d\lambda} \tag{22}$$

Following Tierney & Kadane (1986), Eq. (22) can be approximated as the following form:

$$\hat{g}_{BT}(\alpha,\lambda)=\left[\frac{\det\Sigma^*}{\det\Sigma}\right]^{1/2}\exp\left\{n\left[H^*(\bar{\alpha}^*,\bar{\lambda}^*)-H(\bar{\alpha},\bar{\lambda})\right]\right\} \tag{23}$$

where $(\bar{\alpha}^*,\bar{\lambda}^*)$ and $(\bar{\alpha},\bar{\lambda})$ maximize $H^*(\alpha,\lambda)$ and $H(\alpha,\lambda)$, respectively, and Σ^* and Σ are the negatives of the inverse Hessians of $H^*(\alpha,\lambda)$ and $H(\alpha,\lambda)$ at $(\bar{\alpha}^*,\bar{\lambda}^*)$ and $(\bar{\alpha},\bar{\lambda})$, respectively.

In our case, we have

$$\begin{aligned}H(\alpha,\lambda) &= \frac{1}{n}\{k+(n+c-1)\log\alpha+(n+a-1)\log\lambda-\alpha d\\ &\quad -\lambda b+\sum_{i=1}^{n}\log\int x^{\alpha-1}\exp\left(-\lambda x^{\alpha}\right)\mu_{\widetilde{x}_i}(x)dx\}.\end{aligned}$$

where k is a constant; therefore, $(\bar{\alpha},\bar{\lambda})$ can be obtained by solving the following two equations

$$\begin{aligned}\frac{\partial}{\partial\alpha}H(\alpha,\lambda) &= \frac{1}{n}\{\frac{n+c-1}{\alpha}-d\\ &\quad +\sum_{i=1}^{n}\frac{\int(x^{\alpha-1}-\lambda x^{2\alpha-1})\log x\exp\left(-\lambda x^{\alpha}\right)\mu_{\widetilde{x}_i}(x)dx}{\int x^{\alpha-1}\exp\left(-\lambda x^{\alpha}\right)\mu_{\widetilde{x}_i}(x)dx}\}\end{aligned}$$

$$\frac{\partial}{\partial\lambda}H(\alpha,\lambda)=\frac{1}{n}\left\{\frac{n+a-1}{\lambda}-b-\sum_{i=1}^{n}\frac{\int x^{2\alpha-1}\exp\left(-\lambda x^{\alpha}\right)\mu_{\widetilde{x}_i}(x)dx}{\int x^{\alpha-1}\exp\left(-\lambda x^{\alpha}\right)\mu_{\widetilde{x}_i}(x)dx}\right\}$$

and, from the second derivatives of $H(\alpha,\lambda)$, the determinant of the negative of the inverse Hessian of $H(\alpha,\lambda)$ at $(\bar{\alpha},\bar{\lambda})$ is given by

$$\det\Sigma=(H_{11}H_{22}-H_{12}^2)^{-1}$$

where

$$\begin{aligned}H_{11} &= \frac{1}{n}\{-\frac{n+c-1}{\bar{\alpha}^2}\\ &\quad +\sum_{i=1}^{n}(\frac{\int(\bar{\lambda}^2x^{3\bar{\alpha}-1}-\bar{\lambda}x^{2\bar{\alpha}-1})(\log x)^2\exp\left(-\bar{\lambda}x^{\bar{\alpha}}\right)\mu_{\widetilde{x}_i}(x)dx}{\int x^{\bar{\alpha}-1}\exp\left(-\bar{\lambda}x^{\bar{\alpha}}\right)\mu_{\widetilde{x}_i}(x)dx}\\ &\quad +\frac{\int(x^{\bar{\alpha}-1}-2\bar{\lambda}x^{2\bar{\alpha}-1})(\log x)^2\exp\left(-\bar{\lambda}x^{\bar{\alpha}}\right)\mu_{\widetilde{x}_i}(x)dx}{\int x^{\bar{\alpha}-1}\exp\left(-\bar{\lambda}x^{\bar{\alpha}}\right)\mu_{\widetilde{x}_i}(x)dx})\\ &\quad -\sum_{i=1}^{n}[\frac{\int(x^{\bar{\alpha}-1}-\bar{\lambda}x^{2\bar{\alpha}-1})\log x\exp\left(-\bar{\lambda}x^{\bar{\alpha}}\right)\mu_{\widetilde{x}_i}(x)dx}{\int x^{\bar{\alpha}-1}\exp\left(-\bar{\lambda}x^{\bar{\alpha}}\right)\mu_{\widetilde{x}_i}(x)dx}]^2\ \}\end{aligned}$$

$$\begin{aligned}H_{22} &= \frac{1}{n}\{-\frac{n+a-1}{\bar{\lambda}^2}+\sum_{i=1}^{n}(\frac{\int x^{3\bar{\alpha}-1}\exp\left(-\bar{\lambda}x^{\bar{\alpha}}\right)\mu_{\widetilde{x}_i}(x)dx}{\int x^{\bar{\alpha}-1}\exp\left(-\bar{\lambda}x^{\bar{\alpha}}\right)\mu_{\widetilde{x}_i}(x)dx}\\ &\quad -[\frac{\int x^{2\bar{\alpha}-1}\exp\left(-\bar{\lambda}x^{\bar{\alpha}}\right)\mu_{\widetilde{x}_i}(x)dx}{\int x^{\bar{\alpha}-1}\exp\left(-\bar{\lambda}x^{\bar{\alpha}}\right)\mu_{\widetilde{x}_i}(x)dx}]^2)\}\end{aligned}$$

$$
\begin{aligned}
H_{12} &= \frac{1}{n}\{-\sum_{i=1}^{n}\frac{\int(2x^{2\bar{\alpha}-1}-\bar{\lambda}x^{3\bar{\alpha}-1})\log x\exp\left(-\bar{\lambda}x^{\bar{\alpha}}\right)\mu_{\widetilde{x}_i}(x)dx}{\int x^{\bar{\alpha}-1}\exp\left(-\bar{\lambda}x^{\bar{\alpha}}\right)\mu_{\widetilde{x}_i}(x)dx} \\
&\quad+\sum_{i=1}^{n}(\frac{\int(1-\bar{\lambda}x^{\bar{\alpha}})x^{\bar{\alpha}-1}\log x\exp\left(-\bar{\lambda}x^{\bar{\alpha}}\right)\mu_{\widetilde{x}_i}(x)dx}{\int x^{\bar{\alpha}-1}\exp\left(-\bar{\lambda}x^{\bar{\alpha}}\right)\mu_{\widetilde{x}_i}(x)dx} \\
&\quad\times\frac{\int x^{2\bar{\alpha}-1}\exp\left(-\bar{\lambda}x^{\bar{\alpha}}\right)\mu_{\widetilde{x}_i}(x)dx}{\int x^{\bar{\alpha}-1}\exp\left(-\bar{\lambda}x^{\bar{\alpha}}\right)\mu_{\widetilde{x}_i}(x)dx})\}
\end{aligned}
$$

Now, following the same arguments with $g(\alpha,\lambda)=\alpha$ and λ, respectively, in $H^*(\alpha,\lambda)$, $\hat{\alpha}_{BT}$ and $\hat{\lambda}_{BT}$ in Equation (23) can then be obtained in a straightforward manner.

6. Method of Moments

It is well-known that the kth moment of the Weibull distribution with pdf (1) is

$$E(X^k)=\lambda^{-\frac{k}{\alpha}}\Gamma(1+\frac{k}{\alpha})$$

where $\Gamma(.)$ is the complete Gamma function.

By equating the first and the second sample moments to the corresponding population moments, the following equations can be used to find the estimates of moment method.

$$\lambda^{-\frac{1}{\alpha}}\Gamma(1+\frac{1}{\alpha})=\frac{1}{n}\sum_{i=1}^{n}E_{\alpha,\lambda}(X\mid\widetilde{x}_i) \tag{24}$$

$$\lambda^{-\frac{2}{\alpha}}\Gamma(1+\frac{2}{\alpha})=\frac{1}{n}\sum_{i=1}^{n}E_{\alpha,\lambda}(X^2\mid\widetilde{x}_i) \tag{25}$$

Since the closed form of the solutions to Eqs. (24) and (25) could not be obtained, an iterative numerical process to obtain the parameter estimates is described as follows:

1. Let the initial estimates of α and λ, say $\alpha^{(0)}$ and $\lambda^{(0)}$ with $h=0$.
2. In the $(h+1)$th iteration, we first compute

$$E_{\alpha^{(h)},\lambda^{(h)}}(X^r\mid\widetilde{x}_i)=\frac{\int x^{\alpha^{(h)}+r-1}\exp\left(-\lambda^{(h)}x^{\alpha^{(h)}}\right)\mu_{\widetilde{x}_i}(x)dx}{\int x^{\alpha^{(h)}-1}\exp\left(-\lambda^{(h)}x^{\alpha^{(h)}}\right)\mu_{\widetilde{x}_i}(x)dx},\qquad r=1,2.$$

3. Based on equations (24) and (25), solve the following equation for α

$$\frac{\left[\sum_{i=1}^{n}E_{\alpha^{(h)},\lambda^{(h)}}(X\mid\widetilde{x}_i)\right]^2}{n\left[\sum_{i=1}^{n}E_{\alpha^{(h)},\lambda^{(h)}}(X^2\mid\widetilde{x}_i)\right]}=\frac{\left[\Gamma(1+\frac{1}{\alpha})\right]^2}{\left[\Gamma(1+\frac{2}{\alpha})\right]}$$

to obtain the solution as $\alpha^{(h+1)}$.

4. The solution for λ, say $\lambda^{(h+1)}$, is obtained through the following equation

$$\lambda^{(h+1)} = \left\{ \frac{n\Gamma(1+(1/\alpha^{(h+1)}))}{\sum_{i=1}^{n} E_{\alpha^{(h)},\lambda^{(h)}}(X \mid \widetilde{x}_i)} \right\}^{\alpha^{(h+1)}}$$

5. Setting $h = h + 1$, repeat steps 2 to 4 until convergence occurs and denote the method of moment estimates as $\hat{\alpha}_M$ and $\hat{\lambda}_M$.

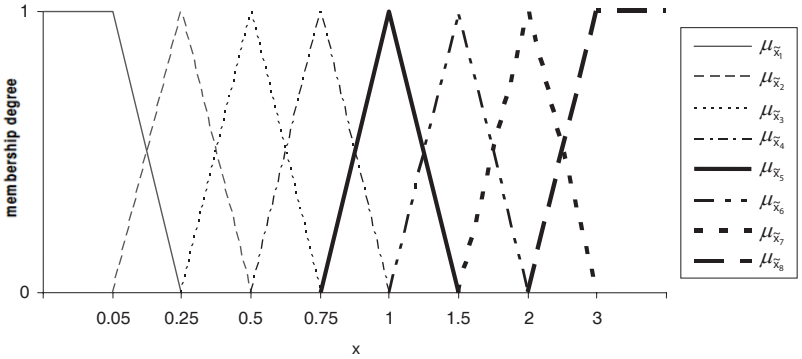


FIGURE 3: Fuzzy information system used to encode the simulated data.

7. Numerical Experiments

7.1. Simulation

In this section, we present some experimental results, mainly to observe how the different methods behave for different sample sizes. We obtain the estimates of the unknown parameters α and λ using the three methods provided in the preceding sections. The computations are performed using R 2.14.0 (R Development Core Team (2011)), which is a non-commercial, open source software package for statistical computing and graphics. First, for different sets of parameter values namely; $(\alpha, \lambda) = (0.5, 1), (1, 1), (2, 1)$, and various choices of n, we have generated i.i.d. random samples, say $\boldsymbol{x}$, from the Weibull distribution. Each realization of $\boldsymbol{x}$ was made fuzzy, using the f.i.s. shown in Fig.3, corresponding to the membership functions

$$\mu_{\widetilde{x}_1}(x) = \begin{cases} 1 & x \leq 0.05, \\ \frac{0.25-x}{0.2} & 0.05 \leq x \leq 0.25, \\ 0 & otherwise, \end{cases} \qquad \mu_{\widetilde{x}_2}(x) = \begin{cases} \frac{x-0.05}{0.2} & 0.05 \leq x \leq 0.25, \\ \frac{0.5-x}{0.25} & 0.25 \leq x \leq 0.5, \\ 0 & otherwise, \end{cases}$$

$$\mu_{\widetilde{x}_3}(x) = \begin{cases} \frac{x-0.25}{0.25} & 0.25 \le x \le 0.5, \\ \frac{0.75-x}{0.25} & 0.5 \le x \le 0.75, \\ 0 & otherwise, \end{cases} \qquad \mu_{\widetilde{x}_4}(x) = \begin{cases} \frac{x-0.5}{0.25} & 0.5 \le x \le 0.75, \\ \frac{1-x}{0.25} & 0.75 \le x \le 1, \\ 0 & otherwise, \end{cases}$$

$$\mu_{\widetilde{x}_5}(x) = \begin{cases} \frac{x-0.75}{0.25} & 0.75 \le x \le 1, \\ \frac{1.5-x}{0.5} & 1 \le x \le 1.5, \\ 0 & otherwise, \end{cases} \qquad \mu_{\widetilde{x}_6}(x) = \begin{cases} \frac{x-1}{0.5} & 1 \le x \le 1.5, \\ \frac{2-x}{0.5} & 1.5 \le x \le 2, \\ 0 & otherwise, \end{cases}$$

$$\mu_{\widetilde{x}_7}(x) = \begin{cases} \frac{x-1.5}{0.5} & 1.5 \le x \le 2, \\ 3-x & 2 \le x \le 3, \\ 0 & otherwise, \end{cases} \qquad \mu_{\widetilde{x}_8}(x) = \begin{cases} x-2 & 2 \le x \le 3, \\ 1 & x \ge 3, \\ 0 & otherwise. \end{cases}$$

Then the estimates of α and λ for the fuzzy sample were computed using the maximum likelihood method (via NR and EM algorithms), the moments method and a Bayesian procedure. For computing the Bayes estimates, we have assumed that λ and α have $Gamma(a, b)$ and $Gamma(c, d)$ priors respectively. To make the comparison meaningful, it is assumed that the priors are non-informative, and they are $a = b = c = d = 0$. Note that in this case the priors are non-proper also. Press (2001) suggested to use very small non-negative values of the hyperparameters in this case, and it will make the priors proper. We have tried $a = b = c = d = 0.0001$. The results are not significantly different than the corresponding results obtained using non-proper priors, and are not reported due to space. From now on, the estimates of parameters obtained by using NR algorithm, EM algorithm, Bayesian procedure and moments method will be denoted by NR, EM, BET and MME, respectively. The average biases (AB) and mean squared errors (MSE) of the estimates over $5,000$ replications are presented in Tables 1-2.

From the experiments, we found that using the NR or EM algorithm for the computation of maximum likelihood estimates of α and λ give similar estimation results, but EM is computationally slower. For small and moderate sample sizes, the Bayesian procedure gives the most precise parameter estimates as shown by ABs and MSEs in Tables 1-2. For large sample sizes ($n = 100, 200$ and 500), the performance of the MLEs, MMEs and Bayes estimates are almost identical. For all the methods, it is observed that as the sample size increases, the biases and MSEs of the estimates decrease as expected.

7.2. Application example

In order to demonstrate the application of proposed methods, let us consider a case study on the light emitting diodes (LED) manufacturing process that focuses on the luminous intensities of LED sources. The process distribution has been justified and has been shown to be fairly close to the Weibull distribution. A sample of size $n = 30$ is taken from the stable process. Since the data given by

TABLE 1: MSE of the estimates of α and λ for different sample sizes.

n	α	λ	Estimation of α				Estimation of λ			
			NR	EM	BET	MME	NR	EM	BET	MME
15	0.5	1	0.0619	0.0620	0.0594	0.0705	0.0870	0.0871	0.0836	0.092
	1	1	0.0987	0.0988	0.0830	0.1246	0.1129	0.1130	0.1091	0.1191
	2	1	0.1263	0.1264	0.1129	0.1380	0.1465	0.1465	0.1421	0.1483
20	0.5	1	0.0558	0.0559	0.0512	0.0631	0.0727	0.0728	0.0639	0.0792
	1	1	0.0942	0.0943	0.0744	0.1193	0.1088	0.1089	0.0966	0.1139
	2	1	0.1017	0.1018	0.0922	0.1240	0.1226	0.1227	0.1182	0.1259
30	0.5	1	0.0366	0.0367	0.0341	0.0394	0.0489	0.0489	0.0422	0.0519
	1	1	0.0614	0.0614	0.0488	0.0828	0.0691	0.0692	0.0646	0.0707
	2	1	0.0721	0.0722	0.0630	0.0895	0.0843	0.0844	0.0819	0.0895
50	0.5	1	0.0285	0.0286	0.0257	0.0335	0.0365	0.0365	0.0342	0.0386
	1	1	0.0361	0.0362	0.0331	0.0451	0.0427	0.0427	0.0419	0.0430
	2	1	0.0488	0.0489	0.0425	0.0536	0.0572	0.0572	0.0558	0.0587
70	0.5	1	0.0214	0.0215	0.0208	0.0232	0.0305	0.0306	0.0291	0.0318
	1	1	0.0282	0.0282	0.0225	0.0346	0.0338	0.0339	0.0328	0.0345
	2	1	0.0327	0.0328	0.0311	0.0387	0.0478	0.0478	0.0460	0.0491
100	0.5	1	0.0154	0.0154	0.0152	0.0156	0.0227	0.0228	0.0220	0.0236
	1	1	0.0191	0.0192	0.0187	0.0195	0.0284	0.0285	0.0282	0.0289
	2	1	0.0270	0.0270	0.0263	0.0271	0.0395	0.0395	0.0390	0.0397
200	0.5	1	0.0104	0.0104	0.0098	0.0109	0.0174	0.0175	0.0168	0.0179
	1	1	0.0127	0.0128	0.0120	0.0134	0.0211	0.0211	0.0202	0.0218
	2	1	0.0214	0.0214	0.0209	0.0225	0.0356	0.0356	0.0348	0.0360
500	0.5	1	0.0055	0.0055	0.0051	0.0058	0.0118	0.0118	0.0113	0.0122
	1	1	0.0086	0.0086	0.0085	0.0088	0.0173	0.0174	0.0161	0.0179
	2	1	0.0142	0.0142	0.0139	0.0153	0.0235	0.0235	0.0230	0.0238

luminous intensity of a particular LED inevitably have some degree of imprecision, the luminous intensities of diodes are reported in the form of lower and upper bounds as well as a point estimate, which are as follows:

DATA SET:

$$(2.163, 2.738, 3.068), (5.972, 6.353, 8.150), (1.032, 1.971, 2.642),$$
$$(0.628, 0.964, 1.735), (2.995, 3.442, 5.066), (3.766, 5.814, 6.212),$$
$$(0.974, 1.839, 2.045), (4.352, 5.206, 5.988), (3.920, 4.762, 6.121),$$
$$(1.375, 2.195, 3.086), (0.618, 0.839, 2.217), (4.575, 6.050, 6.734),$$
$$(1.027, 1.218, 3.116), (6.279, 8.156, 9.435), (2.821, 3.409, 5.272),$$
$$(7.125, 8.470, 9.044), (5.443, 6.231, 7.395), (1.766, 2.190, 2.638),$$
$$(7.155, 8.013, 8.352), (0.830, 1.288, 2.541), (3.590, 4.169, 4.899),$$
$$(5.965, 7.344, 8.019), (3.177, 3.600, 4.213), (4.634, 5.780, 7.058),$$
$$(7.261, 8.325, 8.871), (2.247, 2.990, 4.128), (6.032, 7.746, 8.529),$$
$$(4.065, 5.312, 7.480), (5.434, 7.093, 7.655), (1.336, 2.750, 3.284).$$

In our approach, each triplet is modeled by a triangular fuzzy number $\widetilde{x}_i$, and is interpreted as a possibility distribution related to an unknown value x_i, itself a realization of a random variable X_i. For this data, we employ NR and

Table 2: AB of the estimates of α and λ for different sample sizes.

n	α	λ	Estimation of α				Estimation of λ			
			NR	EM	BET	MME	NR	EM	BET	MME
15	0.5	1	0.1272	0.1273	0.0734	0.1533	0.1180	0.1181	0.1092	0.1230
	1	1	0.1381	0.1382	-0.0783	0.1698	0.1291	0.1292	0.1262	0.1322
2		1	0.1914	0.1915	0.1527	0.2038	0.1570	0.1571	0.1503	0.1637
20	0.5	1	0.1091	0.1092	-0.0617	0.1326	0.0931	0.0931	0.0865	0.1026
	1	1	0.1354	0.1355	-0.0699	0.1633	0.1205	0.1206	0.1177	0.1298
	2	1	0.1775	0.1775	0.1344	0.1851	0.1427	0.1428	0.1321	0.1485
30	0.5	1	0.0922	0.0923	0.0591	0.1130	0.0778	0.0779	0.0631	0.0840
	1	1	0.1228	0.1228	-0.0621	0.1417	0.1086	0.1087	0.1059	0.1152
	2	1	0.1439	0.1439	0.1223	0.1507	0.1162	0.1163	0.1137	0.1218
50	0.5	1	0.0754	0.0755	0.0518	0.0908	0.0620	0.0621	0.0582	0.0685
	1	1	0.0917	0.0918	-0.0571	0.1275	0.0927	0.0927	0.0905	0.0996
	2	1	0.1254	0.1255	0.1033	0.1445	0.1067	0.1067	0.1013	0.1151
70	0.5	1	0.0628	0.0629	0.0435	0.0711	0.0514	0.0514	0.0507	0.0536
	1	1	0.0887	0.0887	0.0494	0.1065	0.0833	0.0834	0.0821	0.0875
	2	1	0.1057	0.1058	-0.0932	0.1126	0.0983	0.0983	0.0970	0.0994
100	0.5	1	0.0413	0.0413	0.0408	0.0419	0.0459	0.0459	0.0455	0.0463
	1	1	0.0438	0.0438	0.0426	0.0440	0.0648	0.0648	0.0642	0.0655
	2	1	0.0906	0.0907	0.0896	0.0918	0.0952	0.0952	0.0948	0.0961
200	0.5	1	0.0287	0.0288	0.0281	0.0290	0.0317	0.0318	0.0314	0.0318
	1	1	0.0349	0.0349	0.0345	0.0353	0.0573	0.0573	0.0570	0.0574
	2	1	0.0855	0.0856	0.0851	0.0859	0.0736	0.0737	0.0733	0.0738
500	0.5	1	0.0211	0.0212	0.0207	0.0225	0.0244	0.0244	0.0241	0.0245
	1	1	0.0267	0.0268	0.0260	0.0271	0.0408	0.0409	0.0404	0.0412
	2	1	0.0762	0.0762	0.0758	0.0766	0.0553	0.0554	0.0550	0.0557

EM algorithms to compute the ML estimates. The stopping criterion is based on the difference between the two consecutive iterates, with a tolerance value $\varepsilon = 10^{-6}$. The final MLEs are $(\hat{\alpha}_{NR}, \hat{\lambda}_{NR}) = (2.1094, 0.0318)$ and $(\hat{\alpha}_{EM}, \hat{\lambda}_{EM}) = (2.1095, 0.0319)$. Also, by using the procedure presented in section 6, the moment estimate of (α, λ) becomes $(\hat{\alpha}_M, \hat{\lambda}_M) = (2.1257, 0.0374)$. For computing the Bayes estimate, we assume that both α and λ have a $Gamma(0.0001, 0.0001)$ prior. Therefore, using the Tierney and Kadane's approximation, the Bayes estimate of the parameters becomes $(\hat{\alpha}_{BT}, \hat{\lambda}_{BT}) = (2.1036, 0.0287)$.

8. Conclusions

Some work has been done in the past on the estimation of Weibull distribution parameters based on complete and censored samples. But, traditionally it is assumed that the available data are performed in exact numbers. However, some collected data might be imprecise and are represented in the form of fuzzy numbers. Therefore, we need suitable statistical methodology to handle these data as well. In this paper, we have discussed different estimation procedures for the Weibull distribution when the obtained data are fuzzy numbers. They include the maximum likelihood method (via NR and EM algorithms), a Bayesian procedure and the method of moments. We have then carried out a simulation study to assess

the performance of all these procedures. The recommendations of an estimator based on minimum biases and MSEs are as follows:

i) For small and moderate sample sizes, the performance of the Bayes estimates is generally best followed by the MLEs and then the MMEs. Thus, it would seem reasonable to recommend the use of the Bayesian procedure for estimating the unknown parameters α and λ.

ii) For large sample sizes, the three estimation procedures behave in similar manner.

Acknowledgments

The authors are thankful to the referees for their valuable comments which led to a considerable improvement in the presentation of this article.

[Recibido: febrero de 2013 — Aceptado: septiembre de 2013]

References

Ageel, M. I. (2002), 'A novel means of estimating quantiles for 2-parameter Weibull distribution under the right random censoring model', *Journal of Computational and Applied Mathematics* **149**(2), 373–380.

Al-Baidhani, P. A. & Sinclair, C. (1987), 'Comparison of methods of estimation of parameters of the Weibull distribution', *Communications in Statistics-Simulation and Computation* **16**(2), 373–384.

Balakrishnan, N. & Kateri, M. (2008), 'On the maximum likelihood estimation of parameters of Weibull distribution based on complete and censored data', *Statistics and Probability Letters* **78**(17), 2971–2975.

Balakrishnan, N. & Mitra, D. (2012), 'Left truncated and right censored Weibull data and likelihood inference with an illustration', *Computational Statistics and Data Analysis* **56**, 4011–4025.

Banerjee, A. & Kundu, D. (2012), 'Inference based on type-II hybrid censored data from a Weibull distribution', *IEEE Transactions on Reliability* **57**(2), 369–378.

Denoeux, T. (2011), 'Maximum likelihood estimation from fuzzy data using the EM algorithm', *Fuzzy Sets and Systems* **183**(1), 72–91.

Dubois, D. & Prade, H. (1980), *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York.

Gertner, G. Z. & Zhu, H. (1996), 'Bayesian estimation in forest surveys when samples or prior information are fuzzy', *Fuzzy Sets and Systems* **77**, 277–290.

Helu, A., Abu-Salih, M. & Alkam, O. (2010), 'Bayes estimation of Weibull distribution parameters using ranked set sampling', *Communications in Statistics-Theory and Methods* **39**(14), 2533–2551.

Joarder, A., Krishna, H. & Kundu, D. (2011), 'Inferences on Weibull parameters with conventional type-I censoring', *Computational Statistics and Data Analysis* **55**, 1–11.

Lin, C., Chou, C. & Huang, Y. (2012), 'Inference for the Weibull distribution with progressive hybrid censoring', *Computational Statistics and Data Analysis* **56**, 451–467.

Marks, N. B. (2005), 'Estimation of Weibull parameters from common percentiles', *Journal of Applied Statistics* **32**(1), 17–24.

Nandi, S. & Dewan, I. (2010), 'An EM algorithm for estimating the parameters of bivariate Weibull distribution under random censoring', *Computational Statistics and Data Analysis* **54**(6), 1559–1569.

Ng, H. K. T. & Wang, Z. (2009), 'Statistical estimation for the parameters of Weibull distribution based on progressively type-I interval censored sample', *Journal of Statistical Computation and Simulation* **79**(2), 145–159.

Press, S. J. (2001), *The Subjectivity of Scientists and the Bayesian Approach*, Wiley, New York.

Qiao, O. & Tsokos, C. P. (1994), 'Parameter estimation of the Weibull probability distribution', *Mathematics and Computers in Simulation* **37**, 47–55.

R Development Core Team (2011), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
*http://www.R-project.org

Singpurwalla, N. D. & Booker, J. M. (2004), 'Membership functions and probability measures of fuzzy sets', *Journal of the American Statistical Association* **99**(467), 867–877.

Tan, Z. (2009), 'A new approach to MLE of Weibull distribution with interval data', *Reliability Engineering and System Safety* **94**(2), 394–403.

Tierney, L. & Kadane, J. B. (1986), 'Accurate approximations for posterior moments and marginal densities', *Journal of the American Statistical Association* **81**, 82–86.

Watkins, A. J. (1994), 'On maximum likelihood estimation for the two parameter Weibull distribution', *Microelectronics Reliability* **36**(5), 595–603.

Zadeh, L. A. (1968), 'Probability measures of fuzzy events', *Journal of Mathematical Analysis and Applications* **10**, 421–427.

Letter to the Editor

Discussion about the Paper "On the Moment Characteristics for the Univariate Compound Poisson and Bivariate Compound Poisson Process with Applications"

Masood Anwar[a]

Department of Mathematics, COMSATS Institute of Information Technology, Islamabad, Pakistan

---

Dear editor,

I have found some few results that were incorrect in a paper recently published by Özel (2013). Please find them attached.

1. In Section 2.1, the given common moment generating function (mgf) of X_i, $i = 1, 2, \ldots, M_X(u) = \sum_{j=0}^{\infty} u^j p_j$ is actually the probability generating function (pgf) of X_i. With this definition of $M_X(u)$, a recursive formula (on page 62) for the general moments of $\{S_t, t \geq 0\}$ is incorrect because $\frac{d^{r-k} M_X(u)}{du^{r-k}} \mid_{u=0} \neq E(X^{r-k})$, where $E(X^{r-k}) = \xi_{r-k}$. Therefore, the common mgf of X_i, $i = 1, 2, \ldots$, should be replaced by $M_X(e^u) = \sum_{j=0}^{\infty} e^{uj} p_j$ and the mgf of $\{S_t, \geq 0\}$ with $M_{S_t}(u) = \exp(\lambda t[M_X(e^u) - 1])$.

2. In Section 2.1, on page 62, in the first line, "factorial" should be replaced by "general (raw)".

3. In Section 2.1, the central moments of $\{S_t, \geq 0\}$ obtained in (8) are incorrect as the first central moment is always equal to zero i.e $\mu_1 = 0$. We derive the corrected version of central moments as follows;

$$\begin{aligned}
\mu_1 &= 0 \\
\mu_2 &= \lambda t \xi_2 \\
\mu_3 &= \lambda t \xi_3 \\
\mu_4 &= 3(\lambda t \xi_2)^2 + \lambda t \xi_4 \\
\mu_5 &= 10(\lambda t \xi_2)(\lambda t \xi_3) + \lambda t \xi_5
\end{aligned}$$

   where $\xi_r = E(x^r)$, $r = 1, 2, \ldots, n$, is the rth general moment of X_i, $i = 1, 2, \ldots$.

4. In Section 2.1, the skewness (10) of S_t should be considered as

$$\sqrt{\beta_1} = \frac{\xi_3}{\sqrt{\lambda t}(\xi_2)^{3/2}}$$

[a]Professor. E-mail: masoodanwar@comsats.edu.pk

5. In Section 2.1, the expression for kurtosis (11) is $\beta_2 - 3 = \dfrac{\xi_4}{\lambda t(\xi_2)^2}$.

6. In Section 2.1, in Equation (14), the statement $\left(\lambda t \sum_{j=1}^{\infty} r^j p_j\right)$ should be replaced by $\left(\lambda t \sum_{j=1}^{\infty} j^r p_j\right)$.

7. In Section 2.1, after Equation (20), "From (21)" should be replaced by "From (20)" and the expression "ξ_r" should be changed with "$\xi_{[r]}$".

8. In Section 2.3, in Example 1, the four central moments of Neyman type A or Poisson-Poisson process given in Table 1 (Neyman 1939, Özel & Inal 2012) should be

$$\begin{aligned}
\mu_1 &= 0 \\
\mu_2 &= [\lambda t(\upsilon + \upsilon^2)] \\
\mu_3 &= [\lambda t(\upsilon + \upsilon^2 + \upsilon^3)] \\
\mu_4 &= [3(\lambda t)^2(\upsilon^2 + 2\upsilon^3 + \upsilon^4)] + [\lambda t(\upsilon + 7\upsilon^2 + 6\upsilon^3 + \upsilon^4)]
\end{aligned}$$

9. In Section 2.3, in Example 2, the four central moments of Neyman type B or Poisson-binomial process given in Table 2 (Neyman 1939, Özel & Inal 2012) are actually

$$\begin{aligned}
\mu_1 &= 0 \\
\mu_2 &= [\lambda t(mp + m(m-1)p^2)] \\
\mu_3 &= [\lambda t(mp + 3m(m-1)p^2 + m(m-1)(m-2)p^3)] \\
\mu_4 &= 3[\lambda t(mp + 3m(m-1)p^2)]^2 + \left[\lambda t(mp + 7m(m-1)p^2)\right. \\
&\quad \left. + 6m(m-1)(m-2)p^3 + m(m-1)(m-2)(m-3)p^4\right]
\end{aligned}$$

10. Similarly, in Example 3, the four central moments of Pólya-Aeppli or geometric Poisson process given in Table 3 (Getis 1974, Özel & Inal 2012) are actually as follows;

$$\begin{aligned}
\mu_1 &= 0 \\
\mu_2 &= [\lambda t(1-\theta)(2-\theta)/\theta^2] \\
\mu_3 &= [\lambda t(1-\theta)(6+\theta(\theta-6))/\theta^3] \\
\mu_4 &= 3[\lambda t(1-\theta)(2-\theta)/\theta^2]^2 + [\lambda t(1-\theta)(2-\theta)(12+\theta(\theta-12))/\theta^4]
\end{aligned}$$

11. In Section 4, the central moments for the Pólya-Aeppli process presented in Table 4 for the parameters $\lambda = 9.84$, $\theta = 0.62$ and values of t are incorrect. The Table 1 should be considered for the central moments. Similarly, the results of skewness and kurtosis computed for the Pólya-Aeppli process in Table 5 should be replaced with Table 2.

TABLE 1: The corrected central moments of the Pólya-Aeppli process for the traffic accidents in Groningen (Meintanis, 1997, Özel & Inal, 2010).

t	μ_1	μ_2	μ_3	μ_4
0.5	0	6.71	20.90	221.40
1	0	19.45	41.80	1307.58
2	0	38.91	83.61	4887.29
3	0	58.36	125.41	10735.67
4	0	77.82	167.21	18858.52

TABLE 2: The skewness and kurtosis of Pólya-Aeppli process for the traffic accidents in Groningen (Meintanis, 1997, Özel & Inal, 2010).

T	$\sqrt{\beta_1}$	$\beta_2 - 3$
0.5	1.20	1.92
1	0.49	0.45
2	0.34	0.23
3	0.28	0.15
4	0.24	0.11

A second dataset illustrates the usage of bivariate compound Poisson processes and it comes from earthquakes in Turkey, which is given by Özel (2011*a*) y Özel (2011*b*).

References

Getis, A. (1974), Representation of spatial point pattern processes by Polya models, *in* M. Yeates, ed., 'Proceedings of the 1972 Meeting of the IGU Commission on Quantitative Geography, McGill-Queens', University Press, Montreal, pp. 76–100.

Meintanis, S. G. (1997), 'A new goodness of fit test for certain bivariate distributions applicable to traffic accidents', *Statistical Methodology* **1**(4), 22–34.

Neyman, J. (1939), 'On a new class of contagious distributions applicable in entomology and bacteriology', *Annals of Mathematical Statistics* **10**(1), 35–57.

Özel, G. (2011*a*), 'A bivariate compound Poisson model for the occurrence of foreshock and aftershock sequences in Turkey', *Environmetrics* **22**(7), 847–856.

Özel, G. (2011*b*), 'On certain properties of a class of bivariate compound Poisson distributions and an application to earthquake data', *Revista Colombiana de Estadistica* **34**(3), 545–566.

Özel, G. (2013), 'On the moment characteristics for the univariate compound Poisson and bivariate compound Poisson processes with applications', *Revista Colombiana de Estadistica* **36**(1), 59–77.

Özel, G. & Inal, C. (2010), 'The probability function of a geometric Poisson distribution', *Journal of Statistical Computation and Simulation* **80**(5), 479–487.

Özel, G. & Inal, C. (2012), 'On the probability function of the first exit time for generalized Poisson processes', *Pakistan Journal of Statistics* **28**(1), 27–40.

Información para los autores

La **Revista Colombiana de Estadística** publica artículos originales de carácter teórico o aplicado en cualquiera de las ramas de la estadística. Los artículos puramente teóricos deberán incluir la ilustración de las técnicas presentadas con datos reales o por lo menos con experimentos de simulación, que permitan verificar la utilidad de los contenidos presentados. Se consideran también artículos divulgativos de gran calidad de exposición sobre metodologías o técnicas estadísticas aplicadas en diferentes campos del saber. Únicamente se publican artículos en español e inglés, si el autor escribe en una lengua diferente a la nativa debe enviar un certificado de un traductor oficial o de un corrector de estilo que haya revisado el texto.

El Comité Editor únicamente acepta trabajos para evaluación que no han sido publicados previamente y que no están siendo propuestos simultáneamente para publicación en otros medios, ni lo serán sin previo consentimiento del Comité, a menos que, como resultado de la evaluación, se decida no publicarlos en la Revista. Se supone además que cuando los autores hacen entrega de un documento con fines de publicación en la **Revista Colombiana de Estadística**, conocen las condiciones anteriores y que están de acuerdo con ellas.

Material

Los artículos remitidos a la **Revista Colombiana de Estadística** deben ser presentados en archivo PDF o PS, con textos, gráficas y tablas en color negro y, además, los autores deben agregar una versión del artículo sin nombres ni información de los autores, que se utilizará para el arbitraje. Se debe enviar una carta firmada por cada uno de los autores, donde manifiesten estar de acuerdo con someter el artículo y con las condiciones de la Revista. Si un artículo es aceptado, los autores deben poner a disposición del Comité Editorial los archivos: fuente en LaTeX y de gráficas en formato EPS en blanco y negro.

Para facilitar la preparación del material publicado se recomienda utilizar MiKTeX[1], usando los archivos de la plantilla y del estilo *revcoles* disponibles en la página Web de la Revista[2] y siguiendo las instrucciones allí incorporadas.

Todo artículo debe incluir:

- Título en español y su traducción al inglés.
- Los nombres completos y el primer apellido, la dirección postal o electrónica y la afiliación institucional de cada autor.
- Un resumen con su versión en inglés (*abstract*). El resumen en español no debe pasar de 200 palabras y su contenido debe destacar el aporte del trabajo en el tema tratado.

[1] http://www.ctan.org/tex-archive/systems/win32/miktex/
[2] http://www.estadistica.unal.edu.co/revista

- Palabras clave (*Key words*) en número entre 3 y 6, con su respectiva traducción al inglés, siguiendo las recomendaciones del *Current Index to Statistics* (CIS)[3].
- Cuando el artículo se deriva de una tesis o trabajo de grado debe indicarse e incluirse como una referencia.
- Si se deriva de un proyecto de investigación, se debe indicar el título del proyecto y la entidad que lo patrocina.
- Referencias bibliográficas, incluyendo solamente las que se hayan citado en el texto.

Referencias y notas al pie de página

Para las referencias bibliográficas dentro del texto se debe utilizar el formato autor-año, dando el nombre del autor seguido por el año de la publicación dentro de un paréntesis. La plantilla LaTeX suministrada utiliza, para las referencias, los paquetes BibTeX y Harvard[4]. Se recomienda reducir el número de notas de pie de página, especialmente las que hacen referencia a otras notas dentro del mismo documento y no utilizarlas para hacer referencias bibliográficas.

Tablas y gráficas

Las tablas y las gráficas, con numeración arábiga, deben aparecer referenciadas dentro del texto mediante el número correspondiente. Las tablas deben ser diseñadas en forma que se facilite su presentación dentro del área de impresión de la Revista. En este sentido, los autores deben considerar en particular la extensión de las tablas, los dígitos representativos, los títulos y los encabezados. Las gráficas deben ser visualmente claras y debe ser posible modificar su tamaño. Cuando el artículo sea aceptado para su publicación, los autores deben poner la versión definitiva a disposición del Comité Editorial. Todos los elementos como barras, segmentos, palabras, símbolos y números deben estar impresos en color negro.

Responsabilidad legal

Los autores se hacen responsables por el uso de material con propiedad intelectual registrada como figuras, tablas, fotografías, etc.

Arbitraje

Los artículos recibidos serán revisados por el Comité Editorial y sometidos a arbitraje por pares especializados en el tema respectivo. El arbitraje es "doble ciego" (árbitros anónimos para los autores y viceversa). El Comité Editorial decide aceptar, rechazar o solicitar modificaciones a los artículos con base en las recomendaciones de los árbitros.

[3] http://www.statindex.org/CIS/homepage/keywords.html
[4] http://tug.ctan.org/tex-archive/macros/latex/contrib/harvard

Revista Colombiana de Estadística

Índice de autores del volumen 36, 2013