Austrian Journal of Statistics

AUSTRIAN STATISTICAL SOCIETY

Volume 44, Issue 4, 2015

ISSN: 1026597X, Vienna, Austria

Österreichische Zeitschrift für Statistik

ÖSTERREICHISCHE STATISTISCHE GESELLSCHAFT

Austrian Journal of Statistics; Information and Instructions

GENERAL NOTES

The Austrian Journal of Statistics is an open-access journal with a long history and is published approximately quarterly by the Austrian Statistical Society. Its general objective is to promote and extend the use of statistical methods in all kind of theoretical and applied disciplines. Special emphasis is on methods and results in official statistics.

Original papers and review articles in English will be published in the Austrian Journal of Statistics if judged consistently with these general aims. All papers will be refereed. Special topics sections will appear from time to time. Each section will have as a theme a specialized area of statistical application, theory, or methodology. Technical notes or problems for considerations under Shorter Communications are also invited. A special section is reserved for book reviews.

All published manuscripts are available at

http://www.ajs.or.at

(old editions can be found at http://www.stat.tugraz.at/AJS/Editions.html)

Members of the Austrian Statistical Society receive a copy of the Journal free of charge. To apply for a membership, see the website of the Society. Articles will also be made available through the web.

PEER REVIEW PROCESS

All contributions will be anonymously refereed which is also for the authors in order to getting positive feedback and constructive suggestions from other qualified people. Editor and referees must trust that the contribution has not been submitted for publication at the same time at another place. It is fair that the submitting author notifies if an earlier version has already been submitted somewhere before. Manuscripts stay with the publisher and referees. The refereeing and publishing in the Austrian Journal of Statistics is free of charge. The publisher, the Austrian Statistical Society requires a grant of copyright from authors in order to effectively publish and distribute this journal worldwide.

OPEN ACCESS POLICY

This journal provides immediate open access to its content on the principle that making research freely available to the public supports a greater global exchange of knowledge.

ONLINE SUBMISSIONS

Already have a Username/Password for Austrian Journal of Statistics?
Go to http://www.ajs.or.at/index.php/ajs/login

Need a Username/Password?
Go to http://www.ajs.or.at/index.php/ajs/user/register

Registration and login are required to submit items and to check the status of current submissions.

AUTHOR GUIDELINES

The original LaTeX-file guidelinesAJS.zip (available online) should be used as a template for the setting up of a text to be submitted in computer readable form. Other formats are only accepted rarely.

SUBMISSION PREPARATION CHECKLIST

- The submission has not been previously published, nor is it before another journal for consideration (or an explanation has been provided in Comments to the Editor).
- The submission file is preferable in LaTeXfile format provided by the journal.
- All illustrations, figures, and tables are placed within the text at the appropriate points, rather than at the end.
- The text adheres to the stylistic and bibliographic requirements outlined in the Author Guidelines, which is found in About the Journal.

COPYRIGHT NOTICE

Manuscripts should be unpublished and not be under consideration for publication elsewhere. By submitting an article, the author(s) certify that the article is their original work, that they have the right to submit the article for publication, and that they can grant the above license.

Austrian Journal of Statistics

Volume 44, Number 4, 2015

Editor-in-chief: Matthias TEMPL

http://www.ajs.or.at

Published by the AUSTRIAN STATISTICAL SOCIETY

http://www.osg.or.at

Österreichische Zeitschrift für Statistik

Jahrgang 44, Heft 4, 2015

ÖSTERREICHISCHE STATISTISCHE GESELLSCHAFT

Impressum

Printed by Statistics Austria, A-1110 Vienna

Published approximately quarterly by the Austrian Statistical Society, C/o Statistik Austria
Guglgasse 13, A–1110 Wien

Contents

Editorial

The Rayleigh distribution is especially used in reliability theory and survival analysis. The first paper of this issue extends exising methodology for progressive type-II censored data, considering the removal of surviving experimental units before the termination of a test.

In the second contribution a restricted graphical log-linear model – a special case of hierarchical log-linear models – is extended by setting equality restrictions not only on the main effects and first-order interactions, but also on higher-order interactions.

In 1958, Lindley proposed a distribution which is named after himself. A lot of work has been done to show the properties of this distribution, and modifications of these distributions were published in recent years. In the third paper, it is shown that the proposed generalized Poisson-Lindley distribution can be considered as an alternative to well known distributions such as the Poisson, the Poisson-Lindley, the negative binomial, the weighted generalized Poisson distribution, and in general to model highly skewed data. This new distribution can be applied in applications of actuarial sciences or in the determination of bonus–malus benefits for non-life insurance products, for example.

Response adaptive allocation is a useful technique in clinical trials for comparing two treatments, which is the content of the fourth paper. Under the presence of treatment covariate interaction, several exact and asymptotic properties are studied and compared with a reasonable alternative hypothesis.

The picture on the cover page is dedicated to Rudolf Dutter. It shows a sign on a hiking trail where he went often. Rudolf Dutter retired in October 2015. He was a pioneer in computational statistics who brought a new thinking to the statistics department at the Vienna University of Technology – the research and method development in statistics with support of computers and software. Because of him, several Austrian researchers got in touch with the important field of Robust Statistics. His career, but also his personality that postively influenced many students and colleagues, is reflected in a Laudatio by his close collaborator Peter Filzmoser.

Matthias Templ
(Editor-in-Chief)

Statistics Austria & Vienna University of Technology
Wiedner Hauptstr. 8–10
A–1040 Vienna, Austria
E-mail: matthias.templ@gmail.com

Vienna, December 2015

Austrian Journal of Statistics
December 2015, Volume 44, 3–15.
http://www.ajs.or.at/
doi:10.17713/ajs.v44i4.71

Bayes Shrinkage Estimation of the Parameter of Rayleigh Distribution for Progressive Type-II Censored Data

Sanku Dey
St. Anthony's College

Tanujit Dey
Cleveland Clinic
(corresponding author)

Sudhansu S. Maiti
Visva-Bharati University

Abstract

This paper derives Bayes shrinkage estimator of Rayleigh parameter and its associated risk function based on conjugate prior under the assumption of general entropy loss function for progressive Type-II censored data. Risk function of maximum likelihood estimate, Bayes estimate and Bayes shrinkage estimate have also been derived and compared. An empirical Bayes estimate procedure has been suggested to include a guess value in case of the Bayes shrinkage estimation. Risk function of empirical Bayes estimate and empirical Bayes shrinkage estimate have also been derived and compared. In conclusion, an illustrative example is presented to assess how the Rayleigh distribution fits a real data set.

Keywords: Bayes estimate, conjugate prior, empirical Bayes estimate, Rayleigh distribution, risk function.

1. Introduction

Rayleigh distribution first introduced in the literature by Lord Rayleigh (1980) has been widely used in reliability theory and survival analysis because of it's flexibility and simplicity. An important characteristic of the Rayleigh distribution is that its failure rate is a linear function of time. The reliability function of Rayleigh distribution decreases at a much higher rate than the reliability function of exponential distribution. The probability density function of one parameter Rayleigh distribution has the form:

$$f(x \mid \theta) = \frac{x}{\theta} exp\left(-\frac{x^2}{2\theta}\right); \quad x > 0, \ \theta > 0. \tag{1}$$

and the cumulative density function (cdf) has the form:

$$F(x \mid \theta) = 1 - exp\left(-\frac{x^2}{2\theta}\right); \quad x > 0, \ \theta > 0. \tag{2}$$

In recent years, several authors have carried out extensive studies in relation to the estimation, prediction and other inferences with respect to Rayleigh distribution. Sinha and Howlader

(1983) obtained credible and highest posterior density (HPD) intervals of the parameter and reliability of Rayleigh distribution. Ariyawansa and Templeton (1984) have also discussed some of its applications. Howlader (1985) obtained HPD prediction intervals for the Rayleigh distribution. Howlader and Hossain (1995) obtained Bayes estimators for the scale parameter and the reliability function in the case of Type-II censored sampling. Abd Elfattah, Hassan, and Ziedan (2006) studied the efficiency of the maximum likelihood estimates of the parameter under three cases, namely, Type-I, Type-II and progressive Type-II censored sampling schemes. Wu, Chen, and Chen (2006) obtained Bayes estimators and highest posterior density credible intervals for parameter and reliability function of the Rayleigh distribution, as well as the Bayes predictive estimator and prediction interval for future observations based on progressively Type-II censored samples. Hendi, Abu-Youssef, and Alraddadi (2007) obtained Bayes estimators of the scale parameter, reliability function and failure rate by using non-informative prior and Hartigan prior based on upper record values. Dey and Das (2007) obtained Bayesian predictive intervals of the parameter of Rayleigh distribution. Dey (2009) also obtained Bayes estimators for the parameter and reliability function of the Rayleigh distribution under different loss function. Dey (2009) also studied the Bayes estimators for the parameter and reliability function of Rayleigh distribution based on complete as well as Type-II censored samples, also compared relative risk functions.

Our key role is to obtain Bayes shrinkage estimator for the parameter of Rayleigh distribution, which is different than the approaches referenced above. The shrinkage estimator are valuable as in many practical situations, the experimenter has some prior information about the parameter value in the form of a point guess value and that value can be used to make inference of the parameter. In this condition, our parameter of interest is θ and thereby the guess value θ_0 can be used to make inference for θ. In this article, we use an empirical Bayes estimation procedure for θ_0 based on sample observation.

Several authors have considered the use of the point guess value for inferences with regard to the parameter. For instance; Prakash and Singh (2006) studied shrinkage estimators for the inverse dispersion for inverse Gaussian distribution under LINEX loss function. Singh, Prakash, and Singh (2007) studied shrinkage estimators for the shape parameter of Pareto distribution using the LINEX loss function. Singh, Singh, Singh, and Upadhyay (2008) studied the Bayes estimators of the failure rate and reliability function for a one-parameter exponential distribution by utilizing a point guess estimate of the parameter. Prakash (2009) obtained some shrinkage estimators and the Bayes estimators for the shape parameter of a Pareto distribution under the general entropy loss function. Al-Hemyari and Al-Dabag (2012) studied a class of shrinkage estimators for the shape parameter of the Weibull lifetime model. Salman and Shareef (2014) studied preliminary test single stage Bayesian shrinkage estimator for the scale parameter of an exponential distribution under the improper prior distribution using the quadratic loss function.

Perhaps the most popular technique that utilizes the point guess value is the shrinkage technique, originally suggested by Thompson (1968). The shrinkage estimator performs better than the usual estimator when a guess value is approximately the true value of the parameter given the sample size is small. Thompson (1968) considered the problem of shrinking an unbiased estimator $\hat{\zeta}$ of the parameter ζ toward a natural origin ζ_0 and suggested a shrinkage type estimator $k\hat{\zeta} + (1-k)\zeta_0$, where k $(0 < k < 1)$is a constant.

The shrinkage technique has been utilized in numerous studies namely; mean survival time in epidemiological studies (Harris and Shakarki 1979), projecting the money supply (Tso 1990), estimating mortality rates (Marshall 1991) and improving estimation in sample surveys (Wooff 1985). In life-testing and reliability experiments, units that are subject to test are sometimes lost or removed from the experiment before failure. Such units are usually called the censored units.The most common censoring schemes are Type-I censoring and Type-II censoring but one of the drawbacks of the conventional Type-I and Type-II censoring schemes is that they are inflexible in removing units at that point of execution but rather at the end of the experiment. One censoring scheme known as progressive Type-II censoring scheme overcomes this shortcoming and this has led to its popularity in recent years.

1.1. Review of progressive Type-II sampling

Progressive censoring is useful in both industrial life testing applications and clinical settings; it allows the removal of surviving experimental units before the termination of the test. Balakrishnan and Aggarwala (2000) provided a comprehensive reference on the subject of progressive censoring and its applications. For further reading, the readers are referred to Kundu (2008), Kundu and Pradhan (2009), Ng, Kundu, and Chan (2009) and the references cited therein. A schematic representation of progressively Type-II right censored sample is depicted in Figure 1.1 (Cramer and Iliopoulos 2010).

Under this censoring scheme, n units are placed on a test at time zero and m failures are to be observed. When the first failure is observed, r_1 of surviving units are randomly selected and removed from the experiment. At the time of second failure, r_2 of the remaining $n - r_1 - 1$ units are randomly selected and removed from the experiment. Finally, at the mth failure all the remaining surviving units $r_m = n - m - r_1 - r_2 - \ldots - r_{m-1}$ are removed from the experiment. In this censoring scheme, $r_1, r_2, \ldots, r_m$ are all prefixed.

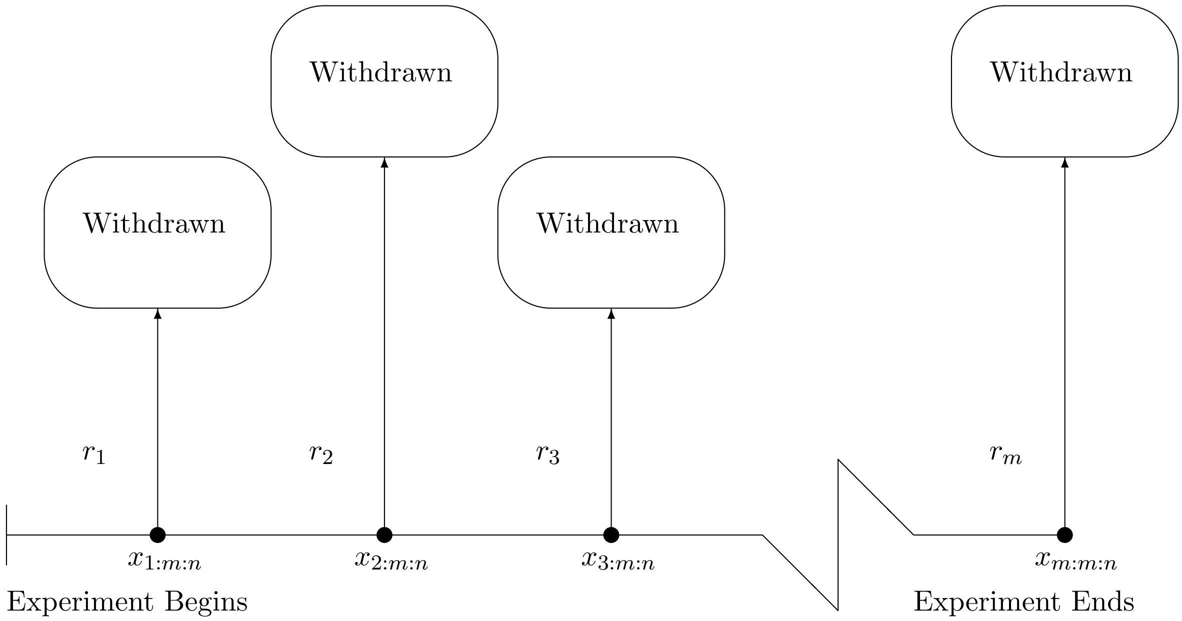


Figure 1: Schematic representation of a progressively Type-II right censored sample where $x_{1:m:n}, x_{2:m:n}, \ldots, x_{m:m:n}$ denote the observed failure times and $r_1, r_2, \ldots, r_m$ denote the corresponding numbers of items removed (withdrawn) from the test.

The key goal of our article is to obtain Bayes estimator and Bayes shrinkage estimator for the parameter of Rayleigh distribution with conjugate prior distribution based on progressively Type-II censored samples. Since we have no idea about the true value of the parameter θ and thus the guess value θ_0, so we propose to obtain the empirical Bayes estimate of θ_0 on the basis of sample observation to approximate the guess value θ_0. Unlike the publications cited above where the focus is on obtaining shrinkage estimator based on complete or Type-II censored samples using exponential, Pareto or Weibull distribution, we have considered Rayleigh distribution with progressive Type-II censoring schemes. Moreover, an empirical Bayes estimate has been taken as guess value for obtaining Bayes shrinkage estimates. Also, risk function of empirical Bayes estimate and empirical Bayes shrinkage estimate have been obtained for comparison.

The rest of the paper is organized as follows. Section 2 discusses prior and loss function used in our Bayesian estimation. In Section 3, we obtained the Bayes' estimators of θ and risk function of maximum likelihood estimator, Bayes estimator and Bayes shrinkage estimator. A simulation study is performed in Section 4, and a real life data is used in Section 5 for the evaluation of classical Bayes estimate and Bayes shrinkage estimate and their estimated risks. We conclude the paper in Section 6.

2. Prior and loss function

In the Bayesian approach, θ is considered a random variable with some specified distribution. In this paper, we consider conjugate prior distribution of the form

$$g(\theta \mid \alpha, \beta) \propto \theta^{-(\alpha+1)} exp\left(-\frac{\beta}{\theta}\right), \theta > 0, \tag{3}$$

where $\alpha > 0$ and $\beta > 0$. The advantage of using natural conjugate prior is that the resulting posterior distribution will also belong to the same family. Furthermore, Jeffreys' prior can be obtained as a special case of (3) by substituting $\alpha = \beta = 0$ and Hartigan's prior can be obtained by substituting $\alpha = 2, \beta = 0$.

In many practical situations it is more realistic to express the loss in terms of ratio $\hat{\theta}/\theta$. In this case, Calabria and Pulcini (1996) proposed a loss function, the general entropy loss function of the form:

$$L(\hat{\theta},\ \theta) \propto \left[\left(\frac{\hat{\theta}}{\theta}\right)^{p} - p \ln\left(\frac{\hat{\theta}}{\theta}\right) - 1\right]; \quad p \neq 0, \tag{4}$$

whose minimum occurs at $\hat{\theta} = \theta$. This loss is a generalization of the entropy loss used by several authors [see, for example, Dey, Ghosh, and Srinivasan (1987) and Dey and Liu (1992)], where the shape parameter p is taken to be equal to 1. If we assume, $\ln\left(\frac{\hat{\theta}}{\theta}\right) = \hat{\theta} - \theta$ i.e., $\left(\frac{\hat{\theta}}{\theta}\right) = e^{(\hat{\theta}-\theta)}$, we get the linear exponential (LINEX) loss function of the form, $\left[e^{p(\hat{\theta}-\theta)} - p(\hat{\theta}-\theta) - 1\right]$ which is proposed by Zellner (1986). Following Calabria and Pulcini (1996), the Bayes estimators for the parameter θ given data $\underline{x}$ under general entropy loss function (GELF) may be defined as

$$\hat{\theta}_{GB} = \left[E\left(\theta^{-p} \mid \underline{x}\right)\right]^{-\frac{1}{p}}.$$

3. Estimation

Let $X = (X_{1:m:n}, X_{2:m:n}, \ldots, X_{m:m:n})$ be a progressively Type-II right censored sample from a life test of effective sample size m from a sample of size n, where the ordered lifetimes have a Rayleigh distribution with pdf and cdf as given by (1) and (2) with a pre-determined number of removal of units from the test, say $R_1 = r_1, R_2 = r_2, \ldots, R_m = r_m$. For convenience, we write $X_{i:m:n}$ as $X_{(i)}$, the ith ordered lifetime. Here the likelihood function of θ is given by

$$\begin{aligned} l(\theta) &= A.\prod_{i=1}^{m} f(x_{(i)} \mid \theta)[1 - F(x_{(i)} \mid \theta)]^{r_i}, \text{ where } A = n(n - r_1 - 1)...(n - \sum_{j=1}^{m-1} r_j - m + 1) \\ &= A.\theta^{-m} e^{-\frac{1}{2\theta}\sum_{i=1}^{m}(1+r_i)x_{(i)}^2} \prod_{i=1}^{m} x_{(i)} \\ &= A.\theta^{-m} e^{-\frac{T}{2\theta}} \prod_{i=1}^{m} x_{(i)}, \end{aligned} \tag{5}$$

where $T = \sum_{i=1}^{m}(1+r_i)x_{(i)}^2$, and $x_{(i)}$ is the observation of $X_{(i)}; i = 1, 2, ...m$. Here the maximum likelihood estimate of θ is $\hat{\theta}_{MLE} = \frac{T}{2m}$. It is to be noted that $\hat{\theta}_{MLE}$ is an unbiased estimator of θ.

Lemma 3.1 *Under progressive Type-II censoring, $\frac{T}{\theta}$ follows a chi-square distribution with degrees of freedom $2m$.*

Proof: Let $Z_i = \frac{X_{(i)}^2}{2\theta}; i = 1, \ldots, m$, then $Z_1 < Z_2 < \ldots < Z_m$ is a progressive Type-II censored sample from the standard exponential distribution. Considering the following transformations

$W_1 = nZ_1$

$W_2 = (n - r_1 - 1)(Z_2 - Z_1)$

$W_3 = (n - r_1 - r_2 - 2)(Z_3 - Z_2)$

....

$W_m = (n - r_1 - r_2 - \ldots - r_{m-1} - m + 1)(Z_m - Z_{m-1})$

Hence $W_1, W_2, \ldots, W_m$ are all independent and identically distributed as standard exponential. Then we have

$2\sum_{i=1}^{m} W_i \sim \chi_{2m}^2$, that is

$2\sum_{i=1}^{m} W_i = 2\sum_{i=1}^{m}(1+r_i)Z_i = 2\sum_{i=1}^{m}(1+r_i)\frac{X_{(i)}^2}{2\theta} = \frac{T}{\theta} \sim \chi_{2m}^2$. □

Combining the prior density (3) with the likelihood function (5) and by using Bayes theorem the joint posterior distribution is derived as follows:

$$\pi(\theta \mid T) = \frac{\left(\beta + \frac{T}{2}\right)^{m+\alpha} \theta^{-(m+\alpha+1)} e^{-\frac{1}{\theta}\left(\beta+\frac{T}{2}\right)}}{\Gamma(m+\alpha)}.$$

The posterior distribution of θ is inverse gamma with parameters $(m+\alpha)$ and $\left(\beta + \frac{T}{2}\right)$. Under the general entropy loss function, the Bayes estimator of θ is the posterior expectation

$$\begin{aligned} \hat{\theta}_{GB} &= \left[E\left(\theta^{-p} \mid T\right)\right]^{-\frac{1}{p}} \\ &= \left[\int_0^\infty \theta^{-p}\pi(\theta \mid T)d\theta\right]^{-\frac{1}{p}}. \end{aligned}$$

On simplification we get,

$$\hat{\theta}_{GB} = \left[\frac{\Gamma(m+\alpha)}{\Gamma(m+\alpha+p)}\right]^{\frac{1}{p}} \left(\beta + \frac{T}{2}\right). \tag{6}$$

We choose the parameters of the prior distribution $g(\theta)$ such that $E(\hat{\theta}_{GB}) = \theta_0$, where θ_0 is the point guess value of θ. This gives

$$E(\hat{\theta}_{GB}) = \theta_0 \quad \Rightarrow \quad \beta = \left(\frac{1}{c} - m\right)\theta_0,$$

where $c = \left[\frac{\Gamma(m+\alpha)}{\Gamma(m+\alpha+p)}\right]^{\frac{1}{p}}$.

Substituting this value of β in $\hat{\theta}_{GB}$, we obtain the Bayes shrinkage estimator for θ under GELF as

$$\hat{\theta}_{GB}^{S} = c.m.T_1 + (1 - c.m)\theta_0 = k_1.T_1 + (1 - k_1)\theta_0, \tag{7}$$

where, $k_1 = c.m = \left[\frac{\Gamma(m+\alpha)}{\Gamma(m+\alpha+p)}\right]^{\frac{1}{p}} m$ and $T_1 = \frac{T}{2m}(= \hat{\theta}_{MLE})$.

The pdf of $W = \frac{\hat{\theta}_{GB}}{\theta} = a(\theta) + b.\frac{T}{\theta}$ is given by

$$f_W(w) = \begin{cases} \frac{1}{b.2^m\Gamma(m)} e^{-\frac{1}{2}\left(\frac{w-a(\theta)}{b}\right)} \left(\frac{w-a(\theta)}{b}\right)^{m-1}, & w > a(\theta), \\ 0, & \text{elsewhere}, \end{cases}$$

where $a(\theta) = \frac{c.\beta}{\theta}$ and $b = \frac{c}{2}$.
The pdf of $V = \frac{\hat{\theta}^S_{GB}}{\theta} = e(\theta) + d.\frac{T}{\theta}$ is given by

$$f_V(v) = \begin{cases} \frac{1}{d.2^m\Gamma(m)} e^{-\frac{1}{2}\left(\frac{v-e(\theta)}{d}\right)} \left(\frac{v-e(\theta)}{d}\right)^{m-1}, & v > e(\theta) \\ 0, & \text{elsewhere}, \end{cases}$$

where, $e(\theta) = \frac{(1-k_1)\theta_0}{\theta}$ and $d = \frac{k_1}{2m}$.
The risk function of $\hat{\theta}_{GB}$ under GELF is given by

$$\begin{aligned} R_{GB}(\theta) &= E[L(\hat{\theta}_{GB}, \theta)] \\ &= E\left[\left(\frac{\hat{\theta}_{GB}}{\theta}\right)^p - p\ln\left(\frac{\hat{\theta}_{GB}}{\theta}\right) - 1\right] \\ &= E\left[\left(a(\theta) + b.\frac{T}{\theta}\right)^p - p\ln\left(a(\theta) + b.\frac{T}{\theta}\right) - 1\right] \\ &= \sum_{j=0}^{p} \frac{p(p-1)\dots(p-j+1)}{j!} a(\theta)^j (2b)^{p-j} \frac{\Gamma(m+j)}{\Gamma(m)} \\ &\quad -p\int_0^\infty \ln(a(\theta) + b.w).\frac{1}{2^m\Gamma(m)} e^{-\frac{w}{2}} w^{m-1} dw - 1 \quad \text{for } p > 0. \end{aligned} \tag{8}$$

Similarly, we will have the risk function of $\hat{\theta}^S_{GB}$ under GELF as $R_{GBS}(\theta)$ and is given by

$$\begin{aligned} R_{GBS}(\theta) &= E\left[\left(\frac{\hat{\theta}^S_{GB}}{\theta}\right)^p - p\ln\left(\frac{\hat{\theta}^S_{GB}}{\theta}\right) - 1\right] \\ &= E\left[\left(e(\theta) + d.\frac{T}{\theta}\right)^p - p\ln\left(e(\theta) + d.\frac{T}{\theta}\right) - 1\right] \\ &= \sum_{j=0}^{p} \frac{p(p-1)...(p-j+1)}{j!} \left(\frac{(1-k_1)\theta_0}{\theta}\right)^j \left(\frac{k_1}{m}\right)^{p-j} \frac{\Gamma(m+j)}{\Gamma(m)} \\ &\quad -p\int_0^\infty \ln\left(\left(\frac{(1-k_1)\theta_0}{\theta}\right) + \left(\frac{k_1}{2m}\right).w\right).\frac{1}{2^m\Gamma(m)} e^{-\frac{w}{2}} w^{m-1} dw - 1 \quad \text{for } p > 0. \end{aligned} \tag{9}$$

The risk function of $\hat{\theta}_{MLE}$ under GELF is given by

$$\begin{aligned} R_{MLE}(\theta) &= E\left[\left(\frac{\hat{\theta}_{MLE}}{\theta}\right)^p - p\ln\left(\frac{\hat{\theta}_{MLE}}{\theta}\right) - 1\right] \\ &= \frac{\Gamma(m+p)}{\Gamma(m)} - p[\psi(m) - \ln(2m)] - 1 \quad \text{for } p > 0, \end{aligned} \tag{10}$$

where $\psi(.)$ is the digamma function.
To find the Bayes shrinkage estimator, we need a guess value of θ_0 that at times could be unknown. We will use the empirical Bayes estimate of θ in place of θ_0. The Bayes' estimator given in (6) depends on α and β, and that in (7) depends on α, which are the parameters of the prior distribution of θ. The parameters α and β, could be estimated by means of empirical Bayes' procedure (see Lindley (1969) and Awad and Gharraf (1986)). Given the

random sample $\underline{x}$, the likelihood function of θ has an inverse gamma density with shape parameter $(m-1)$ and scale parameter $\frac{T}{2}$. Following this, we can use the estimates of the prior parameters α and β from the sample by $(m-1)$ and $\frac{T}{2}$ respectively. Hence we can take $\hat{\theta}_0$ as

$$\begin{aligned}\hat{\theta}_0 = \hat{\theta}_{GB.emp} &= \left[\frac{\Gamma(2m-1)}{\Gamma(2m+p-1)}\right]^{\frac{1}{p}} T \\ &= c_{emp}.T, \end{aligned} \tag{11}$$

with $c_{emp} = \left[\frac{\Gamma(2m-1)}{\Gamma(2m+p-1)}\right]^{\frac{1}{p}}$.

To find out $\hat{\theta}_{GBS.emp}$, we substitute $\hat{\theta}_0$ in place of θ_0 and hence

$$\begin{aligned}\hat{\theta}_{GBS.emp} &= c_{emp}.m.T_1 + (1 - c_{emp}.m)\hat{\theta}_0 \\ &= k_{1.emp}.T_1 + (1 - k_{1.emp})\hat{\theta}_0 \\ &= k_{1.emp}.\hat{\theta}_{MLE} + (1 - k_{1.emp})\hat{\theta}_0 \\ &= \left(\frac{3}{2}c_{emp} - m.c^2_{emp}\right) T = k_{emp}T. \end{aligned} \tag{12}$$

The risk functions of empirical Bayes estimator and empirical Bayes shrinkage estimator are

$$R_{GB.emp}(\theta) = \frac{(2c_{emp})^p.\Gamma(m+p)}{\Gamma(m)} - p[\psi(m) + \ln\{2c_{emp}\}] - 1 \tag{13}$$

and

$$R_{GBS.emp}(\theta) = \frac{(2k_{emp})^p.\Gamma(m+p)}{\Gamma(m)} - p[\psi(m) + \ln\{2k_{emp}\}] - 1 \tag{14}$$

respectively.

4. Simulation study

We present some experimental results to observe the performance of Bayes estimation of θ using the prior (3) for different sample sizes, different effective sample sizes, different priors, and for different sampling schemes. We have considered different sample sizes $n = 20, 25, 30, 35$, different effective sample sizes; $m = 10$, 15, and 20, different hyper-parameter values for α and β, and six ([1] - [6]) different sampling schemes as described in Table 1. We have simulated progressive Type-II samples using the algorithm proposed by Balakrishnan and Sandhu (1995) from the Rayleigh model with $\theta = 2$. The estimators are compared based on the average value of estimates and their corresponding risk performances. For Bayesian computation, we have considered two different set of hyperparameters: $\alpha = \beta = 0$, which is a non-informative prior (labeled as "Prior 1") and $\alpha = \beta = 2$, which is an informative prior (labeled as "Prior 2"). The two different values of p are considered for the loss function: $p = 1, 2$. The simulation results are summarized in Table 2 and 3. All results are based on $10,000$ repetitions.

From Tables 2 and 3 we notice that risk does not depend on sample size n but it depends on effective sample size m, and as effective sample size increases, the risk decreases and average Bayes estimates come closer to the true value of θ. When a guess value θ_0 is chosen in the neighborhood of θ, then obviously the risk is minimum and Bayes shrinkage estimate is better. But in practice, θ is unknown and choosing a guess value is difficult one. Even though the risk for empirical Bayes estimates are a bit more, but from a pragmatic point of view, empirical Bayes shrinkage estimator is worthwhile.

Table 1: Different progressive Type-II censoring schemes used in the simulation study

Number	Scheme
[1]	(0 0 0 0 0 0 0 0 0 10)
[2]	(0 0 0 0 0 0 0 0 0 15)
[3]	(0 0 0 0 0 0 0 0 0 0 0 0 0 0 10)
[4]	(0 0 0 0 0 0 0 0 0 20)
[5]	(0 0 0 0 0 0 0 0 0 0 0 0 0 0 15)
[6]	(0 2 1 0 0 2 1 2 1 1 0 0 2 0 8)

* In the Table, Scheme (0 0 0 r) indicates that at 1st, 2nd and 3rd failure, no active unit is withdrawn or removed but at 4th failure, r active units are withdrawn or removed from the test.

Table 2: The average MLE and the Bayes estimators ($\hat{\theta}$) from different estimation procedures for the parameter θ under the general entropy loss function with $p = 1$. The true risk (in bold) and the estimated risk are presented within parentheses.

		(n, m, Censoring scheme)					
	Estimate	(20,10,[1])	(25, 10, [2])	(25, 15, [3])	(30, 10, [4])	(30, 15, [5])	(35, 20, [6])
	$\hat{\theta}_{MLE}$	2.004	1.971	1.997	2.008	1.995	1.993
		(**0.051**, 0.049)	(**0.051**, 0.052)	(**0.034**, 0.036)	(**0.050**, 0.047)	(**0.033**, 0.031)	(**0.025**, 0.026)
Prior 1	$\hat{\theta}_{GB}$	2.004	1.971	1.997	2.008	1.995	1.993
		(**0.051**, 0.049)	(**0.051**, 0.052)	(**0.034**, 0.036)	(**0.050**, 0.047)	(**0.033**, 0.031)	(**0.025**, 0.026)
	$\hat{\theta}_{GBS}$	2.004	1.971	1.997	2.008	1.995	1.993
		(**0.051**, 0.049)	(**0.051**, 0.052)	(**0.034**, 0.036)	(**0.050**, 0.047)	(**0.033**, 0.031)	(**0.025**, 0.026)
	$\hat{\theta}_{GB.emp}$	2.109	2.074	2.066	2.114	2.064	2.052
		(**0.052**, 0.051)	(**0.052**, 0.052)	(**0.034**, 0.037)	(**0.052**, 0.049)	(**0.034**, 0.031)	(**0.025**, 0.026)
	$\hat{\theta}_{GBS.emp}$	2.054	2.019	2.031	2.058	2.028	2.019
		(**0.051**, 0.049)	(**0.051**, 0.052)	(**0.033**, 0.037)	(**0.051**, 0.048)	(**0.033**, 0.031)	(**0.025**, 0.026)
Prior 2	$\hat{\theta}_{GB}$	1.837	1.837	1.874	1.813	1.885	1.902
		(**0.045**, 0.046)	(**0.045**, 0.046)	(**0.031**, 0.032)	(**0.045**, 0.044)	(**0.031**, 0.029)	(**0.025**, 0.026)
	$\hat{\theta}_{GBS}$	1.921	1.921	1.934	1.896	1.943	1.949
		(**0.038**, 0.039)	(**0.038**, 0.039)	(**0.028**, 0.029)	(**0.038**, 0.037)	(**0.028**, 0.026)	(**0.023**, 0.025)
	$\hat{\theta}_{GB.emp}$	2.110	2.110	2.060	2.079	2.071	2.069
		(**0.052**, 0.054)	(**0.052**, 0.054)	(**0.034**, 0.035)	(**0.052**, 0.049)	(**0.034**, 0.032)	(**0.022**, 0.024)
	$\hat{\theta}_{GBS.emp}$	2.054	2.054	2.024	2.024	2.035	2.031
		(**0.051**, 0.053)	(**0.051**, 0.053)	(**0.033**, 0.034)	(**0.051**, 0.049)	(**0.033**, 0.031)	(**0.025**, 0.026)

5. Data analysis

We are considering the data that appeared in tests on endurance of deep grove ball bearings [Lawless (1982), p.228]. The data are the number of hundreds of million revolutions before failure for each of the 23 ball bearings in the life test:
0.1788, 0.2892, 0.33, 0.4152, 0.4212, 0.456, 0.4848, 0.5184, 0.5196, 0.5412, 0.5556, 0.678, 0.6864, 0.6864, 0.6888, 0.8412, 0.9312, 0.9864, 1.0512, 1.0584, 1.2792, 1.2804, 1.734.
To study the goodness of fit of the Rayleigh model, we compute the χ^2 statistic (with 3 degrees of freedom) and it is 1.0312 with the corresponding p-value 0.7937. Therefore, the high p-value clearly suggests that the one parameter Rayleigh model can be used to analyze this data set. Besides this, we also plotted the scaled TTT (TTT stands for "total time on test") transformed (Aarset 1987) of the ball bearing data. Usually this plot is used to identify whether a random sample is from a lifetime distribution with constant against bathtub-type hazard rate; for further details see Aarset (1987). Figure 2 indicates that the empirical hazard function is unimodal and therefore it is reasonable to use Rayleigh distribution to analyze the

Table 3: The average MLE and the Bayes estimators ($\hat{\theta}$) from different estimation procedures for the parameter θ under the general entropy loss function with $p = 2$. The true risk (in bold) and the estimated risk are presented within parentheses.

		(n, m, Censoring scheme)					
	Estimate	(20,10,[1])	(25, 10, [2])	(25, 15, [3])	(30, 10, [4])	(30, 15, [5])	(35, 20, [6])
	$\hat{\theta}_{MLE}$	2.002	1.997	2.017	1.988	1.986	1.979
		(**0.201**, 0.198)	(**0.201**, 0.198)	(**0.134**, 0.128)	(**0.201**, 0.196)	(**0.134**, 0.139)	(**0.104**, 0.109)
Prior 1	$\hat{\theta}_{GB}$	1.909	1.904	1.953	1.896	1.923	1.916
		(**0.196**, 0.198)	(**0.196**, 0.193)	(**0.131**, 0.125)	(**0.196**, 0.193)	(**0.131**, 0.138)	(**0.106**, 0.103)
	$\hat{\theta}_{GBS}$	1.979	1.974	2.001	2.039	1.971	1.967
		(**0.185**, 0.193)	(**0.185**, 0.182)	(**0.126**, 0.121)	(**0.185**, 0.181)	(**0.126**, 0.132)	(**0.105**, 0.103)
	$\hat{\theta}_{GB.emp}$	2.054	2.049	2.051	2.114	2.020	2.001
		(**0.208**, 0.205)	(**0.208**, 0.204)	(**0.136**, 0.132)	(**0.208**, 0.202)	(**0.136**, 0.142)	(**0.104**, 0.102)
	$\hat{\theta}_{GBS.emp}$	2.028	2.022	2.033	2.013	2.002	1.998
		(**0.204**, 0.201)	(**0.204**, 0.201)	(**0.135**, 0.131)	(**0.204**, 0.199)	(**0.135**, 0.140)	(**0.104**, 0.101)
Prior 2	$\hat{\theta}_{GB}$	1.762	1.773	1.827	1.729	1.982	1.978
		(**0.176**, 0.183)	(**0.176**, 0.176)	(**0.122**, 0.124)	(**0.176**, 0.179)	(**0.122**, 0.122)	(**0.100**, 0.104)
	$\hat{\theta}_{GBS}$	1.901	1.912	1.926	1.868	1.813	1.802
		(**0.139**, 0.146)	(**0.139**, 0.141)	(**0.104**, 0.115)	(**0.139**, 0.137)	(**0.104**, 0.116)	(**0.094**, 0.102)
	$\hat{\theta}_{GB.emp}$	2.052	2.067	2.031	2.011	2.015	2.001
		(**0.208**, 0.218)	(**0.208**, 0.213)	(**0.136**, 0.127)	(**0.208**, 0.198)	(**0.136**, 0.124)	(**0.082**, 0.100)
	$\hat{\theta}_{GBS.emp}$	2.026	2.040	2.014	1.984	1.999	1.982
		(**0.204**, 0.214)	(**0.204**, 0.209)	(**0.135**, 0.126)	(**0.204**, 0.196)	(**0.135**, 0.123)	(**0.101**, 0.107)

data.

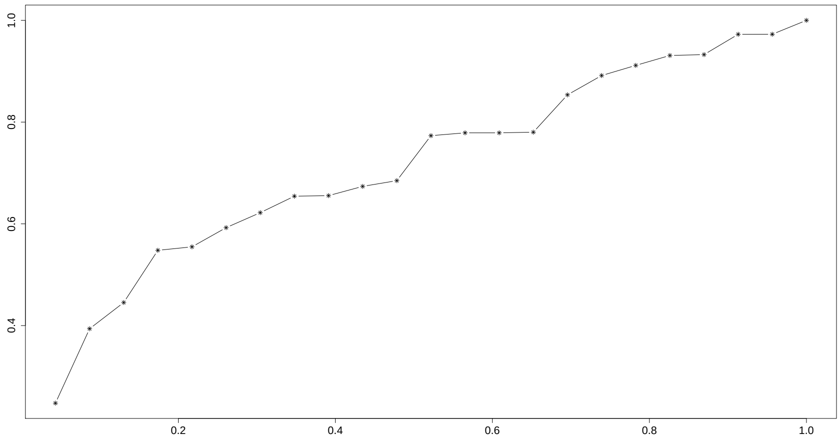


Figure 2: Scaled TTT transformed plot of the ball bearing data

For the purposes of illustrating the methods discussed in this article, a progressively Type-II censored sample was generated from this data set with effective sample size $m = 13$. The observations (in hundreds of millions) and removed numbers are reported in Table 4 [see Wu *et al.* (2006)].

It is observed from Table 5 that the Empirical Bayes Shrinkage Estimate of θ is better in terms of minimum risk sense. As p increases, estimates decrease and their risks increase.

Table 4: Progressively Type II censored sample for Example.

i	1	2	3	4	5	6	7
$x_{(i)}$	0.1788	0.2892	0.33	0.4212	0.456	0.4848	0.5184
r_i	0	0	3	0	0	2	0
i	8	9	10	11	12	13	
$x_{(i)}$	0.5196	0.6780	0.6864	0.8412	0.9312	1.2792	
r_i	0	2	0	2	1	0	

Table 5: Empirical Bayes and Empirical Bayes Shrinkage Estimate and their risks for the ball bearing data.

p	Empirical Bayes Estimate	Empirical Bayes risk	Empirical Bayes Shrinkage Estimate	Empirical Bayes Shrinkage risk
1	0.3809	0.0397	0.3733	0.0391
2	0.3735	0.1586	0.3698	0.1565

6. Conclusion

In this article, we have derived Bayes shrinkage estimate of the parameter of the Rayleigh distribution under conjugate prior assuming general entropy loss function. The Bayes estimate, Bayes shrinkage estimate assuming a point guess value are derived and their risks have been studied. Bayes shrinkage estimate is reasonably good from a risk perspective. In practice, as the true value of the parameter is unknown, getting a point guess value is difficult. An empirical Bayes procedure has been followed to get an estimated guess value of the parameter and utilizing so, empirical Bayes and empirical Bayes shrinkage estimates and their risks have been calculated. The performance of the empirical Bayes shrinkage estimate is fairly reasonable and competitive. We recommend empirical Bayes shrinkage estimate for practical purposes intended for enhanced outcomes. Real life data analysis also echos similar trend as observed in the simulation study.

Acknowledgments

The authors would like to thank the editor and the reviewer who helped to substantially improve the paper.

References

Aarset M (1987). "How to Identify a Bathtub Hazard Function." *IEEE Transactions on Reliability*, **36**, 106–108.

Abd Elfattah AM, Hassan AS, Ziedan DM (2006). "Efficiency of Maximum Likelihood Estimators under Different Censored Sampling Schemes for Rayleigh Distribution." *Interstat*. URL http://interstat.statjournals.net/YEAR/2006/articles/0603001.pdf.

Al-Hemyari ZA, Al-Dabag HA (2012). "A Class of Shrinkage Testimators for the Shape Parameter of the Weibull Lifetime Model." *Pakistan Journal of Statistics and Operational Research*, **8**(2), 167–184.

Ariyawansa KA, Templeton JGC (1984). "Structural Inference on the Parameter of the Rayleigh Distribution from Doubly Censored Samples." *Statist. Hefte*, **25**, 181–199.

Awad AM, Gharraf MK (1986). "Estimation of $P[Y < X]$ in the Burr Case: A Comparative Study." *Communications in Statistics- Simulation and Computation*, **15**(15), 189–203.

Balakrishnan N, Aggarwala R (2000). *Progressive Censoring: Theory, Methods and Applications.* Birkhäuser, Boston.

Balakrishnan N, Sandhu RA (1995). "A Simple Simulation Algorithm for Generating Progressive Type-II Censored Samples." *The American Statistician*, **49**(2), 119–230.

Calabria R, Pulcini G (1996). "Point Estimation under Asymmetric Loss Functions for Left-Truncated Exponential Samples." *Communications in Statistics-Theory & Methods*, **25**(3), 585–600.

Cramer E, Iliopoulos G (2010). "Adaptive Progressive Type-II Censoring." *Test*, **19**(2), 342–358.

Dey DK, Ghosh M, Srinivasan C (1987). "Simultaneous Estimation of Parameters under Entropy Loss." *Journal of Statistical Planning and Inference*, **15**, 347–363.

Dey DK, Liu PL (1992). "On Comparison of Estimators in a Generalized Life Model." *Microelectronics Reliability*, **32**, 207–221.

Dey S (2009). "Comparison of Bayes Estimators of ohe Parameter and Reliability Function for Rayleigh Distribution under Different Loss Functions." *Malaysian Journal of Mathematical Sciences*, **3**, 247–264.

Dey S, Das MK (2007). "Note on Prediction Interval for a Rayleigh Distribution: Bayesian Approach." *American Journal of Mathematical and Management Science*, **1-2**, 43–48.

Harris E, Shakarki G (1979). "Use of the Population Distribution to Improve Estimation of Individual Mean in Epidemiological Studies." *Journal of Chronical Disease*, **32**, 233–243.

Hendi ML, Abu-Youssef SE, Alraddadi AA (2007). "A Bayesian Analysis of Record Statistics from the Rayleigh Model." *International Mathematical Forum*, **2**(13), 619–631.

Howlader HA (1985). "HPD Prediction Intervals for Rayleigh Distribution." *IEEE Transactions on Reliability*, **34**, 121–123.

Howlader HA, Hossain A (1995). "On Bayesian Estimation and Prediction from Rayleigh Distribution Based on Type-II Censored Data." *Comm. Stat. Theory Methods*, **24**(9), 2249–2259.

Kundu D (2008). "Bayesian Inference and Reliability Sampling Plan for Weibull Distribution." *Technometrics*, **50**, 144–154.

Kundu D, Pradhan B (2009). "Bayesian Inference and Life Testing Plans for Generalized Exponential Distribution." *Science in China, Series A: Mathematics*, **52**(6), 1373–1388.

Lawless JF (1982). *Statistical Models and Methods for Lifetime Data.* John Wiley and Sons, New York.

Lindley DV (1969). *Introduction to Probability and Statistics from a Bayesian View Point. Vol.1.* Cambridge University Press.

Marshall RJ (1991). "Mapping Disease and Mortality Rates using Empirical Bayes Estimators." *Journal of Applies Statistics*, **40**, 283–294.

Ng HKT, Kundu D, Chan PS (2009). "Statistical Analysis of Exponential Lifetimes under an Adaptive Type-II Progressive Censoring Scheme." *Naval Research Logistics*, **56**, 687–698.

Prakash G (2009). "Some Estimators for the Pareto Distribution." *Journal of Scientific Research*, **1**(2), 236–247.

Prakash G, Singh DC (2006). "Shrinkage Estimators for the Inverse Dispersion of the Inverse Gaussian Distribution under the LINEX Loss Function." *Austrian Journal of Statistics*, **35**(4), 463–470.

Rayleigh JWS (1980). "On the Resultant of a Large Number of Vibrations of the Some Pitch and of Arbitrary Phase." *Philosophical Magazine, 5-th Series*, **10**, 73–78.

Salman AN, Shareef RA (2014). "Bayesian Shrinkage Estimator for the Scale Parameter of Exponential Distribution under Improper Prior Distribution." *International Journal of Statistics and Applications*, **4**(3), 135–143.

Singh DC, Prakash G, Singh P (2007). "Shrinkage Estimators for the Shape Parameter of Pareto Distribution using the LINEX Loss Function." *Communication in Statistics - Theory and Methods*, **36**(4), 741–753.

Singh GP, Singh SK, Singh U, Upadhyay SK (2008). "Bayes Estimators of Exponential Parameters from a Censored Sample using a Guess Estimate." *Data Science Journal*, **7**, 106–114.

Sinha SK, Howlader HA (1983). "Credible and HPD Intervals of the Parameter and Reliability of Rayleigh Distribution." *IEEE Transactions on Reliability*, **32**, 283–294.

Thompson JR (1968). "Some Shrinkage Techniques for Estimating the Mean." *Journal of American Statistical Association*, **63**, 113–122.

Tso G (1990). "Forecasting money supply in Hong Kong with a multiple shrinkage estimator." In *Proceeding of the ASA Section on Business and Commerce*. ASA.

Wooff A (1985). "Bounds on Reciprocal Moments with Applications and Developments in Stein Estimation and Post Stratification." *Journal of Royal statistical Society, B*, **47**, 362–371.

Wu SJ, Chen DH, Chen ST (2006). "Bayesian Inference for Rayleigh Distribution under Progressive Censored Sample." *Applied Stochastic Models in Business and Industry*, **22**, 269–279.

Zellner A (1986). "Bayesian Estimation and Prediction using Asymmetric Loss Function." *Journal of American statistical Association*, **81**, 446–451.

Affiliation:

Sanku Dey
Department of Statistics
St. Anthony's College
Shillong, Meghalaya, India
E-mail: sanku_dey2k2003@yahoo.co.in

Tanujit Dey
Department of Quantitative Health Sciences
Cleveland Clinic
Cleveland, OH USA
E-mail: deyt@ccf.org

Sudhansu S. Maiti
Department of Statistics
Visva-Bharati University
Santiniketan, West Bengal, India
E-mail: dssm1@rediffmail.com

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 44 *Submitted:* 2014-09-15
December 2015 *Accepted:* 2015-01-27

Austrian Journal of Statistics
December 2015, Volume 44, 17–33.
http://www.ajs.or.at/
doi:10.17713/ajs.v44i4.74

Restricted Graphical Log-linear Models

Ricardo Ramírez-Aldana[1]
Department of Mathematics, Faculty of Sciences, UNAM, Mexico.

Guillermina Eslava-Gómez
Department of Mathematics, Faculty of Sciences, UNAM, Mexico.

Abstract

We introduce a new type of graphical log-linear model called restricted graphical log-linear model. This model is obtained by imposing equality restrictions on subsets of main effects and of first-order interactions. These restrictions are obtained through partitions of the variable and first-order interaction sets. The vertices or variables in the same class have the same main effects in all their categories, and the first-order interactions in the same class are equal. We study its properties and derive its associated likelihood equations and provide some applications. A graphical representation is possible through a coloured graph.

Keywords: graph colourings, graphical Gaussian models with symmetries, graphical log-linear models, iterative proportional fitting, log-linear models, symmetry and quasi-symmetry models.

1. Introduction

In this paper we introduce a new type of model for discrete variables called restricted or coloured graphical log-linear model (*RGLL model*), which combines symmetry with discrete graphical models. Symmetry is considered through specific parameter restrictions not used before, but inspired by those used in the continuous case. It can be seen as a tool to model symmetry and could help to get a better understanding of the data. In a specific model we may obtain improved knowledge about the distribution; for instance, conditional independences as in graphical models or the relationship between cells in a contingency table. RGLL can be fitted to different kinds of data such as panel data, where the variables have the same categories. The model may be useful for researchers of linear models, symmetry, and graphical models.

RGLL models are special cases of hierarchical log-linear models and can be represented graphically, including equality restrictions between certain parameters. Restrictions are imposed on two kind of parameters: the first ones correspond to main effects terms and the second ones to first-order interactions. For the former, restriction classes consist of variables; for the latter, the classes are defined by the parameters themselves. At the same time, marginal and

[1]External lecturer

conditional independences derived from a graph are considered. A graphical representation involves colouring the associated graph.

Symmetry and quasi-symmetry models for square contingency tables were introduced by Bowker (1948) and Caussinus (1965), respectively; and afterwards they were further studied by other authors; for instance, Agresti (2002, p. 423-431). For the same type of tables, Tahata, Yamamoto, Nagatani, and Tomizawa (2009) defined a measure and three types of symmetry through it and distinguished departures from usual symmetry assuming conditional symmetry, and Yamamoto, Iwashita, and Tomizawa (2007) studied alternative ways to decompose symmetry into models that include quasi-symmetry. More recently, Kateri, Mohammadi, and Sturmfels (2014) defined one model in the same two-dimensional case in which algebraic and graphical concepts are used, particular cases corresponding to quasi-symmetry and the Pearsonian quasi-symmetry model, the latter defined through a divergence measure. Lovison (2000) introduced generalized symmetry models (*GS models*), which are defined for any dimension when the same variable is observed several times and where some or even all interaction parameters of any order are invariant under any permutation of the associated index. The other parameters are free to vary and restrictions on the main effects as the ones derived from RGLL models are allowed. However, only GS models for the three-dimensional case are presented.

The relationship between graphical models and symmetry has not been studied as extensively in the discrete case. In the continuous case, Højsgaard and Lauritzen (2008) introduced Gaussian graphical models with symmetries; in one of them (*RCON model*), equalities are set among elements of the concentration matrix. In the discrete case, quasi-symmetric (*QS*) and symmetric and quasi-symmetric (*SQS*) graphical models were introduced by Gottard, Marchetti, and Agresti (2011) and Gottard (2009), respectively.

Symmetry and quasi-symmetry models for square contingency tables can be expressed as RGLL models with two variables. In this sense, RGLL models can be considered as a generalization of symmetry models. Any graphical log-linear model (e.g., Lauritzen 1996, ch. 4 and Edwards 2000, ch. 2) can be expressed as an RGLL model by considering that all parameters are unrestricted. Conditional symmetry (Andersen 1991, p. 328-329) is also a particular case of an RGLL model.

RGLL and RCON models are similar though they are defined for different data types. There is also a relationship between RGLL models and QS and SQS graphical models. Let $u_{XY}(ij)$ represent the first-order interaction between variables X and Y when they take the values i and j, respectively, and $u_X(i)$ a main effect. In SQS graphical models, the main effects restrictions are similar to those for RGLL models, and in SQS and QS models there are restrictions of the kind $u_{XY}(ij) = u_{RS}(ij)$ for all $i, j = 1, ..., L$ (considering L levels in each variable), for elements in the same class; including restrictions $u_{XY}(ij) = u_{XY}(ji)$ for all $i, j = 1, ..., L$, and $u_{RS}(ij) = u_{RS}(ji)$ for all $i, j = 1, ..., L$; and implying specific restrictions on the second-order interactions, i.e. on the parameters $u_{XYZ}(ijk)$. QS graphical models are similarly defined, but they allow even higher-order interactions. Unlike RGLL models, interactions of second or higher order are restricted. The first-order interaction restrictions in any SQS or QS graphical model are a particular case of the kind of restrictions defined for those terms in RGLL models. When these are the highest-order interactions, QS and SQS graphical models are RGLL models.

RGLL, SQS, and GS models are different and related as follows. In SQS models, restrictions of the kind $u_{XY}(ij)= u_{RS}(ij)$ are allowed besides invariance under permutation of the indexes, which are the only kind of restrictions allowed in GS models. However, in RGLL models we allow any restriction on the indexes corresponding to first-order interactions, which do not necessarily have to be permutations or the kind of restriction given in SQS models, and higher-order interactions are not restricted. Hence, the three models are equivalent when (1) a model includes at most first order interactions, (2) the restrictions are of the kind $u_{XY}(ij) = u_{XY}(ji)$, and (3) there are no other first-order interaction restrictions.

Twins data. In order to motivate RGLL models, consider the data analyzed by Drton and Richardson (2008) (Table 1). The data consist of 597 observations for monozygotic twins indicating whether twin i ($i = 1, 2$) suffers from major depression (variable D_i, for $i = 1, 2$) or alcoholism (variable A_i, for $i = 1, 2$). The values associated with D_i and A_i correspond to 0 (*no*) and 1 (*yes*). Consider a model in which the parameters are obtained from the set

$$\mathbf{A} = \{\{A_1, A_2\},\ \{A_1, D_1\}, \{A_2, D_2\}, \{D_1, D_2\}\}, \tag{1}$$

called the generating class. Hence, the expected frequency $m_{A_1A_2D_1D_2}(a_1, a_2, d_1, d_2)$ for a specific cell in the table (a_1, a_2, d_1, d_2) has the following log-linear expansion

$$\begin{aligned} \log m(a_1, a_2, d_1, d_2) =& u + u_{A_1}(a_1) + u_{A_2}(a_2) + u_{D_1}(d_1) + u_{D_2}(d_2) + \\ & u_{A_1A_2}(a_1a_2) + u_{A_1D_1}(a_1d_1) + u_{A_2D_2}(a_2d_2) + u_{D_1D_2}(d_1d_2). \end{aligned} \tag{2}$$

Let V be the set of variables or vertex set and E the set of first-order interactions. A model in which the values of variables A_1 and D_1 are permuted with those of variables A_2 and D_2, respectively, in such a way that the distribution is preserved corresponds to an RGLL model. It is the model with generating class $\mathbf{A}$ and graph G=(V, E) where V is partitioned into two sets $V = (V_1, V_2)$ and E= $(E_1, E_2, ..., E_{10})$. Here, $V_1 = \{A_1, A_2\}$ and $V_2 = \{D_1, D_2\}$; and E_1= $\{u_{A_1D_1}(00),\ u_{A_2D_2}(00)\}$, E_2= $\{u_{A_1D_1}(01),\ u_{A_2D_2}(01)\}$, E_3= $\{u_{A_1D_1}(10),\ u_{A_2D_2}(10)\}$, E_4= $\{u_{A_1D_1}(11),\ u_{A_2D_2}(11)\}$, E_5= $\{u_{A_1A_2}(01),\ u_{A_1A_2}(10)\}$, E_6= $\{u_{D_2D_1}(01),\ u_{D_2D_1}(10)\}$, E_7= $\{u_{A_1A_2}(00)\}$, E_8= $\{u_{A_1A_2}(11)\}$, E_9= $\{u_{D_2D_1}(00)\}$, and E_{10}= $\{u_{D_2D_1}(11)\}$. The partition in V means that the main effects for all the values taken by the variables in each set V_1 and V_2 are the same, whereas the partition in E means that the first-order interactions in each class are the same. The associated graph G is given in Figure 1, in which the partitions are represented by using colours; for instance, variables A_1 and A_2 have the same colour. Observe that one edge of the graph is associated with each first-order interaction. To avoid using too many colours, lines in black represent unrestricted first-order interactions.

The distribution is preserved after the permutation of the values of the variables because the expected frequency $m_{A_1A_2D_1D_2}(a_1, a_2, d_1, d_2)$, given the log-linear expansion in equation (2) and the equality restrictions given by the partitions, is equal to the expected frequency for the cell (a_2, a_1, d_2, d_1); thus $m_{A_1A_2D_1D_2}(a_1, a_2, d_1, d_2)$= $m_{A_1A_2D_1D_2}(a_2, a_1, d_2, d_1)$. The model also implies that A_1 and D_2 are conditionally independent given D_1 and A_2, $A_1 \perp D_2 | D_1, A_2$, and $A_2 \perp D_1 | A_1, D_2$.

The remainder of the paper is organized as follows. In Section **2** we formally define RGLL models. Section **3** is devoted to symmetry and quasi-symmetry models and their representation as RGLL models including an example. In Section **4** we derive the likelihood equations and consider model selection. Finally, in the Appendix we provide a method for getting a numerical solution.

Table 1: Alcohol dependence and major depression for 597 pairs of female twins. Left panel: Observed counts. Right panel: Fitted frequencies under an RGLL model.

		A_2				**A_2**			
		0		**1**		**0**		**1**	
		D_2				**D_2**			
A_1	**D_1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**
0	**0**	288	80	15	9	285.29	84.87	12.60	11.24
	1	92	51	7	10	84.87	55.97	3.75	7.41
1	**0**	8	4	3	2	12.60	3.75	3.52	3.14
	1	8	9	4	7	11.24	7.41	3.14	6.21

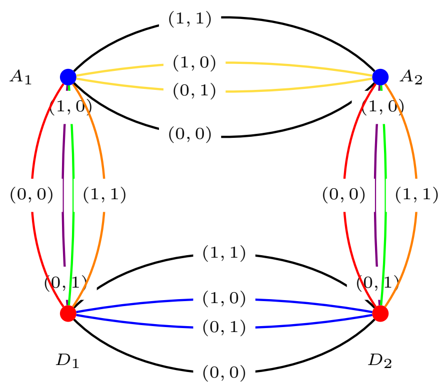


Figure 1: An RGLL model with generating class $\mathbf{A}= \{\{A_1, A_2\}, \{A_1, D_1\}, \{A_2, D_2\}, \{D_1, D_2\}\}$ associated with the twins data.

2. Preliminaries and definition

Consider V as a set of variable labels or a set of vertices. For each element δ in V we associate a discrete random variable I_δ, so that the set of random variables can be expressed as a vector $I = (I_\delta)_{\delta \in V}$. The values taken by those variables, known as levels or categories, are included in a set I_δ. Then all possible values that I takes are given by I, with $\mathrm{I} = \times_{\delta \in V} \mathrm{I}_\delta$. A cell i corresponds to one of these values, i.e. $i \in \mathrm{I}$. Seen as a q-way contingency table, a specific cell can be denoted as $i = (i_1, i_2, ..., i_q)$. A subvector of I is denoted as I_a, where $a \subseteq V$, and a value taken by it is denoted as i_a, where $i_a \in \mathrm{I}_a = \times_{\delta \in a} \mathrm{I}_\delta$.

We denote a parameter associated with a set a, $a \subseteq V$, which depends on the corresponding values i_a, as $u_a(i_a)$. When $a = \emptyset$ it corresponds to a constant term, when $|a| = 1$ to a main effect, and when $|a| > 1$ to an interaction. We also denote $m(i)$ as the expected frequency in a cell $i \in \mathrm{I}$, for instance in a two-way contingency table it is $m(i_1, i_2)$.

A saturated log-linear model can be written as

$$\log m(i) = \sum_{a \subseteq V} u_a(i_a). \tag{3}$$

By setting specific parameters $u_a(i_a)$, for all i_a, to zero in (3), different models are obtained. For example, a log-linear model is called hierarchical when the presence of an interaction $u_a(i_a)$ implies the presence of all those interactions $u_b(i_b)$ with $b \subset a$. The parameters included in this model depend on a generating class $\mathbf{A}$, which is a set of subsets of V. The model can be written as

$$\log m(i) = \sum_{a \in K} u_a(i_a), \tag{4}$$

where K is the set of subsets of the elements in the generating class $\mathbf{A}$.

Definition 1. A hierarchical log-linear model,

$$\log m(i) = \sum_{a \in K} u_a(i_a),$$

is a *restricted graphical log-linear model (RGLL model)* with associated graph $G = (V, E)$ if it satisfies two properties: (a) its generating class is C, the cliques set in the associated graph, and (b) the set of variables V and the set of first-order interactions E are partitioned as follows: the set V is partitioned into $V_1, ..., V_T$, with $T \in \{1, 2, ..., |V|\}$, such that the main effects of the variables in the subset V_t, for t=1,..., T, are equal in all their levels in I^{V_t}, where I^{V_t} is the level set associated with $I_{V_t} = (I_\delta)_{\delta \in V_t}$, for t=1,..., T. The set E is partitioned into $E_1, ..., E_S$, with $S \in \{1, 2, ..., |E|\}$, such that the interactions in every subset E_s, for s=1,..., S, are equal.

RGLL models can be defined assuming that all variables have the same categories, i.e. the level set I_δ is the same for all $\delta \in V$ or I^{V_t} is the same for all V_t, for $t = 1, ..., T$. This implies that the number of categories is the same for any variable, i.e. $|\mathrm{I}_\delta|= L$, for all $\delta \in V$. The advantages of such a restriction are the following: (1) it allows for symmetry interpretations, and (2) it simplifies some of the computational programming. This condition could have been relaxed assuming that only those variables in the same vertex class V_t with $|V_t| > 1$ for every $t = 1, ..., T$ should have the same number of categories or even the same categories; however, interpretation of such models might not be straightforward. The interpretation of an RGLL model depends on the equality restrictions. Some RGLL models have a symmetry interpretation; for instance, in Section 1 this was discussed through an example.

Even though RGLL models could be represented through a graph, this representation might be too complex, for instance when there are six or more variables with three or more levels each one. However, a graphical representation is optional, in the sense that it helps to represent a model but does not define it. In fact, this also happens for any graphical model, so that concepts from graph theory are used, but a graphical representation is not necessary.

Consider an RGLL model with set of variables V and set of first-order interactions E. We define the associated diagram G =(V, E^*) as a graph with vertex and edge set V and E^*, respectively. The set E^* is formed by elements $\{X_k, Y_k\}$, for k= 1,..., $|\mathrm{I}_X|\,|\mathrm{I}_Y|$, where $|\mathrm{I}_X|\,|\mathrm{I}_Y|$ corresponds to all combination of categories of X and Y, considering that $u_{XY}(ij) \in E$ for every (i, j), where $i \in \mathrm{I}_X$ and $j \in \mathrm{I}_Y$, for any $\{X, Y\} \in V$. By definition, there is a one-to-one correspondence between the sets E^* and E so that the graph (V, E) is the same as the graph (V, E^*). As a consequence, the first graph is also denoted as G. For example, if we define a model including two binary variables X and Y, so that $\mathrm{I}_X = \mathrm{I}_Y = \{1, 2\}$, and the associated first-order interactions $u_{XY}(ij)$; i, j =1, 2, then the set E contains the elements $u_{XY}(11)$, $u_{XY}(12)$, $u_{XY}(21)$, $u_{XY}(22)$. As a consequence, E^* contains the edges $\{X = 1, Y = 1\}$, $\{X = 1, Y = 2\}$, $\{X = 2, Y = 1\}$, and $\{X = 2, Y = 2\}$, which are multiple edges joining X with Y. The graph G= (V, E) is a multigraph, a graph with multiple edges without loops, which is determined by the first-order interactions. Clearly, there is a one-to-one correspondence between the first-order interactions and the edges, so that it is equivalent to partition over the parameters or over the edges.

The representation of $G = (V, E)$ consists of circles representing the variables, two of which may be united with an edge when a first-order interaction including those variables is part of the model. There are multiple edges between vertices according to the possible combinations of values of the associated variables. We note that the underlying simple graph G^u associated with the graph G obtained from an RGLL model is the independence graph associated with a graphical log-linear model (e.g., Bondy and Murty 1976, p. 103).

Symbolically, X and Y are in the same vertex class if and only if $u_X(l) = u_Y(l)$, for all $l = 1, ..., L$, where u_X and u_Y are main effects and L is the number of categories for variables X and Y. Similarly, two first-order interactions or their corresponding edges, one joining variable X with variable Y at the values (i_1, j_1) and another joining variable Z with variable R at the values (k_1, l_1), are in the same class if and only if $u_{XY}(i_1 j_1) = u_{ZR}(k_1 l_1)$, where $X, Y, Z, R \in V$; with $i_1 \in \mathrm{I}_X$, $j_1 \in \mathrm{I}_Y$, $k_1 \in \mathrm{I}_Z$, and $l_1 \in \mathrm{I}_R$; and where $u_{XY}(i_1 j_1)$ and $u_{ZR}(k_1 l_1)$ are first-order interactions.

To represent an RGLL model, the graph G is coloured using a different colour for each class, as in Figure 1, i.e. using a vertex and edge colouring. Given any coloured multigraph, we can identify the parameter restrictions corresponding to the associated RGLL model. Then all restrictions in an RGLL model can be identified with a vertex and edge colouring.

We have the following observations concerning Definition 1:

a) In the graphical representation of RGLL models, there are multiple edges between variables, so that if there is a clique, it is not important which edges are used to obtain it. For instance, if there is a clique corresponding to a triangle and there are 4 multiples edges between each pair of edges on it, then there are 64 possible different cliques.

b) When $|V_t| = 1$ or $|E_s| = 1$ for some s or t, where $t = 1, ..., T$ and $s = 1, ..., S$, we have atomic classes; otherwise, they are composite. Atomic classes are those where the corresponding parameters are not restricted, and an RGLL model with only atomic classes is a graphical log-linear model.

c) Independences between variables can be read off from the graph by using separator sets as in graphical log-linear models.

d) Interactions of order higher than one are not restricted.

It is important to notice that we are not assuming any identification constraint on the parameters. As it is defined, any set of restrictions is possible even though they may generate redundant equations among the likelihood equations. However, there are specific RGLL models that do not generate such redundant equations, for instance when the parameter restriction do not change when the identification constraint is effect coding instead of using the parameters without such a constraint. This occurs, for example, when the restrictions on the model are of the form $u_{XY}(ij) = u_{ZR}(ij)$ or $u_{XY}(ij) = u_{ZR}(ji)$ or when the restrictions are imposed on the main effects.

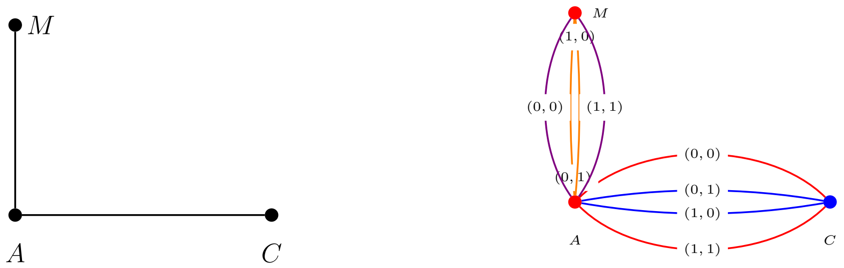
Example 1. Consider three binary variables A, C, and M with categories coded as 0 and 1 and the hierarchical log-linear model with generating class $\mathbf{A} = \{\{A, C\}, \{A, M\}\}$:

$$\log m(i, j, k) = u + u_A(i) + u_C(j) + u_M(k) + u_{AC}(ij) + u_{AM}(ik), \; i, j, k = 0, 1; \qquad (5)$$

and the equality restrictions

$$u_A(0) = u_M(0),\ u_A(1) = u_M(1),$$
$$u_{AM}(0,1) = u_{AM}(1,0),\ u_{AC}(0,1) = u_{AC}(1,0),$$
$$u_{AM}(0,0) = u_{AM}(1,1),\ u_{AC}(0,0) = u_{AC}(1,1).$$

Observe that model (5), without equality restrictions, is a graphical log-linear model with graph $G^u= (V^u, E^u)$, where $V^u= \{A, C, M\}$ and $E^u= \{\{A, C\}, \{A, M\}\}$, shown in Figure 2(a).



(a) Graphical model (b) RGLL model

Figure 2: Two hierarchical log-linear models for Example 1: (a) a graphical log-linear model; (b) an RGLL model.

According to the parameter restrictions, we see that the vertex set $V = \{A, C, M\}$ in the restricted model is partitioned as $V = (V_1, V_2)$, with $V_1 = \{A, M\}$ and $V_2 = \{C\}$. The set of first-order interactions or its corresponding edge set E is partioned as (E_1, E_2, E_3, E_4), where

$E = \{u_{AM}(00), u_{AM}(01), u_{AM}(10), u_{AM}(11), u_{AC}(00), u_{AC}(01), u_{AC}(10), u_{AC}(11)\}$, and $E_1 = \{u_{AM}(01), u_{AM}(10)\}$, $E_2 = \{u_{AC}(01), u_{AC}(10)\}$, $E_3 = \{u_{AM}(00), u_{AM}(11)\}$, and $E_4 = \{u_{AC}(00), u_{AC}(11)\}$.

We obtain the corresponding graph $G= (V, E)$ shown in Figure 2(b) whose underlying simple graph $G^u= (V^u, E^u)$ is shown in Figure 2(a). Using any of these graphs, we see that the generating class corresponds to the cliques set, which together with the restrictions on the parameters of the model define an RGLL model.

The following section shows symmetry and quasisymmetry as instances of RGLL models.

3. Symmetry and quasi-symmetry models as RGLL models

Symmetry and quasi-symmetry models were introduced to analyze square contingency tables with a symmetrical pattern. A square table formed by cells (i, j); for i, $j = 1, 2, ..., L$, satisfies symmetry if

$$m(i, j) = m(j, i), \text{ for all } i \neq j.$$

If we consider two variables X and Y, symmetry can be represented using the log-linear model

$$\log m(i, j) = u + u_X(i) + u_Y(j) + u_{XY}(ij) \tag{6}$$

with additional restrictions

$$u_X(i) = u_Y(i), \; i = 1, 2, ..., L;$$
$$u_{XY}(ij) = u_{XY}(ji), \; i, j = 1, 2, ..., L.$$

Quasi-symmetry is used to analyze cases where there is no symmetry due to marginal heterogeneity, which means that the main effects in the symmetry model differ. Such a model can be written as (6) with restrictions

$$u_{XY}(ij) = u_{XY}(ji), \; i, j = 1, 2, ..., L.$$

The parameterization used to define these models is not unique (e.g., Meiser, von Eye, and Spiel 1997), some parameterizations can be obtained through those log-linear models classified as non-standard as discussed by Mair (2007).

Example 2. Migration patterns data. For a sample of U.S. residents, Agresti (Agresti 2002, p. 423) presents some data given in Table 2 based on data by the U.S. Census Bureau that compare region of residence in 1985 with that of 1980. The variables involved are place of residence in 1980 and 1985. Each variable has four possible values: Northeast, Midwest, South and West. We can see from the observed counts that the table is more or less symmetric; however, the symmetry model is rejected because we get a deviance of 243.55 with 6 degrees of freedom so that the deviance is greater than the associated χ^2 quantile at a significance level $\alpha = 0.05$.

A closer look at the table reveals that 366 people moved from Northeast to South whereas only 172 people moved from South to Northeast. A model which fits well these data according to the deviance is the quasi-symmetry model. It means that the lack of symmetry in the table is caused by some marginal heterogeneity. In other words, the number of people emigrating from region i in 1980 to region j in 1985 would be similar to the one entering to i in 1985 from j in 1980 if the number of people in each region for 1980 were the same as the number of people in the same region in 1985. RGLL models fitted for these data are presented in section **4**.

Table 2: Observed counts[1] and fitted expected frequencies under symmetry[2] and quasi-symmetry[3] models for the data described in Example 2.

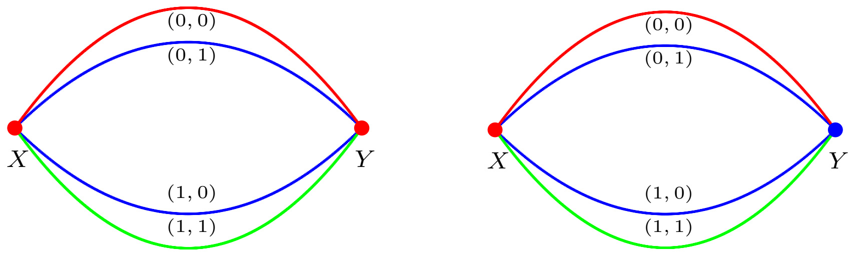
Residence in 1980	**Residence in 1985** Northeast	Midwest	South	West
	11607[1]	100	366	124
Northeast	11607[2]	93.50	269.00	93.50
	11607[3]	95.79	370.44	123.77
	87	13677	515	302
Midwest	93.50	13677	370.00	239.00
	91.21	13677	501.68	311.11
	172	225	17819	270
South	269.00	370.00	17819	278.00
	167.56	238.32	17819	261.12
	63	176	286	10192
West	93.50	239.00	278.00	10192
	63.23	166.89	294.88	10192

Symmetry and quasi-symmetry models can be expressed as RGLL models as follows. In the symmetry model there is only one vertex colour class formed by both vertices and there are L atomic edge colour classes, one for every $u_{XY}(ii)$ interaction, and $\binom{L}{2}$ edge colour classes for the interactions $u_{XY}(ij) = u_{XY}(ji)$, for $i \neq j$. In the quasi-symmetry model there are the same edge colour classes as in symmetry models, but every vertex belongs to a different atomic class.

As an example, consider two binary variables X and Y with categories 0 and 1, and the saturated log-linear model given in (6) for i,j= 0, 1.

Now consider the RGLL model with graph $G = (V, E)$ given in Figure 3(a). From G, we observe that $E = (E_1, E_2, E_3)$, where $E_1 = \{u_{XY}(00)\}$, $E_2 = \{u_{XY}(11)\}$, and $E_3 = \{u_{XY}(01), u_{XY}(10)\}$. Edges in E_3 belong to the same class, indicating that the corresponding interactions are identical. The remaining edges belong to different atomic colour classes. The vertex set, $V = \{X, Y\}$, is not partitioned. Then, the main effects corresponding to X and Y are the same for all the categories. The model associated with Figure 3(a) can be expressed as (6), for i,j= 0, 1, with restrictions

$$u_{XY}(ij) = u_{XY}(ji),\ \ i, j = 0, 1;$$
$$u_X(i) = u_Y(i),\ \ i = 0, 1.$$



(a) Symmetry model (b) Quasi-symmetry model

Figure 3: Colourings for models with two binary variables: (a) symmetry model and (b) quasi-symmetry model.

Hence, the symmetry model for V=$\{X, Y\}$ is the RGLL model generated by $\mathbf{A}$= $\{\{X, Y\}\}$ with vertex class V and edge set E partitioned into $E_1 = \{u_{XY}(00)\}$, $E_2 = \{u_{XY}(11)\}$, and $E_3 = \{u_{XY}(01),\ u_{XY}(10)\}$, whose graph is given in Figure 3(a).

On the other hand, consider the RGLL model whose graph $G = (V, E)$ is shown in Figure 3(b). From G, we have the same edge or first-order interactions partition as in the symmetry model. The vertex set $V = \{X, Y\}$ is partitioned into V_1 and V_2, with $V_1 = \{X\}$ and $V_2 = \{Y\}$, so that the main effects are not restricted. Then, the quasi-symmetry model for V=$\{X, Y\}$ given in (6), for i,j= 0, 1, with restrictions

$$u_{XY}(ij) = u_{XY}(ji),\ \ i, j = 0, 1;$$

is the RGLL model generated by $\mathbf{A}$= $\{\{X, Y\}\}$ with vertex set $V = (V_1, V_2)$, where $V_1 = \{X\}$ and $V_2 = \{Y\}$, and edge set $E = (E_1, E_2, E_3)$, where $E_1 = \{u_{XY}(00)\}$, $E_2 = \{u_{XY}(11)\}$, and E_3= $\{u_{XY}(01),\ u_{XY}(10)\}$, whose associated graph is given in Figure 3(b).

4. Likelihood equations

There is some additional notation that has to be introduced before obtaining the likelihood equations. We denote by n the vector of observed or cell counts $n(i)$, for $i \in \mathrm{I}$, and m corresponds to a vector containing the expected frequencies for each cell $m(i)$, for $i \in \mathrm{I}$, i.e. $m' = (m(i))_{i \in \mathrm{I}}$.

Consider an RGLL model with associated graph $G = (V, E)$ in which V is partitioned into T classes $V = (V_1, ..., V_T)$ with $T \in \{1, 2, ..., |V|\}$. It is useful to identify each variable within a vertex class V_t, for $t = 1, ..., T$, as v_k^t, with $k = 1,..., kver(t)$, where $kver$ is a vector identifying the number of vertices included in each class. Hence, $V_t = \left\{ v_1^t, ..., v_{kver(t)}^t \right\}$. We assume that each variable has L levels and define the associated parameters as

$u_{v_k^t}(l)$: main effect for variable v_k^t in category l, for t=1,..., T; k=1,..., $kver(t)$, and l=1,..., L.

For each vertex class V_t, for t=1,...,T, we obtain the following equality restrictions

$$u_{v_1^t}(l) = ... = u_{v_{kver(t)}^t}(l) = u_{V_t}(l), \text{ for } l = 1, ..., L;$$

where $u_{V_t}(l)$ is the parameter representing all the equated parameters associated with colour class V_t and level l.

We define the marginal total for the k-th variable in class V_t for level l, $n(v_k^t = l)$, as the sum of observed counts in all cells in which v_k^t takes value l:

$$n(v_k^t = l) = \sum_{z : z_{v_k^t} = l} n(z), \; l = 1, ..., L; \; k = 1, ..., kver(t); \; t = 1, ..., T.$$

We similarly define the corresponding marginal total for the expected frequencies $m(v_k^t = l)$.

Consider an RGLL model in which the first-order interactions set E is partitioned into S classes E_1,..., E_S. Each class $E_s = \left\{ e_1^s, ..., e_{ked(s)}^s \right\}$ for s=1, 2,..., S, where e_r^s corresponds to the r-th element within class E_s, for $r = 1, 2, ..., ked(s)$, with s=1,..., S, where ked is a vector consisting of the number of edges in each class. We denote $u_{l_r^s m_r^s}(i_r^s j_r^s)$ as the first-order interaction associated with the edge e_r^s, so that e_r^s is an edge joining variable l_r^s to variable m_r^s at the values (i_r^s, j_r^s). That is, l_r^s and m_r^s are the r-th variables for the class E_s in the first and second entries, respectively, of $u_{l_r^s m_r^s}(i_r^s j_r^s)$, whose associated values are i_r^s and j_r^s, respectively.

For each class E_s, for $s = 1, ..., S$, we obtain the following equality restrictions

$$u_{l_1^s m_1^s}(i_1^s j_1^s) = ... = u_{l_{ked(s)}^s m_{ked(s)}^s}(i_{ked(s)}^s j_{ked(s)}^s) = u_{E_s},$$

where u_{E_s} denotes the parameter representing all parameters in class E_s for levels i_r^s and j_r^s, with $r = 1, 2, ..., ked(s)$ and $s = 1, ..., S$.

The marginal total for e_r^s is defined as the sum of observed counts in all cells in which l_r^s and m_r^s take the values i_r^s and j_r^s, respectively:

$$n(l_r^s = i_r^s, m_r^s = j_r^s) = \sum_{\substack{z : (z_{l_r^s}, z_{m_r^s}) \\ = (i_r^s, j_r^s)}} n(z), \; r = 1, ..., ked(s); \; s = 1, ..., S.$$

The corresponding marginal total for the expected frequencies $m(l_r^s = i_r^s, m_r^s = j_r^s)$ is similarly defined. Finally, the marginal counts for a subset b of the set of variables V, $b \subset V$, for a specific slice i_b corresponds to

$$n_b(i_b) = \sum_{j:j_b=i_b} n(j).$$

Cell counts n may follow a Poisson, multinomial, or multinomial with fixed marginals sampling scheme. The first one corresponds to having independent Poisson-distributed random variables for each cell $i \in \mathrm{I}$. The second one corresponds to having a fixed sample size $|n|$, random counts, and observations that independently belong to cell i with probability $p(i)$, for $i \in \mathrm{I}$, where $p(i) \geq 0$ and $\sum_{i\in\mathrm{I}} p(i) = 1$. The third one corresponds to having counts in specific slices i_b which are independent and multinomially distributed.

Independently of the sampling scheme, the logarithm of the kernel of the likelihood function $L(m)$ is (e.g., Lauritzen 1996, p. 71)

$$\sum_{i\in\mathrm{I}} n(i)\log m(i) - \sum_{i\in\mathrm{I}} m(i). \tag{7}$$

When a hierarchical log-linear model (4) is considered, expression (7) becomes

$$\sum_{a\in K}\sum_{i_a} n_a(i_a)u_a(i_a) - \sum_{i\in\mathrm{I}} \exp\left(\sum_{a\in K} u_a(i_a)\right). \tag{8}$$

This expression is used to obtain the likelihood equations corresponding to an RGLL model by including the corresponding parameter restrictions.

For an RGLL model, expression (8) becomes

$$\begin{aligned}\sum_{t=1}^{T}\sum_{l=1}^{L}\sum_{k=1}^{kver(t)} n(v_k^t = l)u_{V_t}(l) + \sum_{s=1}^{S}\sum_{r=1}^{ked(s)} n(l_r^s = i_r^s, m_r^s = j_r^s)u_{E_s} + \\ \sum_{a\in K, |a|\neq 1,2}\sum_{i_a} n_a(i_a)u_a(i_a) - \sum_{i\in\mathrm{I}}\exp(\sum_{a\in K} u_a(i_a)).\end{aligned} \tag{9}$$

When expression (9) is differentiated with respect to each parameter and equated to zero, we obtain an equation system as follows:

$$\begin{aligned}\sum_{k=1}^{kver(t)} n(v_k^t = l) = \sum_{k=1}^{kver(t)} m(v_k^t = l),\ t = 1, ..., T;\ l = 1, ..., L; \\ \sum_{r=1}^{ked(s)} n(l_r^s = i_r^s, m_r^s = j_r^s) = \sum_{r=1}^{ked(s)} m(l_r^s = i_r^s, m_r^s = j_r^s),\ s = 1, 2, ..., S; \\ n_a(i_a) = m_a(i_a),\ \text{for all } a \in K,\ |a| \neq 1, 2;\end{aligned} \tag{10}$$

where, $m_a(i_a)$ denotes the marginal expected frequency for the slice i_a, i.e. $m_a(i_a) = \sum_{j:j_a=i_a} m(j)$.

Redundant equations can be eliminated by replacing the last equation in (10) with

$$n_a(i_a) = m_a(i_a), \text{ for all } a \in A, \ |a| \neq 1, 2.$$

Hence, the corresponding equation system is solved to obtain the maximum likelihood estimators of the expected frequencies m.

4.1. Solution of likelihood equations

There is a closed expression that solves the likelihood equations for a symmetry model; however, for a general RGLL model an approximated solution through numerical methods is necessary. Log-linear models are fitted using iterative proportional fitting, IPF (e.g., Bishop, Fienberg, and Holland 1975, p. 83-102), or the Newton Raphson (Fisher scoring) method (e.g., Agresti 2002, p. 143-146, p. 342-343). We have written programs to fit an RGLL model based only on the IPF method. The Newton Raphson method can be directly used when the restrictions can be equivalently used for parameters without identification constraints that consider all possible values of the variables (full parameters) or for the parameters under a parametrization. Instances of restrictions of such kind were described before Example 1 in Section **2**. In these cases, an RGLL model and its associated design matrix can be obtained using such a parameterization, and the associated design matrix has full rank. This is straightforward on models with vertex colouring only (main effects restrictions), but when there is an edge colouring (first-order interactions restrictions), equalities between parameters with different parameterizations are not always represented in the same way by simply changing the parameter. When we can not easily find such parametrization, we use the full parameters to obtain a system of equations where some of them are redundant. Hence, we can solve a subset corresponding to linearly independent equations and the rest are automatically solved. The modified IPF method described in the Appendix depends on those equations and can always be used.

A group of Fortran programs has been written to fit and select RGLL models, we refer to them as REGRAPH and are available upon request from the first author. The modified IPF method was implemented using some subroutines from Haberman (1972, 1976) to fit log-linear models, but most of the subroutines used were specifically written for RGLL models. Fitted expected frequencies, deviance, the associated design matrix, and the number of degrees of freedom for the associated asymptotic chi square distribution was also calculated. The number of degrees of freedom is calculated from the design matrix and is corrected when necessary, for instance, when an estimated expected frequency has a value of zero. Numerical results obtained with REGRAPH have been compared with those obtained by using *MIM 3.2.0.6* (Edwards 2000) or *R* for models in which the comparison is possible, e.g. quasi-symmetry and symmetry models.

4.2. Model selection

In order to fit an RGLL model to some data, three components have to be considered: (1) the graph structure given by the generating class, (2) the colouring or partition of the sets V and E, and (3) the estimated expected frequencies. Usually, the generating class is obtained performing a model search looking for a graphical log-linear model that fits the data using software like *MIM*. Then we apply a model search method to get a colouring, defining in this way an updated RGLL model. We obtained a method to see whether it is convenient to join colour classes using the deviance. Two classes are joined whenever its p value is greater than a significance level. With this method, we implemented an iterative search in REGRAPH to get, from any initial RGLL or graphical log-linear model, a new RGLL model that fits some data. Hence, only the last two components can be obtained through REGRAPH. The iterative search consists of joining pairs of vertex (or edge) classes from an initial model. We iteratively try to join all pairs of classes until a pair is significant according to its p value, obtaining new updated classes. The process is repeated with the new classes and so on until it is not possible to join more classes. We can first apply a model search for the vertex classes and use the final model obtained through this process as an initial model to obtain the edge classes.

Example 2 (continued). First, we label residence in 1980 and residence in 1985 as vertices 1 and 2, respectively. The four possible values, Northeast, Midwest, South and West, are coded as 1, 2, 3, and 4, respectively. The quasi-symmetry model fits these data and corresponds to an RGLL model with generating class $\mathbf{A}$={{1, 2}} with V=(V_1, V_2), E=(E_1,...,E_{10}), where

$$V_1 = \{1\},\ V_2 = \{2\};\ E_i = \{u_{12}(ii)\}, i = 1,..,4,\ E_5 = \{u_{12}(12), u_{12}(21)\},$$
$$E_6 = \{u_{12}(13), u_{12}(31)\},\ E_7 = \{u_{12}(14), u_{12}(41)\},\ E_8 = \{u_{12}(23), u_{12}(32)\},$$
$$E_9 = \{u_{12}(24), u_{12}(42)\},\ E_{10} = \{u_{12}(34), u_{12}(43)\}.$$

The deviance for this model obtained with REGRAPH is 2.99 with three degrees of freedom and a p-value of 0.39, indicating that we do not reject the null hypothesis that the model fits the data.

Starting from the quasi-symmetry model, we apply a selection method that preserves the vertex colour classes using REGRAPH. Assuming a significance level of 0.05, we get an RGLL model with generating class $\mathbf{A}$={{1, 2}} with $V = (V_1, V_2)$ and E=(E_1,E_2,...,E_7), where

$$V_1 = \{1\}, V_2 = \{2\};\ E_1 = \{u_{12}(22)\},\ E_2 = \{u_{12}(12), u_{12}(21)\},\ E_3 = \{u_{12}(13), u_{12}(31)\},$$
$$E_4 = \{u_{12}(14), u_{12}(41)\},\ E_5 = \{u_{12}(23), u_{12}(32)\},\ E_6 = \{u_{12}(33), u_{12}(11), u_{12}(34), u_{12}(43)\},$$
$$E_7 = \{u_{12}(44), u_{12}(24), u_{12}(42)\}.$$

This model has a deviance of 2.98 with three degrees of freedom, p-value of 0.39, while the value of the Pearson X^2 statistic is 2.98. Of these classes, only two differ with respect to those in the quasi-symmetry model. The class containing $u_{12}(34)$ and $u_{12}(43)$ now also contains $u_{12}(11)$ and $u_{12}(33)$ and the class containing $u_{12}(24)$ and $u_{12}(42)$ now also contains $u_{12}(44)$. The graphs associated with this and the quasi-symmetry model are not presented here because they involve too many edges and colours to be helpful.

Example 2 shows the fitting of a quasi-symmetry model as an RGLL model and offers a second fitting of another RGLL model. Both indicate that the expected cell frequencies in the same class would be equal if the marginal effects were not considered or that the variables and values in the same class are equally associated.

Example 3 (twins data, continued). In Section 1 we defined an RGLL model with two vertex and ten edge colour classes. The associated restrictions are as usual for the main effects and of the kind $u_{XY}(ij)$=$u_{ZR}(ij)$ or $u_{XY}(ij)$=$u_{ZR}(ji)$ for the first-order interactions. This allows us to equivalently apply the restrictions on the full parameters or on those under effect coding as described in Section **4.1**. Hence, the model can be fitted with R using the *glm* function by generating the effect coding variables and summing those whose associated full parameters are equal. Similar statistics were obtained with REGRAPH, which uses an IPF algorithm, and R, where a Newton-Raphson algorithm is used. Using R we got a residual deviance of 9.49 with 10 degrees of freedom, and p value of 0.49, while using REGRAPH the statistic is 9.18 with the same degrees of freedom, and a p value of 0.51. When compared with the graphical model (the unrestricted model defined only by the generating class) whose deviance is 4.75 with seven degrees of freedom, the deviance between models is 4.74 (or 4.43 with REGRAPH) with three degrees of freedom and associated p value of 0.19 (or 0.21 with

REGRAPH). Hence, the model fits the data well when compared with both the saturated or graphical models. It improves the inference because it allows the symmetric interpretations discussed earlier. The associated fitted expected values are shown in Table 1.

5. Discussion and perspectives

We have introduced RGLL models mainly as a way of generalizing symmetry in graphical log-linear models. We have used both the parameterization and the kind of restrictions used by Agresti (Agresti 2002, p. 423-426) for symmetry and quasi-symmetry models. We could have used any other parameterization to define the parameter restrictions in RGLL models; for example, parameters under effect coding. In some particular cases, the restrictions with any kind of parameter are exactly the same. The graphical representation can be helpful when there are few variables and levels. RGLL models can be fitted to specific panel data, which can be represented using contingency tables as discussed by Lovison (2000).

RGLL models can be fitted using REGRAPH. Care should be taken with the interpretation of a fitted model. Indeed, in some cases the fitted model could be too complex to be interpreted in terms of the relationship between the cells. We have to be careful with which restrictions are imposed, for instance if all first-order interactions are in one class, then the corresponding parameter becomes a constant term and the model is non-hierachical. Even though we proved in Ramírez-Aldana (2010) that the IPF method converges to the estimators, we have to be careful with the initial values. The problem with this type of numerical methods is that for higher-dimension contingency tables, the solution can slightly differ depending on those values. This can be more evident on contingency tables in which there are cells with very small counts and others with very large counts. Future work could include to improving the computational algorithm in this sense.

There are two types of models defined in Ramírez-Aldana (2010) for graphical log-linear models with associated triangle-free graphs called label and level invariant graphical log-linear models. They represent four types of symmetry: (a) graph symmetry, (b) model and distribution preservation after permuting subsets of variables, (c) expected frequencies equalities, and (d) scale invariance. They can be expressed as RGLL models though they have different properties. The RGLL model associated with a label invariant model can be easily obtained using graphical and algebraic concepts. The example given in Section **1** corresponds to a particular level invariant graphical log-linear model, which is almost equivalent to a label invariant model but without scale invariance. The advantage of such a model is that its associated RGLL model can be obtained using similar concepts as the ones used for a label invariant model. The classes for the RGLL model in Section 1 can be obtained in this way.

We could extend RGLL models by setting equality restrictions not only on the main effects and first-order interactions, but also on higher-order interactions. This generalization requires additional work in both computational and theoretical terms. A graphical representation of these models, even with multi-graphs, is not possible, much less when we consider that the equalities involve not only variables but also their values. If we obtained these generalizations, more symmetry generalizations, for example complete symmetry or quasi-symmetry (Bishop *et al.* 1975, p. 299-306), would be particular cases.

Acknowledgments

Most of this work corresponds to doctoral research from the first author, which was supported by a grant from the National Council of Science and Technology of Mexico (CONACYT). The authors have benefited from conversations with Dr. Johan Van Horebeek and with Prof. Steffen Lauritzen.

A. Solution to the likelihood equations

The IPF method (e.g., Lauritzen 1996, p. 83-84) consists of three steps applied to every cell i, for $i \in \mathrm{I}$. For RGLL models, there are likelihood equations that have to be solved in addition to the ones for the generating class $\mathbf{A}$. These equations correspond to first-order interactions and main effects, first and second equations in (10). The algorithm then solves RGLL models similarly to the IPF method; however, in step 2 we add transformations corresponding to the restrictions or colourings. The usual IPF method consists of the following steps:

1. Assign an initial value, $m_0(i)$, for $i \in \mathrm{I}$. For instance, $m_0(i)$ can be one.
2. Take all the elements in the generating class $\mathbf{A}$, and order them in a set $(b_1, b_2, ..., b_z)$. Define the transformation $T_v = T_{b_v}$, for $v = 1, 2, ..., z$, as

$$(T_{b_v}m)(i) = m(i)\frac{n(i_{b_v})}{m(i_{b_v})},\ i \in \mathrm{I}.$$

   Define recursively

$$m_{r+1}(i) = (T_1T_2...T_z)m_r(i),\ r = 0, 1, 2, ...$$

   In every step, we adjust all the elements in the generating class, so that there are z sub-steps for every step.
3. The process continues until the maximal difference between the marginal counts and the marginal adjusted expected frequencies reaches a predetermined error δ.

For RGLL models, we add other transformations. Considering that all variables have the same categories, L levels in each, we define for every vertex class V_t, where $t = 1, 2, ..., T$ and $l = 1, ..., L$, the following transformations

$$(T_{V_t(l)}m)(i) = m(i)\frac{\sum_{v_k^t \in V_t} n(v_k^t = l)}{\sum_{v_k^t \in V_t} m(v_k^t = l)},\ i \in \{(i_1, i_2, ..., i_{|V|}) \in \mathrm{I} | i_{v_k^t} = l,\ \text{for some } v_k^t \in V_t\}.$$

For each edge or first-order interaction class E_s, with $s = 1, 2, ..., S$, we define the following transformations for $i \in \{(q_1, q_2, ..., q_{|V|}) \in \mathrm{I} | q_{l_v} = i_v,\ q_{r_v} = j_v,\ \text{for some } \{l_v = i_v, r_v = j_v\} \in E_s\}$

$$(T_{E_s}m)(i) = m(i)\frac{\sum_{\{l_v = i_v, r_v = j_v\} \in E_s} n(l_v = i_v, r_v = j_v)}{\sum_{\{l_v = i_v, r_v = j_v\} \in E_s} m(l_v = i_v, r_v = j_v)}.$$

We define $0/0 = 0$. The element $\{l_v = i_v, r_v = j_v\} \in E_s$ denotes an edge on E_s joining the variable l_v to r_v for the value combination (i_v, j_v), for $v = 1, ..., ked(s)$, where $s = 1, ..., S$.

Then, we define for r, where $r = 0, 1, 2, ...$

$$m_{r+1}(i) = (T_1 T_2 ... T_z T_{V_1(1)} T_{V_1(2)} ... T_{V_1(L)} ... T_{V_T(1)} T_{V_T(2)} ... T_{V_T(L)} T_{E_1} T_{E_2} .... T_{E_S}) m_r(i).$$

We emphasize that the transformation $T_{b_v} = T_v$ is applied for $|b_v| \neq 1, 2$ and that not all transformations associated with the colourings are applied to every cell, it depends on the class and cell. For example, if we have a cell whose entries corresponding to all the variables in a colour class V_t are all different from l, then there is no sense in applying to this cell the transformation $T_{V_t(l)}$.

References

Agresti A (2002). *Categorical Data Analysis*. 2nd edition. John Wiley and Sons, New York.

Andersen E (1991). *The Statistical Analysis of Categorical Data*. 2nd edition. Springer-Verlag, Heidelberg.

Bishop YMM, Fienberg SE, Holland PW (1975). *Discrete Multivariate Analysis*. MIT Press, Cambridge, Massachusetts. Reprinted by Springer, New York, 2007.

Bondy JA, Murty USR (1976). *Graph Theory with Applications*. MacMillan Press, London.

Bowker AH (1948). "A Test for Symmetry in Contingency Tables." *Journal of the American Statistical Association*, **43 (244)**, 572–574.

Caussinus H (1965). "Contribution à l'Analyse Statistique des Tableux de Corrélation." *Annales de la Faculté des Sciences de Toulouse*, **29**, 77–182.

Drton M, Richardson TS (2008). "Binary Models for Marginal Independence." *Journal of the Royal Statistics Society Series B*, **70 (2)**, 287–309.

Edwards D (2000). *Introduction to Graphical Modelling*. 2nd edition. Springer-Verlag, New York. PC-program *MIM* by the same author and used in the book can be downloaded from http://www.hypergraph.dk.

Gottard A (2009). "*Graphical Log-linear Models for Homologous Factors*." Paper presented at S.Co.2009, Politecnico di Milano, Milano, Italy. Available from http://www2.mate.polimi.it/ocs/viewabstract.php?id=137&cf=7.

Gottard A, Marchetti GM, Agresti A (2011). "Quasi-symmetric Graphical Log-linear Models." *The Scandinavian Journal of Statistics*, **38 (3)**, 447–465.

Haberman SJ (1972). "Algorithm AS 51: Log-linear Fit for Contingency Tables." *Applied Statistics*, **21 (2)**, 218–225.

Haberman SJ (1976). "Correction to AS 51: Log-linear Fit for Contingency Tables." *Applied Statistics*, **25 (2)**, 193.

Højsgaard S, Lauritzen SL (2008). "Graphical Gaussian Models with Edge and Vertex Symmetries." *Journal of the Royal Statistics Society Series B*, **70 (5)**, 1005–1027.

Kateri M, Mohammadi F, Sturmfels B (2014). "A Family of Quasisymmetry Models." Available from arXiv:1403.0547v2.

Lauritzen SL (1996). *Graphical Models*. Clarendon Press, Oxford, England.

Lovison G (2000). "Generalized Symmetry Models for Hypercubic Concordance Tables." *International Statistical Review*, **68 (3)**, 323–338.

Mair P (2007). "A Framework to Interpret Nonstandard Log-linear Models." *Austrian Journal of Statistics*, **36 (2)**, 89–103.

Meiser T, von Eye A, Spiel C (1997). "Log-linear Symmetry and Quasi-symmetry Models for the Analysis of Change." *Biometrical Journal*, **39 (3)**, 351–368.

Ramírez-Aldana R (2010). *Restricted or Coloured Graphical Log-linear Models*. Ph.D. thesis, Graduate studies in Mathematics, National Autonomous University of Mexico, UNAM, Mexico. Available from http://132.248.9.195/ptb2011/enero/0666028/Index.html.

Tahata K, Yamamoto K, Nagatani N, Tomizawa S (2009). "A Measure of Departure from Average Symmetry for Square Contingency Tables with Ordered Categories." *Austrian Journal of Statistics*, **38 (2)**, 101–108.

Yamamoto H, Iwashita T, Tomizawa S (2007). "Decomposition of Symmetry into Ordinal Quasi-symmetry and Marginal Equimoment for Multi-way Tables." *Austrian Journal of Statistics*, **36 (4)**, 291–306.

Affiliation:

Ricardo Ramírez-Aldana
Department of Mathematics, Faculty of Sciences, UNAM
Circuito Exterior, CU,
04510, D.F. Mexico.
E-mail: ricardoramirezaldana@ciencias.unam.mx

Guillermina Eslava-Gómez
Department of Mathematics, Faculty of Sciences, UNAM
Circuito Exterior, CU,
04510, D.F. Mexico.
E-mail: eslava@ciencias.unam.mx

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 44 *Submitted:* 2014-09-23
December 2015 *Accepted:* 2015-04-16

Austrian Journal of Statistics
December 2015, Volume 44, 35–51.
http://www.ajs.or.at/
doi:10.17713/ajs.v44i4.54

A New Generalized Poisson-Lindley Distribution: Applications and Properties

Deepesh Bhati†
Central Unviersity of Rajasthan

D V S Sastry
Central Unviersity of Rajasthan

PZ Maha Qadri
Central Unviersity of Rajasthan

Abstract

A new generalized Poisson-Lindley distribution is obtained by compounding Poisson distribution with a two parameter generalised Lindley distribution. The new distribution is shown to be unimodal and over–dispersed. This distribution has a tendency to accommodate right tail, and for particular values of parameter, the tail tends to zero at a faster rate. Various properties like cumulative distribution function, generating function, moments etc. are derived. Knowledge about the parameters is obtained through method of moments, maximum likelihood method and EM algorithm. Moreover, an actuarial application to collective risk model is shown by considering the proposed distribution as primary distribution and exponential and Erlang as secondary distributions. The model is applied to real data sets and found to perform better than other competing models.

Keywords: compound distribution, Poisson-Lindley distribution, over–dispersion, EM algorithm, collective risk model, Bonus–Malus premium.

1. Introduction

Researchers in the recent past proposed many distributions, most popular being the generalizations of exponential, gamma and Weibull distributions. Lindley (1958) suggested a one parameter distribution to illustrate the difference between fiducial distribution and posterior distribution. Ghitany et al. (2008) studied many statistical properties of the Lindley distribution and showed that the mathematical properties are more flexible than exponential distribution, as such, Lindley distribution is found to be better model than exponential distribution. Many researchers viz. Zakerzadeh and Dolati (2009), Zamani and Ismail (2010), Nadarajah (2011), Bakouch et al. (2012), Elbatal et al. (2013), Hassan (2014), Shanker (2013) have proposed new classes of distributions by modifying the Lindley distribution and discussed various properties of their proposed generalizations.
Sankaran (1970) while modelling count data, introduced Poisson-Lindley distribution, by assuming the parameter of the Poisson distribution, λ, to follow a Lindley distribution. Ghitany et al. (2008) proposed different methods of estimation for the discrete Poisson-Lindley distribution. Ghitany et al. (2008) have used the zero truncated Poisson-Lindley distribution to model count data when the data structurally excludes zero counts. Size-biased Poisson-

Lindley Distribution and its application were also studied by Ghitany et al. (2008). Mahmoudi et al.(2010) generalized the Poisson-Lindley distribution of Sankaran (1970) and showed that their generalized distribution has more flexibility in analysing count data. Hernández et al. (2010) in the context of insurance have used the Poisson-Lindley to study the collective risk model when number of claims and size of a single claim are incorporated into the model. Another form of discrete Lindley distribution was introduced by Gómez et al. (2011) by discretizing the continuous Lindley distribution. The model is over-dispersed and competitive with the Poisson distribution to fit automobile claim frequency data sets. After revising some of its properties a compound discrete Lindley distribution is obtained. Gómez et al. (2013) analysed the collective risk model by assuming Erlang distribution to the loss data and generalised Lindley distribution to the claim frequency data.

In many potential branches like insurance, clinical trials, engineering, biology etc. the variable of interest is a count variable. Though classical models like Poisson, geometric, negative binomial, and their generalizations (see for example Philippou (1983), Gómez (2010, 2011), Chandra et al. (2013), Sastry et.al. (2014)), are available for count data analysis, it is found that these models are not supportive to capture the right tail behaviour of the data set properly. Few attempts have been made by researchers to capture this right tail behaviour (see for example Gomez et al. (2013)), but not much work has been done. The importance of right tail behaviour is illustrated in the following example.

In insurance industry, the actuary while fixing the premium, assumes a claim distribution which may or may not be realised after a claim is made, therefore it becomes a challenge to the actuary to develop new statistical distributions which take care of this subtle difference. Besides properly fitting to the available data set, the proposed distribution has to capture specific characteristics like shape and tail behaviour. The tail can either tapers out quickly or slowly. When the tail approaches zero slowly, the extreme values can be properly understood. Therefore, besides fitting properly to the data, the shape of the distribution should capture the behaviour of extreme values also with their tendency to approach zero.

This critical observation has motivated the authors to search for a discrete distribution which fades away to zero much more slowly/faster than the classical compound Poisson-Lindley distribution by Sankaran (1970).

In this paper we propose a new generalized Poisson-Lindley distribution which is obtained from Poisson distribution when its parameter λ, follows a two parameter Lindley distribution ($\mathcal{TPLD}(\theta, \alpha)$) as suggested by Shanker et al. (2013) defined as

$$f\left(x;\alpha,\theta\right)=\frac{\theta^{2}}{(\theta+\alpha)}(1+\alpha x)e^{-\theta x},\qquad x,\alpha,\theta>0$$

The paper is structured as follows: In Section 2, a new distribution is proposed. Various properties like shape, quantile function, generating function and moments are presented in Section 3. An algorithm for simulation of random variable is shown in Section 4. Estimation methods like method of moments, maximum likelihood estimation and EM algorithm are discussed in Section 5. Section 6 addresses the actuarial application to auto mobile insurance of the proposed model. In Section 7, applicability of the proposed model is shown, and compared with other competing probability models. Moreover, by using Bayesian methodology, Bonus-Malus premium for automobile insurance product is being computed and shown in same Section 7.

2. Proposed model

Definition 1: A random variable X is said to be a new generalized Poisson-Lindley distribution if it follows the stochastic representation

$$X|\lambda \sim Po(\lambda)$$

$$\lambda|\theta, \alpha \sim \mathcal{TPLD}(\theta, \alpha)$$

for $\lambda > 0$ and $\theta, \alpha > 0$. We denote unconditional distribution of by $\mathcal{NGPL}(\theta, \alpha)$.

Theorem 1: If $X \sim \mathcal{NGPL}(\theta, \alpha)$, then probability mass function (pmf) of X is

$$f(x; \theta, \alpha) = \frac{\theta^2}{(\theta+\alpha)(1+\theta)^{x+1}} \left(1 + \frac{\alpha(x+1)}{(1+\theta)}\right), x = 0, 1, 2, ...$$

with $\theta, \alpha > 0$.

Proof: If $X|\lambda \sim Po(\lambda)$ and $\lambda|\theta, \alpha \sim \mathcal{TPLD}(\theta, \alpha)$, then pmf of unconditional random variable X is given as

$$Pr(X = x) = \int_0^\infty Pr(X = x|\lambda) f(\lambda; \theta, \alpha) d\lambda$$

where $f(\lambda; \theta, \alpha)$ is two parameter Poisson Lindley distribution.

$$\begin{aligned}
Pr(X = x) &= \int_0^\infty \frac{\theta^2 e^{-\lambda} \lambda^x (1 + \alpha\lambda) e^{-\theta\lambda}}{(\theta+\alpha)x!} d\lambda \\
&= \frac{\theta^2}{(\theta+\alpha)x!} \int_0^\infty e^{-\lambda(1+\theta)} \lambda^x d\lambda + \frac{\alpha\theta^2}{(\theta+\alpha)x!} \int_0^\infty e^{-\lambda(1+\theta)} \lambda^{x+1} d\lambda \\
&= \frac{\theta^2}{(\theta+\alpha)x!} \frac{\Gamma(x+1)}{(1+\theta)^{x+1}} + \frac{\alpha\theta^2}{(\theta+\alpha)x!} \frac{\Gamma(x+2)}{(1+\theta)^{x+2}} \\
Pr(X = x) &= \frac{\theta^2}{(\theta+\alpha)(1+\theta)^{x+1}} \left(1 + \frac{\alpha(x+1)}{(1+\theta)}\right), x = 0, 1, 2, ... \qquad (1)
\end{aligned}$$

where $\theta > 0$, $\alpha > 0$.

Remarks

- i For $\alpha \to 0$, pmf (1) reduces to geometric distribution $\mathcal{G}\left(\frac{\theta}{1+\theta}\right)$.
- ii For $\alpha = 1$, pmf (1) reduces to discrete Poisson-Lindley distribution proposed by Shankaran (1970).
- iii For $\alpha \to \infty$, pmf (1) tends to negative binomial $\mathcal{NB}(2, \frac{\theta}{1+\theta})$.
- iv (1) can also viewed as a mixture of $\mathcal{G}(\frac{\theta}{1+\theta})$ and $\mathcal{NB}(2, \frac{\theta}{1+\theta})$ with mixing proportion $\frac{\theta}{\theta+\alpha}$.

For different values of θ and α, the probability function is evaluated and presented in Figure 1. It can be observed that for fixed θ and increasing α, the distribution accommodates more right tail, whereas, for fixed α and increasing θ, the distribution condenses and the right tail approaches to zero at a faster rate. Our proposed model fits appropriately to those data sets where there is large right tail or the tail approaches to zero at a faster rate. Such data sets are quite common in insurance problems and count data example in biology.

The cumulative density function of $X \sim \mathcal{NGPL}(\theta, \alpha)$ can be given as

$$\begin{aligned}
F_X(x) &= \sum_{n=0}^{x} \frac{\theta^2}{(\theta+\alpha)} \frac{(1+\theta+\alpha n+\alpha)}{(1+\theta)^{n+2}} \\
&= \frac{(\alpha+\theta)(\theta+1)^{x+2} - (2\alpha\theta+\alpha+\theta^2+\theta+\alpha\theta x)}{(\theta+1)^{x+2}(\alpha+\theta)} \qquad (2)
\end{aligned}$$

3. Properties of the new generalized Poisson-Lindley distribution

3.1. Shape of the probability function

It can be seen that

$$p(0) = \frac{\theta^2(1+\alpha+\theta)}{(\alpha+\theta)(1+\theta)^2}$$

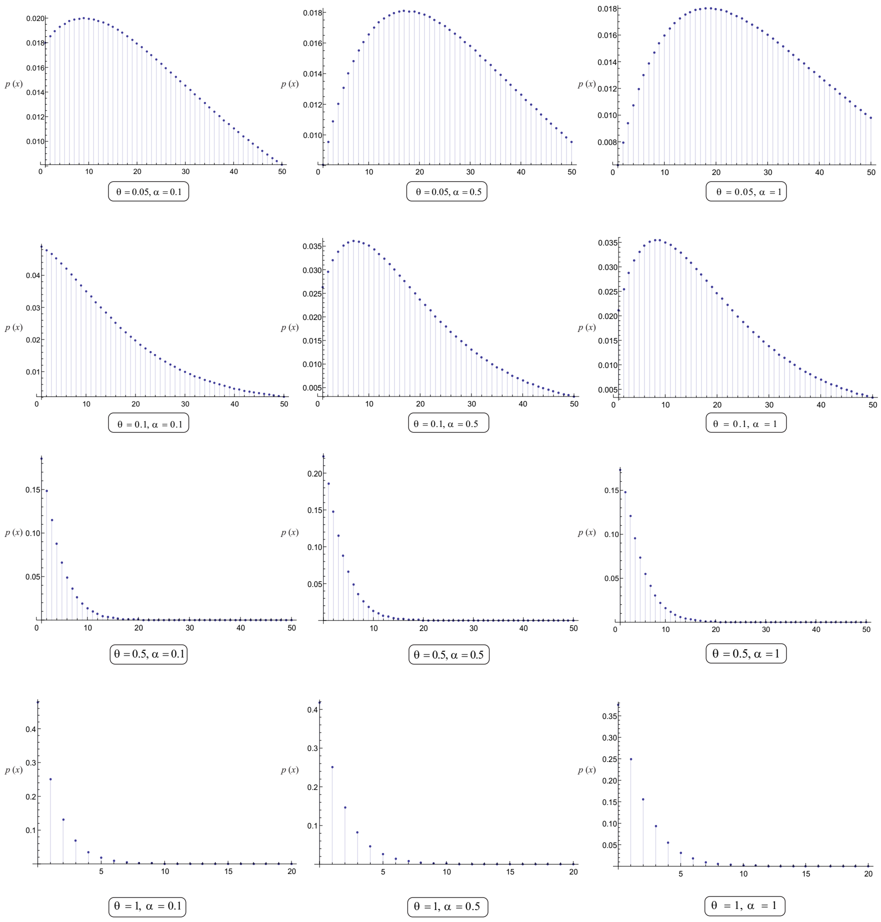


Figure 1: Pmf plots for different values of θ and α

$$\frac{p(x+1)}{p(x)} = \frac{1+(2+x)\alpha+\theta}{(1+\theta)(1+(1+x)\alpha+\theta)}, \qquad x = 1,2,3,\cdots \tag{3}$$

The above expression is a decreasing function in x implying unimodalilty. Further,

$$\frac{Pr(X=x+2)Pr(X=x)}{(Pr(X=x+1))^2} = \frac{1+\frac{\alpha}{1+\theta+\alpha(x+2)}}{1+\frac{\alpha}{1+\theta+\alpha(x+1)}} < 1 \tag{4}$$

As such $Pr(X = x)$ is log-concave and hence the distribution (1) has an increasing failure rate, see Johnson et al.(2005), p. 209. It can also be verified that,

(i) For $\theta > 1$, (1) is unimodal and has a unique mode at 0.

(ii) When $0 < \theta \leq 1$ and $\alpha \geq \frac{\theta(\theta+1)}{1-\theta}$, the mode of pmf (1) is unimodal and unique at $[x^*]+1$, where $[x^*]$ denotes integer part of x^* defined as

$$x^* = \frac{\alpha-\theta-\theta\alpha-\theta^2}{\alpha\theta}$$

(iii) For $0 < \theta \leq 1$ and $\alpha \leq \frac{\theta(\theta+1)}{1-\theta}$, the pmf (1) has mode at 0.

The above facts are shown in Fig. 1 for selected values of θ, α.

Further, if we define a ratio $R = \frac{Pr(N_1=n;\theta,\alpha)}{Pr(N_2=n,\theta)} = \frac{1+\frac{1+\alpha x}{\alpha+\theta}}{1+\frac{1+x}{1+\theta}}$, where $N_1 \sim \mathcal{NGPL}(\theta,\alpha)$ and $N_2 \sim \mathcal{PL}(\theta)$(i.e. Poisson-Lindley), then we can observe, $\mathcal{NGPL}$ will have heavier (thinner) right tail probability as compare to $\mathcal{PL}(\theta)$ i.e $R > (<)1$ for $\alpha > (<)1$ and $x > (<)x_0(= 1/\theta)$.

3.2. Quantile function

Let $X \sim \mathcal{NGPL}(\theta,\alpha)$, then the quantile function $Q_X(\gamma) = F_X^{-1}(\gamma)$, $0 < \gamma < 1$ is x_γ given in
Theorem 2: The quantile function of the $\mathcal{NGPL}(\theta,\alpha)$ is

$$x_\gamma = -\frac{\left(2\alpha\theta+\alpha+\theta^2+\theta\right)\log(\theta+1)+W\left(-\frac{(1-\gamma)(\theta+1)^{-\frac{\alpha+\theta^2+\theta}{\alpha\theta}}(\alpha+\theta)\log(\theta+1)}{\alpha\theta}\right)}{\alpha\theta\log(\theta+1)}$$

where $W(.)$ denotes the Lambert W function.

Proof: The c.d.f of the distribution is

$$F_X(x) = \frac{(\alpha+\theta)(\theta+1)^{x+2}-\left(2\alpha\theta+\alpha+\theta^2+\theta+\alpha\theta x\right)}{(\theta+1)^{x+2}(\alpha+\theta)}$$

the γ^{th} quantile function is obtained by solving $F_X(x) = \gamma$.

$$\gamma = 1-\frac{\alpha\theta x+2\alpha\theta+\theta+\theta^2+\alpha}{(1+\theta)^{x+2}(\alpha+\theta)}$$

$$1-\gamma = \frac{\alpha\theta x+2\alpha\theta+\theta+\theta^2+\alpha}{(1+\theta)^{x+2}(\alpha+\theta)}$$

Multiplying both sides of the above equation by $-\frac{(1+\theta)^{\frac{-\theta-\theta^2-\alpha}{\alpha\theta}}}{\alpha\theta}\log(1+\theta)$ and after rearranging the terms, we get

$$-\frac{(\alpha+\theta)(1-\gamma)(1+\theta)^{\frac{-\theta-\theta^2-\alpha}{\alpha\theta}}}{\alpha\theta}\log(1+\theta) = -\frac{\left(2\alpha\theta+\alpha\theta x+\alpha+\theta+\theta^2\right)}{\alpha\theta}\log(1+\theta)$$
$$*e^{-\frac{\left(2\alpha\theta+\alpha\theta x+\alpha+\theta+\theta^2\right)}{\alpha\theta}\log(1+\theta)}$$

Using definition of Lambert-W function ($W(z)e^{W(z)} = z$, where z is a complex number, see Jońda (2010)), the right hand side of the above expression is the Lambert W function of the real argument $-\frac{(\alpha+\theta)(1-\gamma)(1+\theta)^{\frac{-\theta-\theta^2-\alpha}{\alpha\theta}}}{\alpha\theta}\log(1+\theta)$. Thus we have

$$W\left(-\frac{(1-\gamma)(\theta+1)^{-\frac{\alpha+\theta^2+\theta}{\alpha\theta}}(\alpha+\theta)\log(\theta+1)}{\alpha\theta}\right) = -\frac{1}{\alpha\theta}\log(1+\theta)\left(2\alpha\theta+\alpha\theta x+\alpha+\theta+\theta^2\right)$$

Solving the above equation for x, we get

$$x_\gamma = F^{-1}(\gamma) = -\frac{(2\alpha\theta+\alpha+\theta^2+\theta)\log(\theta+1) + W\left(-\frac{(1-\gamma)(\theta+1)^{-\frac{\alpha+\theta^2+\theta}{\alpha\theta}}(\alpha+\theta)\log(\theta+1)}{\alpha\theta}\right)}{\alpha\theta\log(\theta+1)} \tag{5}$$

The first three quantiles can be obtained by substituting $\gamma = \frac{1}{4}, \frac{1}{2}$ and $\frac{3}{4}$ in equation (5).

$$\begin{aligned}
Q_1 &= -\frac{2\alpha\theta+\alpha+\theta^2+\theta}{\alpha\theta} - \frac{W\left(-\frac{3(\theta+1)^{-\frac{\alpha+\theta^2+\theta}{\alpha\theta}}(\alpha+\theta)\log(\theta+1)}{4\alpha\theta}\right)}{\log(\theta+1)} \\
Median = Q_2 &= -\frac{2\alpha\theta+\alpha+\theta^2+\theta}{\alpha\theta} - \frac{W\left(-\frac{(\theta+1)^{-\frac{\alpha+\theta^2+\theta}{\alpha\theta}}(\alpha+\theta)\log(\theta+1)}{2\alpha\theta}\right)}{\log(\theta+1)} \\
Q_3 &= -\frac{2\alpha\theta+\alpha+\theta^2+\theta}{\alpha\theta} - \frac{W\left(-\frac{(\theta+1)^{-\frac{\alpha+\theta^2+\theta}{\alpha\theta}}(\alpha+\theta)\log(\theta+1)}{4\alpha\theta}\right)}{\log(\theta+1)}
\end{aligned}$$

3.3. Generating functions

The Probability Generating Function of (1) is given by

$$\begin{aligned}
P_X(t) &= \sum_{x=0}^{\infty} t^x \frac{\theta^2}{(\theta+\alpha)(1+\theta)^{x+1}}\left(1+\frac{\alpha(x+1)}{(1+\theta)}\right) \\
&= \left(\frac{\theta^2}{\theta+\alpha}\right)\left(\frac{\alpha\theta+\alpha+\theta^2+2\theta-\theta t-t+1}{(\theta+1)(-\theta+t-1)^2}\right) \\
P_X(t) &= \frac{\theta^2(\alpha+\theta-t+1)}{(\alpha+\theta)(\theta-t+1)^2}
\end{aligned} \tag{6}$$

The Moment Generating Function works out to

$$\begin{aligned}
M_X(t) &= \sum_{x=0}^{\infty} e^{tx} \frac{\theta^2}{(\alpha+\theta)(1+\theta)^{x+1}}\left(1+\frac{\alpha(x+1)}{(1+\theta)}\right) \\
&= -\left(\frac{-\alpha-\theta+e^t-1}{(-\theta+e^t-1)^2}\right)\left(\frac{\theta^2}{\theta+\alpha}\right) \\
M_X(t) &= \frac{\theta^2\left(\alpha+\theta-e^t+1\right)}{(\alpha+\theta)\left(\theta-e^t+1\right)^2}
\end{aligned} \tag{7}$$

3.4. Moments, skewness and kurtosis

The first four raw moments of X can easily be obtained by simple computations and are as follows

$$\mu_1' = \frac{2\alpha+\theta}{\theta(\alpha+\theta)}$$

$$\mu_2' = \frac{2\alpha(\theta+3)+\theta(\theta+2)}{\theta^2(\alpha+\theta)}$$
$$\mu_3' = \frac{\theta\left(\theta^2+6\theta+6\right)+2\alpha\left(\theta^2+9\theta+12\right)}{\theta^3(\theta+\alpha)}$$
$$\mu_4' = \frac{2\alpha\left(\theta^3+21\theta^2+72\theta+60\right)+\theta\left(\theta^3+14\theta^2+36\theta+24\right)}{\theta^4(\theta+\alpha)} \quad (8)$$

and the moments of X about the mean are

$$\mu_2 = \frac{2\alpha^2(1+\theta)+\theta^2(1+\theta)+\alpha\theta(4+3\theta)}{\theta^2(\alpha+\theta)^2}$$
$$\mu_3 = \frac{2\alpha^3\left(\theta^2+3\theta+2\right)+\alpha^2\theta\left(5\theta^2+18\theta+12\right)+\alpha\theta^2\left(4\theta^2+15\theta+12\right)+\theta^3\left(\theta^2+3\theta+2\right)}{\theta^3(\alpha+\theta)^3}$$

The expressions for skewness $\left(\frac{\mu_3}{\mu_2^{3/2}}\right)$ and kurtosis $\left(\frac{\mu_4}{\mu_2^2}\right)$ are large and complicated; however their values for different parametric values can be determine and are presented in Table 1. Moreover, Figure 2 and 3, respectively, represents the surface plot of skewness and kurtosis with α and θ as axes.

Table 1: Skewness (Kurtosis) for various value of parameter (θ, α)

$\theta \downarrow, \alpha \rightarrow$	0.5	1	1.5	2	2.5	3
0.25	1.556(6.676)	1.488(6.336)	1.464(6.225)	1.453(6.171)	1.446(6.139)	1.442(6.119)
0.50	1.708(7.460)	1.598(6.830)	1.550(6.586)	1.525(6.460)	1.509(6.383)	1.498(6.332)
1.00	1.916(8.210)	1.792(7.765)	1.723(7.188)	1.679(6.984)	1.650(6.851)	1.628(6.758)
1.50	2.063(8.240)	1.949(7.832)	1.875(7.547)	1.824(7.351)	1.787(7.212)	1.759(7.108)
2.00	2.186(7.989)	2.083(7.895)	2.010(7.734)	1.956(7.593)	1.915(7.479)	1.884(7.387)
2.50	2.297(7.675)	2.203(7.848)	2.132(7.823)	2.078(7.755)	2.035(7.682)	2.001(7.614)
3.00	2.401(7.387)	2.314(7.765)	2.246(7.865)	2.191(7.872)	2.148(7.846)	2.112(7.810)

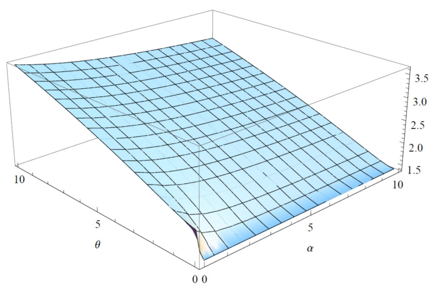


Figure 2: Skewness

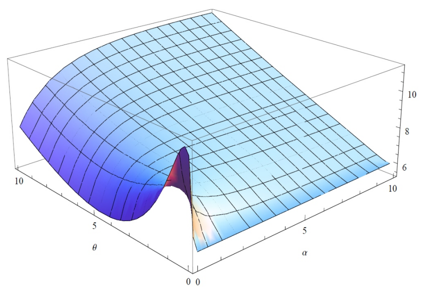


Figure 3: Kurtosis

It can be seen from the above that for fixed θ, as α increases, both skewness and kurtosis are decreasing. For fixed α, as θ increases, skewness is increasing while kurtosis is showing a mixed behaviour.

The coefficient of variation (C.V) of the distribution comes out to be

$$C.V. = \frac{\sigma}{\mu}$$
$$= \left(\sqrt{\frac{2\alpha^2(1+\theta)+\theta^2(1+\theta)+\alpha\theta(4+3\theta)}{\theta^2(\alpha+\theta)^2}}\right)\left(\frac{\theta(\alpha+\theta)}{2\alpha+\theta}\right)$$
$$C.V. = \frac{\sqrt{2\alpha^2(1+\theta)+\theta^2(1+\theta)+\alpha\theta(4+3\theta)}}{(2\alpha+\theta)}$$

The index for dispersion is given by

$$\begin{aligned} r &= \frac{\sigma^2}{\mu} \\ &= \left(\frac{2\alpha^2(1+\theta)+\theta^2(1+\theta)+\alpha\theta(4+3\theta)}{\theta^2(\alpha+\theta)^2}\right)\left(\frac{\theta(\alpha+\theta)}{2\alpha+\theta}\right) \\ r &= \frac{2\alpha^2(\theta+1)+\alpha\theta(3\theta+4)+\theta^2(\theta+1)}{\theta(\alpha+\theta)(2\alpha+\theta)} \end{aligned}$$

As the new distribution comes from a mixture of a Poisson distribution and two parameter Lindley distribution, the variance–to–mean ratio is greater than one (see Karlis (2005) and Sundt and Vernic (2009), p. 66) the proposed distribution is over-dispersed.

4. Simulation of random variables

Note that, If $Y \sim P(\lambda)$ and λ follows a two parameter Lindley distribution, which can also be represented as a mixture of two independent random variable $V_1 \sim exp(\theta)$ and $V_2 \sim gamma(2,\theta)$ with mixture parameter $p = \frac{\theta}{\theta+\alpha}$, then following steps can be used to generate $\mathcal{NGPL}(\theta,\alpha)$ random variates

Step 1: Generate $U_i, i = 1, 2, \cdot, n$ from $U(0,1)$ distribution.

Step 2: If $U_i \leq \frac{\theta}{\theta+\alpha}$, generate $\lambda_i \sim exp(\theta)$; otherwise, generate $\lambda_i \sim gamma(2,\theta)$

Step 3: Generate $Y_i, i = 1, 2, \cdots, n$, where $Y_i \sim P(\lambda_i)$.

5. Estimation

5.1. Method of moments

Given a random sample $x_1, x_2, \cdots, x_n$ of size n from (1), the moment estimates, $\tilde{\alpha}$ and $\tilde{\theta}$, of α and θ can be obtained by solving the following equations

$$m_1 = \mu_1' = \frac{2\alpha+\theta}{\theta(\alpha+\theta)} \tag{9}$$

and

$$m_2 = \mu_2' = \frac{2\alpha(\theta+3)+\theta(\theta+2)}{\theta^2(\alpha+\theta)}. \tag{10}$$

where m_1 and m_2 are the first and second sample moments. Solving the above equations, we get

$$\tilde{\alpha} = \frac{\tilde{\theta} - m_1\tilde{\theta}^2}{m_1\tilde{\theta} - 2} \qquad and \qquad \tilde{\theta} = \frac{2m_1 + \sqrt{4m_1^2 + 2m_1 - 2m_2}}{m_2 - m_1} \tag{11}$$

Theorem 3: For fixed α, the estimator $\tilde{\theta}$ of θ is positively biased, i.e. $E(\tilde{\theta}) > \theta$.
Proof: Let $\tilde{\theta} = g(\bar{X})$ and $g(t) = \frac{1-\alpha t+\sqrt{1+6\alpha t+\alpha^2 t^2}}{2t}$ for $t > 0$. Then

$$g''(t) = \frac{1}{t^3} + \frac{9\alpha t + 3\alpha^3 t^3 + 15\alpha^2 t^2 + 1}{t^3\left(1+6\alpha t+\alpha^2 t^2\right)^{3/2}} > 0$$

therefore, $g(t)$ is strictly convex. Thus, by Jensenś inequality, we have $E\left\{g(\bar{X})\right\} > g\{E(\bar{X})\}$. Finally, since $g\{E(\bar{X})\} = g\left(\frac{2\alpha+\theta}{\theta(\alpha+\theta)}\right) = \theta$, we obtain $E(\tilde{\theta}) > \theta$.

Theorem 4: For fixed values of α, the moment estimate $\tilde{\theta}$ of θ is consistent and asymptotically normal and distributed as

$$\sqrt{n}\left(\tilde{\theta}-\theta\right) \to_d N\left(0,\nu^2(\theta)\right)$$

where $\nu^2(\theta)=\frac{\theta^2(\alpha+\theta)^2\left(2\alpha^2\theta+2\alpha^2+3\alpha\theta^2+4\alpha\theta+\theta^3+\theta^2\right)}{\left(2\alpha^2+4\alpha\theta+\theta^2\right)^2}$.

Proof: Consistency: Since $\mu<\infty$, $\bar{X}\to_P \mu$. Also since $g(t)=\frac{1-\alpha t+\sqrt{1+6\alpha t+\alpha^2t^2}}{2t}$ is a continuous function at $t=\mu$, $g(\bar{X})\to_P g(\mu)$ i.e. $\tilde{\theta}\to_P\theta$.
Asymptotic normality: Since $\sigma^2<\infty$, then by the central limit theorem, we have

$$\sqrt{n}\left(\bar{X}-\mu\right)\to_d N\left(0,\sigma^2\right)$$

Also, since $g(\mu)$ is differentiable and $g'(\mu)\neq 0$, by the delta-method, we have

$$\sqrt{n}\left(g\left(\bar{X}\right)-g(\mu)\right)\to_d N\left(0,\nu^2(\theta)\right)$$

where

$$\begin{aligned}\nu^2(\theta) &= [g'\left(\frac{2\alpha+\theta}{\theta(\alpha+\theta)}\right)]^2\sigma^2\\ &= \frac{\theta^2(\alpha+\theta)^2\left(2\alpha^2\theta+2\alpha^2+3\alpha\theta^2+4\alpha\theta+\theta^3+\theta^2\right)}{\left(2\alpha^2+4\alpha\theta+\theta^2\right)^2}\end{aligned}$$

Hence the theorem.
As a result of this, the asymptotic $100(1-\gamma)\%$ confidence interval for θ is given by $\hat{\theta}\mp z_{\alpha/2}\frac{\nu(\hat{\theta})}{\sqrt{n}}$, where $z_{\alpha/2}$ is the $(1-\alpha/2)$ percentile of the standard normal distribution.

5.2. Maximum likelihood estimation

Let $x_1,x_2,\cdots,x_n$ be random observations of size n from our proposed generalized Poisson-Lindley distribution. The log-likelihood function for the vector of parameters $\Theta=(\theta,\alpha)^\top$ can be written as

$$l_n\left(\alpha,\theta\,|\underline{x}\right)=2n\log\theta-n\log(\theta+\alpha)-\sum_{i=1}^{n}\left(x_i+1\right)\log(1+\theta)+\sum_{i=1}^{n}\log\left(1+\frac{\alpha\left(x_i+1\right)}{1+\theta}\right) \quad (12)$$

The associated score function is given by $U_n=\left(\frac{\partial l_n}{\partial\theta},\frac{\partial l_n}{\partial\alpha}\right)^\top$, where

$$\frac{\partial l_n}{\partial\theta} = \frac{2n}{\theta}-\frac{n}{\theta+\alpha}-\frac{\left(\sum_{i=1}^{n}x_i\right)+n}{1+\theta}-\sum_{i=1}^{n}\frac{\alpha\left(x_i+1\right)}{\left(1+\theta+\alpha\left(x_i+1\right)\right)\left(1+\theta\right)} \quad (13)$$

$$\frac{\partial l_n}{\partial\alpha} = \frac{-n}{\theta+\alpha}+\sum_{i=1}^{n}\frac{\left(x_i+1\right)}{\left(1+\theta+\alpha\left(x_i+1\right)\right)} \quad (14)$$

the ML estimator can be obtained by equating $\left(\frac{\partial l_n}{\partial\theta},\frac{\partial l_n}{\partial\alpha}\right)^\top_{(\hat{\theta},\hat{\alpha})}=(0,0)^\top$, hence we get

$$\hat{\alpha}=\frac{\hat{\theta}^2\bar{x}-\hat{\theta}}{2-\hat{\theta}\bar{x}} \qquad \textit{and} \qquad \frac{-n}{\hat{\theta}+\hat{\alpha}}+\sum_{i=1}^{n}\left(\frac{x_i+1}{1+\hat{\theta}+\hat{\alpha}\left(x_i+1\right)}\right)=0$$

The second partial derivatives are given by

$$\begin{aligned}
\frac{\partial^2 l}{\partial \theta^2} &= \sum_{i=1}^{n} \frac{\alpha\,(x_i+1)\,(\alpha+2\theta+\alpha x_i+2)}{(\theta+1)^2\,(\alpha+\theta+\alpha x_i+1)^{\,2}} + \sum_{i=1}^{n} \frac{x_i+1}{(\theta+1)^2} + \frac{n}{(\alpha+\theta)^2} - \frac{2n}{\theta^2} \\
\frac{\partial^2 l}{\partial \theta \partial \alpha} &= \frac{n}{(\alpha+\theta)^2} - \sum_{i=1}^{n} \frac{x_i+1}{(\alpha+\theta+\alpha x_i+1)^{\,2}} \\
\frac{\partial^2 l}{\partial \alpha^2} &= -\sum_{i=1}^{n} \frac{(x_i+1)^2}{(\theta+1)^2\left(\frac{\alpha(x_i+1)}{\theta+1}+1\right)^2} + \frac{n}{(\alpha+\theta)^2} \qquad (15)
\end{aligned}$$

The above equations can be solved numerically by using open source R software.

5.3. EM algorithm for a new generalize Poisson-Lindley distribution

The MLE of α and θ needs to be computed numerically. Newton-Raphson algorithm is one of the standard methods to estimate the parameters. The algorithm requires second order derivatives of the log-likelihood for all iterations which makes it quite complex. This problem can be handled by another estimation method known as Expectation-Maximization (EM) (see Dempster et al. (1977)). The EM algorithm consists of two steps: the E-step and the M-step. **E-Step** computes the expectation of the unobservable part given the current values of the parameters and **M-step** maximizes the complete data likelihood and updates the parameters using the conditional expectations obtained in E-step. This procedure can be useful when there are no closed-form expressions for estimating the parameters and the derivatives of the likelihood are complicated.

To start with, a hypothetical complete-data distribution is defined with joint probability function

$$g(X,\lambda;\theta,\alpha) = \frac{\theta^2(1+\alpha\lambda)\lambda^x e^{-\lambda(1+\theta)}}{(\theta+\alpha)x!}, \qquad \theta>0;\alpha>0$$

It is straightforward to verify that the E-step of an EM cycle requires the computation of the conditional expectations of $(\lambda|x_i;\theta^{(h)},\alpha^{(h)})$ and $(\frac{\lambda}{1+\alpha\lambda}|x_i;\alpha^{(h)},\theta^{(h)})$, say $t_i^{(h)}$ and $s_i^{(h)}$ respectively, where $(\theta^{(h)},\alpha^{(h)})$ is the current estimate of (θ,α). Using,

$$g(\lambda|X) = \frac{\lambda^x(1+\alpha\lambda)e^{-\lambda(1+\theta)}(1+\theta)^{x+2}}{x!(1+\theta+\alpha(x+1))}$$

$$t_i^{(h)} = E(\lambda|x_i;\theta^{(h)},\alpha^{(h)}) = \frac{(x_i+1)\left(1+\theta^{(h)}+\alpha^{(h)}\,(x_i+2)\right)}{(1+\theta^{(h)})\,(1+\theta^{(h)}+\alpha^{(h)}\,(x_i+1))}$$

$$s_i^{(h)} = E\left(\frac{\lambda}{1+\alpha\lambda}|x_i;\theta^{(h)},\alpha^{(h)}\right) = \frac{x_i+1}{1+\theta^{(h)}+\alpha^{(h)}(\theta^{(h)}+1)}$$

The EM cycle completes with the M-step, involving complete data maximum likelihood over (θ,α), with the missing λ's replaced by their conditional expectations $E(\lambda|x;\theta^(h);\alpha^(h))$. Thus the EM iterates

$$\begin{aligned}
\theta^{(h+1)} &= \frac{-\left(\bar{t}^{(h)}\alpha^{(h)}-1\right)+\sqrt{\left(\bar{t}^{(h)}\alpha^{(h)}-1\right)^2+8\bar{t}^{(h)}\alpha^{(h)}}}{2\bar{t}^{(h)}} \\
\alpha^{(h+1)} &= \frac{1}{\bar{s}^{(h)}} - \theta^{(h)}
\end{aligned}$$

until defined convergence criterion is satisfied.

6. Collective risk model

In non-life insurance portfolio, say, automobile insurance, the aggregate loss (S) is a random variable defined as sum of claims incurred in a certain period of time. Let N be number of claims in a certain period which is a random variable and X_i be claim severity random variable, which is independent of N, is the size of i^{th} claim. Thus, aggregate loss is defined as $S = \sum_{i=0}^{N} X_i$. It is well known that the pdf of S is given as $f_s(x) = \sum_{n=0}^{\infty} p_n f^{n*}(x)$, where p_n denotes the probability of n claims (primary distribution) and $f^{n*}(x)$ is the n^{th} fold convolution of $f(x)$, the claim amount (secondary distribution). For more detail on classic risk model, see Freifelder (1974), Rolski et al. (1999), Nadarajah and Kotz (2006a and 2006b) and Klugman et al. (2012) and reference therein. Here, we consider two such situations: In one case, the primary distribution is as defined in Section 1 and claim severity distribution as exponential distribution with parameter (λ). In the second case, the Erlang distribution with parameters r and λ is the secondary distribution. Erlang loss distribution may arise in insurance settings when the individual claim amount is the sum of exponentially distributed claims.

Theorem 5: If we assume a $\mathcal{NGPL}(\theta, \alpha)$ as primary distribution and an exponential distribution with parameter (λ) as secondary distribution, then the pdf of aggregate loss random variable $S = \sum_{i=0}^{N} X_i$ is given by

$$f_S(x) = \frac{\theta^2 \lambda e^{-\frac{\theta\lambda x}{\theta+1}} \left(2\alpha\theta + 2\alpha + \theta^2 + 2\theta + \alpha\lambda x + 1\right)}{(1+\theta)^4(\theta+\alpha)}, \qquad for \qquad x > 0 \tag{16}$$

whereas,

$$f_S(0) = \frac{\theta^2 \left(1 + \alpha + \theta\right)}{(\theta+\alpha)(1+\theta)^2}$$

Proof: By assuming that the claim severity follows an exponential distribution with parameter $\lambda > 0$, since the n^{th} fold convolution of exponential distribution is gamma distribution with parameter n and λ, the n^{th} fold convolution is given by

$$f^{*n}(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}, \qquad n = 1, 2, ...,$$

Then the pdf of the random variable S is given by

$$\begin{aligned}
f_S(x) &= \sum_{n=1}^{\infty} \frac{\theta^2}{(\theta+\alpha)(1+\theta)^{n+1}} \left(1 + \frac{\alpha(n+1)}{1+\theta}\right) \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} \\
&= \frac{\theta^2 e^{-\lambda x}}{(\theta+\alpha)} \sum_{n=1}^{\infty} \frac{\lambda^n}{(1+\theta)^{n+1}(n-1)!} x^{n-1} \left(1 + \frac{\alpha(n+1)}{1+\theta}\right) \\
&= \frac{\theta^2 e^{-\lambda x}}{(\theta+\alpha)} \frac{e^{\frac{x\lambda}{1+\theta}} \lambda \left(1 + 2\alpha + x\alpha\lambda + 2\theta + 2\alpha\theta + \theta^2\right)}{(1+\theta)^4} \\
f_S(x) &= \frac{\theta^2 \lambda e^{-\frac{\theta\lambda x}{1+\theta}} \left(2\alpha\theta + 2\alpha + \theta^2 + 2\theta + \alpha\lambda x + 1\right)}{(\theta+\alpha)(1+\theta)^4}
\end{aligned}$$

The pdf of the aggregate loss has a jump of size $Pr(S = 0)$ at the origin. The no claim probability comes out to be

$$f_S(0) = \frac{\theta^2(1 + \alpha + \theta)}{(\theta+\alpha)(1+\theta)^2}$$

It is well known in the actuarial literature that the mean of the aggregate loss random variable S can be obtained as $E(S) = E(X)E(N)$ (see eqn. 9.9 of Klugman(2012)). Hence, when $X \sim exp(\lambda)$ then

$$E(S) = \frac{2\alpha + \theta}{\lambda\theta(\alpha+\theta)} \tag{17}$$

Theorem 6: In collective risk model, if the primary distribution follows $\mathcal{NGPL}(\theta, \alpha)$ and the secondary distribution follows an Erlang(2,λ) distribution, then the pdf of aggregate loss random variable $S = \sum_{i=0}^{N} X_i$ is given by

$$f_S(x) = \frac{\theta^2 e^{-\lambda x}\lambda\left((\theta+1)(3\alpha+2\theta+2)\sinh\left(\frac{\lambda x}{\sqrt{\theta+1}}\right)+\alpha\sqrt{\theta+1}\lambda x\cosh\left(\frac{\lambda x}{\sqrt{\theta+1}}\right)\right)}{2(\theta+\alpha)(\theta+1)^{7/2}} \quad for \quad x>0 \tag{18}$$

whereas

$$f_S(0) = \frac{\theta^2\,(1+\alpha+\theta)}{(\theta+\alpha)(1+\theta)^2}$$

Proof: By assuming that the claim severity follows an Erlang(2,λ) distribution, the n^{th} fold convolution of exponential distribution is gamma distribution with parameter $2n$ and λ. The n^{th} fold convolution is given by

$$f^{*n}(x) = \frac{\lambda^{2n}}{(2n-1)!}x^{2n-1}e^{-\lambda x}, \qquad n=1,2,\cdots,$$

Then, the pdf of the aggregate loss random variable S is given by

$$\begin{aligned}
f_S(x) &= \sum_{n=1}^{\infty}\frac{\theta^2}{(\theta+\alpha)(1+\theta)^{n+1}}\left(1+\frac{\alpha(n+1)}{1+\theta}\right)\frac{\lambda^{2n}}{(2n-1)!}x^{2n-1}e^{-\lambda x}\\
&= \frac{\theta^2 e^{-\lambda x}}{(\theta+\alpha)}\sum_{n=1}^{\infty}\frac{\lambda^{2n}}{(1+\theta)^{n+1}(2n-1)!}x^{2n-1}\left(1+\frac{\alpha(n+1)}{1+\theta}\right)\\
&= \frac{\theta^2 e^{-\lambda x}}{(\theta+\alpha)}\frac{\lambda\left((\theta+1)(3\alpha+2\theta+2)\sinh\left(\frac{\lambda x}{\sqrt{\theta+1}}\right)+\alpha\sqrt{\theta+1}\lambda x\cosh\left(\frac{\lambda x}{\sqrt{\theta+1}}\right)\right)}{2(\theta+1)^{7/2}}\\
f_S(x) &= \frac{\theta^2 e^{-\lambda x}\lambda\left((\theta+1)(3\alpha+2\theta+2)\sinh\left(\frac{\lambda x}{\sqrt{\theta+1}}\right)+\alpha\sqrt{\theta+1}\lambda x\cosh\left(\frac{\lambda x}{\sqrt{\theta+1}}\right)\right)}{2(\theta+\alpha)(\theta+1)^{7/2}}
\end{aligned}$$

The pdf of the aggregate loss random variable has a jump of size $Pr(S=0)$ at the origin. The probability of zero claim comes out to be

$$f_S(0) = \frac{\theta^2(1+\alpha+\theta)}{(\theta+\alpha)(1+\theta)^2}$$

7. Application to real data set

In this section, we fit our proposed distribution to 3 data sets so as to illustrate our claim that our proposed model fits well when compared to other competing models. The first data set has a long right tail and approaches to zero slowly and in the other data sets the right tail approaches zero at a faster rate. In one of the illustrations, we propose a method for calculation of automobile insurance premium under Bonus–Malus system.

Illustration 1: A dataset representing epileptic seizure counts (see Chakraborty (2010)) are used. Comparability of the proposed model with other distributions like Poisson distribution ($\mathcal{P}$), negative binomial ($\mathcal{NB}$), Generalized Poisson-Lindley ($\mathcal{GPL}$) and Weighted Generalized Poisson ($\mathcal{WGP}$) Distribution (Chakraborty (2010)) have been shown in Table 2. In each of these distributions, the parameters are estimated by using the maximum likelihood method. Further, based on the values of log likelihood and chi-square, we observe that New Generalized Poisson-Lindley ($\mathcal{NGPL}$) distribution provides a satisfactorily better fit for the data set compared to other distributions.

It can be seen that the log-likelihood and Chi-square statistic for the $\mathcal{NGPL}$ are lower that those of competing models showing that our model satisfactorily fits better to the data set.

Table 2: Distribution of epileptic seizure counts

	Observed	Expected frequency				
Count	frequency	$\mathcal{P}$	$\mathcal{NB}$	$\mathcal{WGP}$	$\mathcal{GPL}$	$\mathcal{NGPL}$
0	126	74.94	91.00	118.11	121.51	122.00
1	80	115.71	86.60	95.81	92.00	91.00
2	59	89.34	63.37	59.89	59.00	58.74
3	42	46.00	42.57	34.49	35.10	35.22
4	24	17.75	27.60	19.24	20.10	20.52
5	8	5.48	17.60	10.59	11.18	11.22
6	5	1.41	10.50	5.81	6.10	6.39
7	4	0.31	6.52	3.18	3.30	3.25
8	3	0.06	5.00	3.88	2.71	2.50
Total	351	351	351	351	351	351
	parameter	$\hat{\lambda}$=1.544	$\hat{r}$=1.757,	$\hat{a}$=1.089,	$\hat{\theta}$=1.139,	$\hat{\theta}$ =1.116,
			$\hat{p}$=.463	$\hat{s}$=-1,$\hat{b}$=.295	$\hat{\alpha}$=1.292	$\hat{\alpha}$ = 2.9061
	Log likelihood	-636.05	-595.22	-595.83	-594.61	-594.48
	Chi square	256.54	22.53	7.12	5.94	5.75

Illustration 2: Klugman et al. (2012, pp. 664) presented the distribution of automobile insurance policyholders according to number of claims. We reproduce the data set in Table 3. It is found that this data set is highly right skewed and over–dispersed (variance is greater than mean). We mentioned in the text that our proposed model can be applied to such over-dispersed data sets. We chose chi-square statistic $\chi^2 = \sum_n \left((O_n - E_n)^2/E_n\right)$ for comparison purposes and calculated this statistic for other competing models Poisson, Negative Binomial, Poisson-Lindley, Generalized Poisson-Lindley Distribution.

Table 3: Number of claims in automobile insurance

	Observed	Expected frequency				
Claim count	frequency	$\mathcal{P}$	$\mathcal{NB}$	$\mathcal{PL}$	$\mathcal{GPL}$	$\mathcal{NGPL}$
0	1563	1544.153	1566.40	1569.525	1566.407	1564.504
1	271	299.772	261.500	256.341	261.384	264.251
2	32	29.1	40.146	41.340	40.202	39.686
3	7	1.883	5.990	6.600	5.985	5.589
4	2	0.0914	0.880	1.044	0.874	0.870
Total	1875	1875	1875	1875	1875	1875
	parameter	$\hat{\lambda}$=0.19413	$\hat{\tau}$= 6.135,	$\hat{\theta}$=5.898,	$\hat{\theta}$=6.6715,	$\hat{\theta}$ =7.8747
			$\hat{a}$=1.191		$\hat{\alpha}$=1.1648	$\hat{\alpha}$=8.835
	χ^2-value	57.04	3.610	3.874	3.655	3.489

From the above calculations, it appears that our proposed model fits better to the data set.

Illustration 3: It is known that Poisson distribution is not a suitable choice for automoble insurance claims because of the restriction that mean equals variance for Poisson distribution. As such, Negative Binomial distribution is preferred in this case. In many countries Bonus–Malus system (BMS) is operative wherein policyholders with no claims are given bonus whereas policyholders with claims are punished. Therefore, next year premium depends on the history of the policyholder till this year irrespective of size of the claim. Gómez and

Vázquez (2003) calculated the premium under BMS by

$$P_{t+1} = 100\frac{E_{f(\lambda|\boldsymbol{n})}(\delta(\lambda))}{E_{f(\lambda)}(\delta(\lambda))}$$

where $\delta(\lambda) = \sum_{n=0}^{\infty} nP(N = n|\lambda)$, $f(\lambda)$ is the prior distribution and $f(\lambda|\boldsymbol{n})$ is the posterior distribution. Thus for sequence of independent and identical distributed claims $\boldsymbol{n} = (n_1, n_2, \cdots, n_t)$, we can see that the posterior distribution is easy to obtain by dividing the mixing distribution by the marginal distribution as follows.

$$f(\lambda|\boldsymbol{n}) = \frac{P(n|\lambda)f(\lambda)}{\int_0^\infty P(n|\lambda)f(\lambda)d\lambda}$$

Thus,

$$P_{t+1} = 100\frac{\frac{n+1}{t+\theta}\left(1+\frac{\alpha}{t+\theta+\alpha(n+1)}\right)}{\frac{1}{\theta}\left(1+\frac{\alpha}{\theta+\alpha}\right)} \tag{19}$$

Using the data presented in Table 3, we have computed the BMPs which are shown in Table 5. Finally, while calculating the premium under BMS, Lemaire (1979) remarked on the problem of overcharges. It can be seen that $\mathcal{NGPL}$ model produces a lower penalization as compared to traditional Poisson-Gamma models (as shown in Table 4 and 5)

Table 4: BMP using Poisson-Gamma model

$t \downarrow n \rightarrow$	0	1	2	3	4
0	100				
1	85.9845	158.1799	230.3752	302.5704	374.7657
2	75.4148	138.7355	202.0561	265.3768	328.6974
3	67.1592	123.5483	179.9372	236.3262	292.7152
4	60.5328	111.3580	162.1832	213.0084	263.8336

Table 5: BMP using new generalized Poisson-Lindley distribution

$t \downarrow n \rightarrow$	0	1	2	3	4
0	100				
1	86.9991	154.7229	217.6116	278.5636	338.5467
2	76.7979	137.7936	194.5001	249.4317	303.4587
3	68.6009	124.0580	175.6913	225.6937	274.8505
4	61.8849	112.7019	160.0944	205.9848	251.0835

8. Final comments

This paper provides a new generalized discrete distribution with an infinite and non-negative integer support. It has been shown that the proposed distribution can be considered as an alternative to well known distribution like Poisson, Poisson-Lindley, negative binomial, weighted generalized Poisson Distribution for the dataset possessing a right tail which approach to zero at a slower/faster rate. Further, in the context to actuarial science, close expression of pdf of aggregate loss random variable is obtained by considering our proposed distribution as primary distribution and exponential and Erlang distributions as secondary distributions. Moreover, our proposed model can also be useful in the determination of Bonus–Malus premium for non-life insurance products.

9. Acknowledgements

The authors acknowledge Editor, Matthias Templ, of this journal for constant encouragement to finalize the paper. Further, the authors acknowledge profound thanks to anonymous referee for giving critical comments which have immensely improved the presentation of the paper.

References

[1] Bakouch H.S., Al-Zahrani B.M., Al-Shomrani A.A., Marchi V.A., and Louzada F. (2012). An Extended Lindley distribution. *Journal of the Korean Statistical Society*, **41**, 75–85.

[2] Chakraborty S. (2010). On Some Distributional Properties of the Family of Weighted Generalized Poisson Distribution. *Communications in Statistics - Theory and Methods*, **39 (15)**, 2767–2788.

[3] Chandra N.K., Roy D. and Ghosh T. (2013). A Generalized Poisson Distribution. *Communications in Statistics Theory and Methods*, **42**, 2786–2797.

[4] Dempster A.P., Laird N.M. and Rubin D. (1977). Maximum Likelihood for Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical society*, **39**, 1–38.

[5] Elbatal I., Merovci F., and Elgarhy M. (2013). A New Generalized Lindley distribution. *Mathematical Theory and Modeling*, **3 (13)**.

[6] Freifelder, F. (1974). Statistical Decision Theory and Procedures. In P.M. Kahn, ed. *Credibility Theory and Applications*, Academic Press, New-York, 71–88.

[7] Ghitany M.E., Atieh B. and Nadarajah S. (2008). Lindley Distribution and Its Application. *Mathematics and Computers in Simulation*, **78**, 493–506.

[8] Ghitany M.E., Atieh B. and Nadarajah S. (2008). Zero-truncated Poisson- Lindley Distribution and Its Application. *Mathematics and Computers in Sim- ulation*, **79**, 279–287.

[9] Ghitany, M.E. and Al-Mutairi, D.K. (2008). Size-biased Poisson-Lindley Distribution and Its Application. *METRON International, Journal of Statistics*, **LXVI (3)**, 299–31.

[10] Ghitany M.E. and Al-Mutairi D.K. (2009). Estimation Methods for the Discrete Poisson-Lindley Distribution. *Journal of Statistical Computation and Simulation*, **79 (1)**, 1–9.

[11] Gómez D.E. and Ojeda E.C. (2011). The Discrete Lindley Distribution: Properties and Applications. *Journal of Statistical Computation and Simulation*, **81 (11)**, 1405–1416.

[12] Gómez D.E. and Calderín-Ojeda E. (2013). The Compound DGL/Erlang Distribution in the Collective Risk Model. *Journal of Quantitative Methods for Economics and Business Administration*, **16**, 121–142.

[13] Gómez-Déniz E., Hernández A., Fernández M.P. (2013). A Suitable Discrete Distribution for Modelling Automobile Claim Frequencies. *Bulletin of the Malaysian Mathematical Sciences Society.* (in press).

[14] Gómez-Déniz E. (2010). Another Generalization of the Geometric Distribution. *Test*, **19**, 399–415.

[15] Gómez-Déniz E., Sarabia J.M., Calderín-Ojeda E. (2011). A New Discrete Distribution with Actuarial Applications. *Insurance, Mathematics and Economics*, **48**, 406–412

[16] Gómez E., Vázquez F. (2003). Robustness in Bayesian Model for Bonus-Malus Systems. In: *Intelligent and Other Computation Techniques in Insurance. Theory and Applications.* World Scientific.

[17] Hassan M.K. (2014). On the Convolution of Lindley Distribution. *Columbia In ternational Publishing Contemporary Mathematics and Statistics*, **2 (1)**,47–54.

[18] Hernández-Bastida A., Fernández-Snchez M.P. and Gómez-Déniz E. (2010). Collective Risk Model: Poisson-Lindley and Exponential Distributions for Bayes Premium and Operational Risk. *Journal of Statistical Computation and Simu- lation*, **81 (6)**, June 2011, 759–778.

[19] Johnson N.L., Kotz S. and Kemp A.W. (2005). *Univariate Discrete Distribu- tions, Second Edition*. John Wiley and Sons.

[20] Jośda P. (2010). Computer Generation of Random Variables with Lindley or Poisson-Lindley Distribution via the Lambert W function. *Mathematics and Computers in Simulation*, **81**, 851–859.

[21] Karlis D. (2005). EM Algorithm for Mixed Poisson and Other Discrete Distributions. *Astin Bulletin*, **35 (1)**, 3–24.

[22] Klugman S., Panjer H., Willmot G. (2012). *Loss Models. From Data to Decisions*. John Wiley and Sons, New York.

[23] Lemaire J. (1979). How to Define a Bonus-Malus System with an Exponential Utility Function. *Astin Bulletin*, **10**, 274–282.

[24] Lindley D.V. (1958). Fiducial Distributions and Bayes' Theorem. *Journal of the Royal Statistical Society*, **20 (1)**, 102–107.

[25] Mahmoudi E. and Zakerzadeh H. (2010). Generalized Poisson-Lindley Distribution. *Communications in Statistics - Theory and Methods*, **39 (10)**, 1785–1798.

[26] Nadarajah S., Bakouch H.S., Tahmasbi R. (2011). A Generalized Lindley Distribution. *Sankhya B*, **73**, 331–359.

[27] Nadarajah S. and Kotz S. (2006a). Compound Mixed Poisson Distributions I. *Scandinavian Actuarial Journal*, **3**, 141–162.

[28] Nadarajah S. and Kotz S. (2006b). Compound Mixed Poisson Distributions II. *Scandinavian Actuarial Journal*, **3**, 163–181.

[29] Philippou A.N., Georghiou C. and Philippou G.N. (1983). A Generalized Geometric Distribution and Some of Its Properties. *Statistics and Probability Letters*, **1**, 171–175

[30] Rolski T., Schmidli H. Schmidt V. and Teugel J. (1999). *Stochastic Processes for Insurance and Finance*. John Wiley & Sons.

[31] Sankaran M. (1970). The Discrete Poisson-Lindley Distribution. *Biometrics*, **26 (1)**, 145–149.

[32] Sastry D.V.S., Bhati D., Rattihalli R.N., Gmez-Dniz E. (2014). On Zero Distorted Generalised Geometric Distribution. *Communications in Statistics: Theory and Methods*. (In press).

[33] Shanker R., Sharma S. and Shanker R. (2013). A Two-Parameter Lindley Distribution for Modeling Waiting and Survival Times Data. *Applied Mathematics*, **4**, 363–368.

[34] Sundt B., Vernic R. (2009). *Recursions for Convolutions and Compound Dis- tributions with Insurance Applications*. SpringerVerlag, New York.

[35] Zakerzadeh H. and Dolati A. (2009). Generalized Lindley Distribution. *Journal of Mathematical Extension*, **3 (2)**, 13–25.

[36] Zamani Hossein and Ismail Noriszura (2010). Negative Binomial-Lindley Distribution and Its Application. *Journal of Mathematics and Statistics*, **6 (1)**, 4–9.

Affiliation:

† Corresponding Author
Deepesh Bhati
Department of Statistics
Central University of Rajasthan
NH-8 Jaipur Ajmer Highway, Kishangarh-Ajmer, India
E-mail: deepesh.bhati@curaj.ac.in

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 44 *Submitted:* 2014-07-28
December 2015 *Accepted:* 2015-06-08

Austrian Journal of Statistics
December 2015, Volume 44, 53–65.
http://www.ajs.or.at/
doi:10.17713/ajs.v44i4.69

On a Class of Optimal Type Covariate Adjusted Response Adaptive Allocations for Normal Treatment Responses

Uttam Bandyopadhyay
Calcutta University

Rahul Bhattacharya
Calcutta University

Abstract

In the context of clinical trials for comparing two treatments, a new class of covariate adjusted response adaptive procedures is developed to achieve a balance between clinical optimality and inferential precision. Assuming normally distributed responses, several exact and asymptotic properties are studied and compared with a reasonable alternative under the presence of treatment covariate interaction.

Keywords: response adaptive design, covariate adjusted allocation, optimality.

1. Introduction

Response adaptive allocation is a useful technique in clinical trials to skew the allocation of incoming subjects towards the better performing treatment by using the available information. Assuming homogeneity among the subjects, a number of response adaptive allocation designs were developed in the last few decades. But heterogeneity with respect to some covariate (e.g. age, sex etc.) is quite obvious in trials involving human beings. The presence of covariate information often influences the responses of the subjects and this suggests to incorporate such information into the process of randomization. Therefore, a sensible allocation design should randomize any entering subject on the basis of the data accrued thus far together with his/her covariate information. Such allocations are referred to as *covariate adjusted allocation designs* (Hu and Rosenberger, 2006).
A number of attempts (see, for example, Begg and Iglewicz (1980), Atkinson (1982) and Wong and Zhu (2008)) is made to incorporate covariates into the randomisation process using the Efron's biased coin methodology in the framework of optimum design theory. But these schemes do not allow to incorporate the information of the previous treatment outcomes, and hence are not response adaptive and consequently lacks the ethical requirement of reducing the assignment fraction to the inferior treatment. However, in the context of continuous response trials, only a few response adaptive procedures incorporates covariate information, in addition. Although Bandyopadhyay and Biswas (2001) and Biswas et al. (2006) incorporated available responses, allocations and covariate information in randomisation, ignored the current covariate information. Therefore, the allocation designs developed with covari-

ates either minimize the variability of the estimated treatment effect incorporating covariates (see, for example, Begg and Iglewicz (1980), Atkinson (1982) and Wong and Zhu (2008)) without caring allocation of patients or favour the better performing treatment for allocation (see, for example, Bandyopadhyay and Biswas (2001) and Biswas et al. (2006)) accepting a significant loss in statistical precision. Although Jennison and Turnbull (2000) provided an unified approach to find an allocation ratio as a trade off between the ethical need of optimizing a clinically relevant criteria (e.g., number of allocations to the inferior treatment, total expected response or number of failures) and preservation of power for testing the equality of treatment effects, ignored the possibility of any covariate information. Along the line of Jennison and Turnbull (2000), Biswas and Mandal (2004) made an attempt to incorporate covariate information with an aim to minimise treatment failure but ignored the assessment of the so claimed optimality. It is worth mentioning at this point that, recently Azriel et al. (2012) derived an optimal proportion considering only the maximization of power of a test of equality of treatment effects but in the context of covariate independent binary response trials.

The existence of interaction between treatments and covariates is a common phenomena in any clinical trial. But the effect of such interactions under a response adaptive set up is not explored completely. In fact, most of the works related to covariate either ignored the possibility of such interactions (e.g., Bandyopadhyay and Biswas (2001), Biswas and Mandal (2004)) or, if incorporated, avoid to assess the effect in details. However, the ethical goal consists always in assigning a greater fraction of subjects to the better performing treatment independently of the presence or absence of treatment-covariate interaction. But the only difference is that in the presence of such interactions, the benefit of the treatments depends on the kind of patients (Friedman et al. (2010)). The present work, in the presence of treatment covariate interaction, is an attempt to develop a class of target allocation proportions by combining the views of both statistician and clinician when the response variable is continuous. The relevant measures of optimality, optimal target proportions and their implementation, along with some limiting results, are all given in Section 2. In Section 3 we explore and compare the small sample properties of the proposed allocation for a hypothetical clinical trial. Section 4 concludes with a discussion of some related issues.

2. The proposed allocation

2.1. Clinical optimality

Suppose two treatments, denoted by A and B , are under a clinical trial and the patients appear in the clinic sequentially. Each patient is treated once, the responses are immediate and a lower response indicates a favourable situation. Let Y_{ki} denote the potential outcome of the ith subject on treatment k, $k = A, B$ with $\mathbf{Z}_i$ as the associated p component vector of covariates. In practical situations either Y_{Ai} or Y_{Bi} is observed for the ith subject. We assume that the conditional distribution of Y_{ki} given $\mathbf{Z}_i = \mathbf{z}$ is normal with mean $\mu_k(\mathbf{z}) = \alpha_k + \boldsymbol{\beta}_k^T\mathbf{z}$ and variance σ_k^2. We further assume that the variables $\mathbf{Z}_i$'s are independently and identically distributed (iid) according to a p-variate distribution which is completely/partially known. However, in real clinical trials, the covariates are the measurements on some clinical characteristics of human beings (e.g. age, blood pressure, glucose level), and hence they vary within certain limits. Therefore, it is clinically meaningful to assume that the covariates $\{\mathbf{Z}_j, j \geq 1\}$ are uniformly bounded.

As indicated earlier, the first step in deriving an optimal allocation is to define a measure of ethical sensitivity. Treatment failures, allocations to the inferior treatment and the expected treatment responses are often used to measure the ethical sensitivity. But, before quantifying such measures, we need to define a treatment success and an effective treatment. Unlike a binary response trial, a treatment success, in the context of a continuous response trial, is not uniquely defined. However, if we look at real clinical trials to reduce hypertension, blood pressure or blood sugar, we find that the response is continuous and a lower response, irrespective of the covariate value, is favourable. Therefore, if a lower response indicates a favourable situation, a treatment producing a response lower than a clinically relevant threshold can be regarded as successful. Thus treatment k is successful/beneficial for the ith subject if $Y_{ki} \leq c$ for some clinically relevant threshold c. Moreover, if a lower response is favourable, treatment k is more effective for the ith patient than treatment k' if $Y_{ki} < Y_{k'i},\ k \neq k'$. Therefore, we can set the following choices of the ethical metric $\gamma_k(\mathbf{Z})$ for treatment k: $\gamma_{k1}(\mathbf{Z}) = P(Y_{k1} > Y_{k'1}|\mathbf{Z})$, $\gamma_{k2}(\mathbf{Z}) = P(Y_{k1} \geq c|\mathbf{Z})$ and $\gamma_{k3}(\mathbf{Z}) = P(Y_{k1} \geq c, Y_{k1} > Y_{k'1}|\mathbf{Z})$, where lower values are desired for any treatment k. Suppose n_k subjects are assigned to treatment k $(= A, B)$ in a non-randomized manner and it remains to define a combined metric. However, different covariate profiles are of different importance and hence deriving a combined metric requires further consideration. For a better apprehension, assume that the covariate $\mathbf{Z}$ is categorical taking only the values z_1, z_2 and z_3 with the consideration that z_1 (z_3) is the most (least) favorable covariate profile. That is, $z_1 < z_2 < z_3$ and hence, the response of a subject with covariate profile z_1 to a particular treatment is expected to be the smallest. Then an ethical metric corresponding to treatment k can be expressed as $\sum_{j=1}^{3} w_j \gamma_k(z_j)$, where w_j is the non-negative weight assigned to covariate profile $z_j,\ j = 1, 2, 3$. Since, a lower value of $\gamma_k(z)$ indicates higher treatment effectiveness for covariate profile z, $\gamma_k(z)$ must be increasing in z. Now, under the presence of treatment covariate interaction, treatment effectiveness depends on the covariate profile and in the current context we assume that treatment A is most effective for $Z = z_1$ and least effective for $Z = z_3$ but for $Z = z_2$ both of the treatments are equally effective. Thus the relations $\gamma_A(z_1) \leq \gamma_B(z_1)$, $\gamma_A(z_2) = \gamma_B(z_2)$ and $\gamma_A(z_3) \geq \gamma_B(z_3)$, are natural. Although different choices of weights provide different metrics but we use $w_j = P(Z = z_j),\ j = 1, 2, 3$ in the current work. Such a choice reduces the treatment effectiveness measure for treatment k to $E_{\mathbf{Z}}\left\{\gamma_k(\mathbf{Z})\right\}$ and in the current work, we continue our development with this. Now, if we consider the first criterion, it can be observed that $E_{\mathbf{Z}}\left\{\gamma_{A1}(\mathbf{Z})\right\} = P(Y_{A1} \geq Y_{B1})$ represents the probability that treatment A is inferior than treatment B. Naturally, $n_A P(Y_{A1} \geq Y_{B1}) + n_B P(Y_{B1} \geq Y_{A1})$ can be looked upon as the expected number of assignments to the inferior treatment. Again, $E_{\mathbf{Z}}\left\{\gamma_{A2}(\mathbf{Z})\right\} = P(Y_{A1} \geq c)$ gives the probability that treatment A is not beneficial and hence the expected number of subjects not benefited under the trial can be expressed as $n_A P(Y_{A1} \geq c) + n_B P(Y_{B1} \geq c)$. In a similar fashion, $n_A P(Y_{A1} \geq Y_{B1}, Y_{A1} \geq c) + n_B P(Y_{B1} \geq Y_{A1}, Y_{B1} \geq c)$ represents the expected number of subjects assigned to the inferior and non-beneficial treatment. Therefore, a combined measure of the ethical sensitivity corresponding to any ethical metric $\gamma_k(\mathbf{Z})$ can be expressed as

$$G(n_A, n_B) = n_A E_{\mathbf{Z}}\left\{\gamma_A(\mathbf{Z})\right\} + n_B E_{\mathbf{Z}}\left\{\gamma_B(\mathbf{Z})\right\}.$$

Clearly, with these choices of $\gamma_k(\mathbf{Z})$, minimization of $G(n_A, n_B)$ is required to maintain the ethical requirements. Under normality of responses, the ethical metrics can be expressed as $\gamma_{k1}(\mathbf{Z}) = \Phi\left(\frac{\mu_k(\mathbf{Z}) - \mu_{k'}(\mathbf{Z})}{\sqrt{\sigma_A^2 + \sigma_B^2}}\right)$ and $\gamma_{k2}(\mathbf{Z}) = \Phi\left(\frac{\mu_k(\mathbf{Z}) - c}{\sigma_k}\right)$. For the derivation of the third criterion, we note that the conditional (i.e. for given $\mathbf{Z}$) joint distribution of Y_{k1} and $Y_{k1} - Y_{k'1}$ is bivariate normal with means $\mu_k(\mathbf{Z})$ and $\mu_k(\mathbf{Z}) - \mu_{k'}(\mathbf{Z})$, variances σ_k^2 and $\sigma_A^2 + \sigma_B^2$ and correlation coefficient $\zeta_k = \frac{\sigma_k}{\sqrt{\sigma_A^2 + \sigma_B^2}}$. Hence $\gamma_{k3}(\mathbf{Z})$ can be expressed as $\Phi_2\left(\frac{\mu_k(\mathbf{Z}) - c}{\sigma_k}, \frac{\mu_k(\mathbf{Z}) - \mu_{k'}(\mathbf{Z})}{\sqrt{\sigma_A^2 + \sigma_B^2}}, \zeta_k = \frac{\sigma_k}{\sqrt{\sigma_A^2 + \sigma_B^2}}\right)$, where $\Phi_2(.,.,\zeta)$ is the distribution function of a standard bivariate normal distribution with correlation coefficient ζ. Now it is observed that $\gamma_{k1}(\mathbf{Z})$ is sensitive to only the difference in the treatment effectiveness and $\gamma_{k2}(\mathbf{Z})$ is sensitive only to the departure from the threshold whereas $\gamma_{k3}(\mathbf{Z})$ is sensitive to both difference in the treatment effectiveness and the departure from the threshold.

2.2. Statistical optimality

In a statistical optimal procedure one can search for an allocation rule by controlling the variability of a distance measure between the two parameters $\boldsymbol{\eta}_A$ and $\boldsymbol{\eta}_B$, where $\boldsymbol{\eta}_k = (\alpha_k, \boldsymbol{\beta}_k^T)^T,\ k = A, B$. However, the true parameters are never known in advance and hence, as an alternative, we suggest to control the variability of a relevant sample distance measure. In the current work we suggest to control the variability of the Wald type distance measure, used in testing the equality of $\boldsymbol{\eta}_A$ and $\boldsymbol{\eta}_B$. If $\hat{\boldsymbol{\eta}}_{kn}$ denotes the maximum likelihood (ML) estimate of $\boldsymbol{\eta}_k$ based on $n = n_A + n_B$ observations, then such distance is defined by the statistic

$$W_n = (\hat{\boldsymbol{\eta}}_{An} - \hat{\boldsymbol{\eta}}_{Bn})^T \hat{V}_n^{-1} (\hat{\boldsymbol{\eta}}_{An} - \hat{\boldsymbol{\eta}}_{Bn}),$$

where $\widehat{V}_n$ is the ML estimate of V_n, the conditional (given all the covariate values) dispersion matrix of $\hat{\boldsymbol{\eta}}_{An} - \hat{\boldsymbol{\eta}}_{Bn}$. Under the normal response model, the conditional distribution (given all the covariate values) of $\hat{\boldsymbol{\eta}}_{An} - \hat{\boldsymbol{\eta}}_{Bn}$ is $(p+1)$ variate normal with mean vector $\boldsymbol{\eta}_A - \boldsymbol{\eta}_B$ and dispersion matrix

$$V_n = \frac{\sigma_A^2}{n_A} S_A^{-1} + \frac{\sigma_B^2}{n_B} S_B^{-1},$$

where $S_k = \frac{1}{n_k} \sum_{j=1}^{n_k} \boldsymbol{\xi}_j \boldsymbol{\xi}_j^T$ and $\boldsymbol{\xi}_j = (1, \mathbf{Z}_j^T)^T$ and is independent of $\hat{V}_n$. Then, since W_n is a quadratic form involving $\hat{\boldsymbol{\eta}}_{An} - \hat{\boldsymbol{\eta}}_{Bn}$ and $\hat{V}_n$, the conditional variance of W_n can be expressed as (see, for example, Seber and Lee (2003))

$$4(\boldsymbol{\eta}_A - \boldsymbol{\eta}_B)^T \hat{V}_n^{-1} V_n \hat{V}_n^{-1} (\boldsymbol{\eta}_A - \boldsymbol{\eta}_B) + 2Trace(\hat{V}_n^{-1} V_n \hat{V}_n^{-1} V_n).$$

We now assume that $n \to \infty$ and $\frac{n_k}{n}$ converges to a limit belonging to $(0,1)$. Then, if $E_{\mathbf{Z}_1}(\boldsymbol{\xi}_1 \boldsymbol{\xi}_1^T)$ is positive definite, $\widehat{V}_n$ is asymptotically equivalent in probability to V_n and hence the asymptotic variance of W_n comes out to be a decreasing function of $\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}$. Therefore, fixing the asymptotic variance of W_n or equivalently $\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}$, a statistically optimal procedure can be obtained.

2.3. The optimal target

Thus we have in our hand the measures of both statistical and clinical optimality, the prerequisites to derive an optimal allocation ratio. But minimisation of $G(n_A, n_B)$ alone provides degenerate rules and this necessitates the consideration of statistical optimality. Naturally, to find the optimal ratio $\frac{n_A}{n_A + n_B}$, we need to minimise $G(n_A, n_B)$ fixing $\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}$ below a suitable quantity. We, therefore, formulate the following optimization problem:

$$\min_{n_A, n_B} \quad [n_A E_{\mathbf{Z}} \{\gamma_A(\mathbf{Z})\} + n_B E_{\mathbf{Z}} \{\gamma_B(\mathbf{Z})\}]$$

subject to

$$\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B} = \kappa,$$

where κ is a prefixed positive quantity. With $\Psi_k = E_{\mathbf{Z}} \{\gamma_k(\mathbf{Z})\} (> 0)$, the optimal solution to the optimisation problem above is

$$\pi_A = \left(\frac{n_A}{n_A + n_B} \right)_{opt} = \frac{\sigma_A \sqrt{\Psi_B}}{\sigma_A \sqrt{\Psi_B} + \sigma_B \sqrt{\Psi_A}}.$$

2.4. CARA allocation in practice

It is worth mentioning at this point that the presence of treatment covariate interaction implies that a treatment is not uniformly effective for every covariate profile and hence the allocation probability to a treatment must depend on the patient's covariate profile. For example, suppose treatment A is more (less) effective than treatment B for a given covariate profile. Naturally, the allocation probability of a subject with that covariate profile must be higher (lower) for treatment A than treatment B. However, the optimal ratio π_k does not depend on the covariate profile $\mathbf{Z}$ and hence to assign patients according to their covariate profile, we need to use the current patient's covariate for his/her assignment. In most of the trials the covariate information is available at the time of randomisation, and hence it is not difficult to incorporate such information in the randomisation procedure. In addition, the optimal ratio π_k depends on the unknown parameter $\boldsymbol{\theta} = (\boldsymbol{\theta}_A^T, \boldsymbol{\theta}_B^T)$ with $\boldsymbol{\theta}_k = (\boldsymbol{\eta}_k^T, \sigma_k)^T$ and hence we need to replace them by suitable estimates. Therefore, in order to develop a reasonable randomization procedure based on the optimal ratio, we need to use the current patient's covariate and the available parameter estimates in a reasonable way. In the present work, for the assignment of the $(i+1)$th subject, we suggest to replace the unknown quantity $E_{\mathbf{Z}}\left\{\gamma_k(\mathbf{Z})\right\}$ by its projection $\hat{\gamma}_{ki}(\mathbf{Z}_{i+1})$, where the unknown parameters are replaced by the estimates, calculated on the basis of the response, allocation and covariate data prior to the randomization of the $(i+1)$th subject and $\mathbf{Z}$ by the current covariate value $\mathbf{Z}_{i+1}$. If δ_{kj} is the treatment indicator of the jth subject, that is, $\delta_{kj} = 1$ if treatment k is applied and 0 otherwise and $\mathcal{F}_i,\ i \geq 1$ denotes the information on allocations, responses and covariates up to and including the ith subject then the suggested CARA randomization corresponding to any metric $\gamma_k(\mathbf{Z})$ can be described by the allocation probability

$$P(\delta_{k,i+1} = 1|\mathcal{F}_i, \mathbf{Z}_{i+1}) = \pi_{ki}(\widehat{\boldsymbol{\theta}}_i, \mathbf{Z}_{i+1})$$

where $\hat{\boldsymbol{\theta}}_i$ is the ML estimate of $\boldsymbol{\theta}$ based on history of the data up to the ith entry of the incoming subject and

$$\pi_{Ai}(\widehat{\boldsymbol{\theta}}_i, \mathbf{Z}_{i+1}) = 1 - \pi_{Bi}(\widehat{\boldsymbol{\theta}}_i, \mathbf{Z}_{i+1}) = \frac{\hat{\sigma}_{Ai}\sqrt{\hat{\gamma}_{Bi}(\mathbf{Z}_{i+1})}}{\hat{\sigma}_{Bi}\sqrt{\hat{\gamma}_{Ai}(\mathbf{Z}_{i+1})} + \hat{\sigma}_{Ai}\sqrt{\hat{\gamma}_{Bi}(\mathbf{Z}_{i+1})}}.$$

Clearly $\pi_{ki}(\widehat{\boldsymbol{\theta}}_i, \mathbf{Z}_{i+1})$ is an estimate based on the available data where $E_{\mathbf{Z}}\left\{\gamma_k(\mathbf{Z})\right\}$ is replaced by its projection $\hat{\gamma}_{ki}(\mathbf{Z}_{i+1})$ and σ_k is replaced by its available estimate $\hat{\sigma}_{ki}, k = A, B$. However, for implementation in practice, we need to assign an initial number n_0 of subjects to each treatment arm to get the starting estimates of the parameters. For the estimation of parameters, we suggest to use sequentially updated maximum likelihood estimates. Specifically, with the response model as $Y_{kj} \sim N(\boldsymbol{\eta}_k^T\boldsymbol{\xi}_j, \sigma_k^2),\ j \geq 1$ and observations up to stage i, the likelihood function of $\boldsymbol{\theta}$, ignoring the constant of proportionality, can be expressed as

$$\mathcal{L}_i(\boldsymbol{\theta}) = \mathcal{L}_i(\boldsymbol{\theta}_A)\mathcal{L}_i(\boldsymbol{\theta}_B),$$

where

$$log\mathcal{L}_i(\boldsymbol{\theta}_k) = -\frac{\sum_{j=1}^{i}\delta_{kj}}{2}log\sigma_k^2 - \frac{1}{2}\sum_{j=1}^{i}\delta_{kj}\left(\frac{Y_{kj} - \boldsymbol{\eta}_k^T\boldsymbol{\xi}_j}{\sigma_k}\right)^2.$$

Now, writing

$$N_{ki} = \sum_{j=1}^{i}\delta_{kj}, \bar{y}_{ki} = \frac{\sum_{j=1}^{i}\delta_{kj}Y_{kj}}{N_{ki}}, \bar{\mathbf{Z}}_{ki} = \frac{\sum_{j=1}^{i}\delta_{kj}\mathbf{Z}_j}{N_{ki}}$$

$$S_{zz,i}^{(k)} = \sum_{j=1}^{i}\delta_{kj}(\mathbf{Z}_j - \bar{\mathbf{Z}}_{ki})(\mathbf{Z}_j - \bar{\mathbf{Z}}_{ki})^T$$

$$\mathbf{s}_{zY,i}^{(k)} = \sum_{j=1}^{i}\delta_{kj}(\mathbf{Z}_j - \bar{\mathbf{Z}}_{ki})Y_{kj},$$

we can express the ML estimates based on the data available up to ith stage as

$$\widehat{\boldsymbol{\beta}}_{ki} = S_{zz,i}^{(k)^{-1}}\mathbf{s}_{zY,i}^{(k)}, \quad \widehat{\alpha}_{ki} = \bar{Y}_{ki} - \widehat{\boldsymbol{\beta}}_{ki}^T\bar{\mathbf{Z}}_{ki}, \quad \widehat{\sigma}_{ki}^2 = \frac{1}{N_{ki}}\sum_{j=1}^{i}\delta_{kj}(Y_{kj} - \widehat{\alpha}_{ki} - \widehat{\boldsymbol{\beta}}_{ki}^T\mathbf{Z}_j)^2.$$

Let $\widehat{\boldsymbol{\theta}}_{kn}$ denote the ML estimate of $\boldsymbol{\theta}_k$ based on the data available up to the nth stage. Then irrespective of any metric $\gamma_k(\mathbf{Z})$ we have the following results, proved in Appendix A.

Result 2.1: *As $n \to \infty$,*

$$\widehat{\boldsymbol{\theta}}_{kn} \to \boldsymbol{\theta}_k$$

almost surely, $k = A, B$.

Result 2.2: *As $n \to \infty$,*

$$\frac{N_{kn}}{n} \to E_{\mathbf{Z}_1} \pi_k(\boldsymbol{\theta}, \mathbf{Z}_1)$$

almost surely, $k = A, B$.
Now to examine the extent of benefited proportion of subjects under the proposed allocation, we define the indicator variable $F_i = \delta_{Ai} I_{[Y_{Ai} \leq c]} + \delta_{Bi} I_{[Y_{Bi} \leq c]}$, $i \geq 1$ for the ith patient. We use $\bar{F}_n = n^{-1} \sum_{i=1}^{n} F_i$ as an indicative measure of the proportion benefited under the trial, which takes into account the covariate information. Then we have the following result, which is proved in Appendix B.

Result 2.3: *As $n \to \infty$,*

$$\bar{F}_n \to \psi = E_{\mathbf{Z}_1} \left\{ \sum_{k=A,B} \pi_k(\boldsymbol{\theta}, \mathbf{Z}_1) P(Y_{k1} \leq c \mid \mathbf{Z}_1) \right\},$$

almost surely.

However, the role of covariate information can be best explained if we can separate the covariate levels. For example, considering $\mathbf{Z}$ to be categorical, we denote the number of subjects having covariate level $\mathbf{z}$ assigned to treatment k by $N_{kn}(\mathbf{z})$ and the corresponding total number of subjects for a given covariate level $\mathbf{z}$ by $N_n(\mathbf{z})$, after the completion of n assignments. Then we get $\frac{N_{kn}(\mathbf{Z})}{N_n(\mathbf{Z})}$ as the the conditional proportion of subjects assigned to treatment k at $\mathbf{Z} = \mathbf{z}$. Now we have the following result, the proof of which follows from Zhang et al. (2007).

Result 2.4: *As $n \to \infty$,*

$$\frac{N_{kn}(\boldsymbol{z})}{N_n(\boldsymbol{z})} \to \pi_k(\boldsymbol{\theta}, \mathbf{z})$$

almost surely, for a given covariate, provided $P(\boldsymbol{Z}_1 = \boldsymbol{z}) > 0$.

In a similar fashion, the conditional proportion of subjects benefited in the trial can be expressed as $\frac{F_n(\mathbf{Z})}{N_n(\mathbf{Z})}$, where $F_n(\mathbf{z})$ denote the number of subjects benefited under the trial for a given covariate category $\mathbf{z}$. Proceeding as in Zhang et al. (2007), for a given covariate $\mathbf{z}$, we find the almost sure limit of the conditional proportion of benefited subjects as $\sum_{k=A,B} \pi_k(\boldsymbol{\theta}, \mathbf{z}) P(Y_{k1} \leq c \mid \mathbf{z})$, provided $P(\mathbf{Z}_1 = \boldsymbol{z}) > 0$.
However, assessing the optimality in the presence of covariate is not only critical but also a debatable issue. In a covariate ignored procedure, optimality means achieving the optimal target proportions in the limit. But there is no real guideline on deriving and defining an optimal allocation in the presence of covariate information. The current work is a reasonable extension of the framework developed in Jennison and Turnbull (2000) to derive the optimal target proportions. Although the derivation of the optimal proportion is based on average patients but in actual implementation, the current patient's covariate is used to calculate his/her assignment probability. In particular, we suggested to replace $E_{\mathbf{Z}} \{\gamma_k(\mathbf{Z})\}$ by its projection $\hat{\gamma}_{ki}(\mathbf{Z}_{i+1})$ to allow the assignment probability of the $(i+1)$th patient to depend on his/her covariate. This makes the allocation *covariate adjusted* but leads to a limiting allocation proportion which is different from the target optimal proportion. On the other hand, if one

replaces $E_{\mathbf{Z}}\{\gamma_k(\mathbf{Z})\}$ by the conventional estimate $i^{-1}\sum_{j=1}^{i}\hat{\gamma}_{kj}(\mathbf{Z}_j)$ for the assignment of the $(i+1)$th subject, the optimal target proportion π_k is realized in the limit and makes the allocation *conventionally optimal*. But in any clinical trial the ethical imperative lies in ensuring the best possible care for the subjects using all available information, and a *conventionally optimal* procedure does not incorporate the current patient's covariate for his/her allocation even if it is available prior to randomization and hence lacks sensibility. Naturally, the proposed allocation seemed to be more desirable than the conventionally optimal allocation in the context of a real clinical trial. In fact, we have modified the un- conditional optimal proportion π_k in accordance with the clinician's requirement to get the conditional (conditional on the covariate) treatment allocation probabilities and hence, refer the allocation as *near* or *pseudo* optimal.

3. Performance evaluation

3.1. A hypothetical clinical trial

To assess the performance of the proposed allocation, we consider a hypothetical clinical trial on reducing systolic blood pressure of the patients. Since, in general, the response to a treatment of blood pressure reduction varies with age, we consider age as the only covariate and assume a realistic situation, where the experimental treatment (i.e. treatment A) is more efficient than the Control (i.e. treatment B) for all age groups but the effectiveness for both the treatments is the highest for the young patients, moderate for the middle aged and the lowest for the old aged. This naturally convinces the presence of a quantitative treatment-covariate interaction. The patients enter the study sequentially , their ages are noted and depending on the age they are classified either Young (age $<$ 30 years) or Middle-aged (age falls in the interval 30-60 years) or Old-aged (age $>$ 60 years). Naturally, the effect of the treatment is expected to be the most for the young patients and the least for the old-aged patients and hence the young patient category is the most favourable (MF), the middle-aged category is only favourable (F) whereas the old-aged category is least favourable (LF). Since we are categorizing a continuous variable (i.e. age), it becomes necessary to assign appropriate values to different categories. If the relative weights of different categories are known, then one can use the well known Likert's category scaling (Garret (1966)) to determine the representative numerical values. Specifically, depending on the proportion of patients in the *MF*, *F* and *LF* categories, the corresponding values of the categorizing variable Z can be determined.

3.2. Performance measures

For the present hypothetical trial a treatment is naturally beneficial if it is capable of reducing the systolic blood pressure of the subject below the *Normal* limit. Recent reports (e.g. "Understanding blood pressure readings", American Heart Association, 2011; http://www.heart.org/HEARTORG/Conditions/HighBloodPressure) reveal that a systolic blood pressure of less than 120 mmHg is defined as *Normal* and this suggests to take c as 120 (in mmHg), irrespective of the other prognostic factors. Being consistent with the development so far, we assume that the response of a subject assigned to treatment k has a normal distribution with mean $\alpha_k + \beta_k z$ and constant variance σ^2, for given covariate value z of Z following a discrete probability distribution. Suppose the proportion of patients in the *MF*, F and *LF* categories are respectively $\frac{1}{6}$, $\frac{1}{3}$ and $\frac{1}{2}$. The categories are ordered and hence to run the allocation, we need to assign appropriate numerical weights to each category. Then Likert's method under the assumption of normality of age gives, respectively, the weights -1.5, -.45 and .80 for the most favourable, favourable and the least favourable categories. Thus we can take

$$Z = \begin{cases} -1.5 & \text{for a young patient} \\ -.45 & \text{for a middle-aged patient} \\ +.80 & \text{for an old-aged patient} \end{cases} .$$

with $p_{MF} = P(Z = -1.5) = \frac{1}{6}$, $p_F = P(Z = -.45) = \frac{1}{3}$, and $p_{LF} = P(Z = .80) = \frac{1}{2}$. For the evaluation of performance, we consider

- The overall and conditional allocation proportions to treatment A together with the standard errors.
- The overall and conditional proportions of benefited subjects together with the relevant standard errors.
- The power of a test of equality of treatment effects $H_0 : \alpha_A = \alpha_B, \beta_A = \beta_B$.

Now to assess the performance of the class of allocation procedures in small samples, we consider the allocation designs corresponding to each of the metrics γ_{k1}, γ_{k2} and γ_{k3} where the allocation design corresponding to the metric γ_{kj} is indicated by Dj. We, in addition, consider equiprobability randomization (denoted by $D4$), where either of the treatments is assigned with probability $\frac{1}{2}$ and conduct a simulation study and calculate the sample size necessary to attain 80% power using $D4$ for a test of equality of treatment effects under specific alternatives and evaluate the performance at this sample size. Specifically, for each fixed $\mathbf{z}$, we examine $\frac{N_{kn}(\mathbf{Z})}{N_n(\mathbf{Z})}$, the conditional proportion of allocations (CAP) to treatment k and $\frac{F_n(\mathbf{Z})}{N_n(\mathbf{Z})}$, the conditional proportion of subjects benefited (CBP) in the trial when the trial is conducted with n subjects. The conditional proportion of allocations to treatment A (denoted by CAP_A), expected allocation proportions to treatment A (denoted by EAP_A), the conditional and overall benefited proportions (indicated, respectively, by CBP and EBP) are also computed together with their standard errors. The relevant test statistic is W_n^*,

which is the same as W_n except that the estimates used are obtained following the adaptive procedure with n_k replaced by the corresponding observed number N_{kn}. A larger value of W_n^* indicates possible rejection of the null hypothesis. We always take $\alpha_B = 121, \alpha_A = 119.5$ and vary σ, β_k and $\mathbf{p} = (p_{MF}, p_F, p_{LF})^T$. For the purpose of computation, we always choose β_k's to ensure highest (lowest) expected response for the patients with most (least) favourable covariate categories. Now the covariate categories may be of equal as well as unequal importance. But equal proportions in $\mathbf{p}$ is not convincing in practice and therefore, we always assume that p_{MF} is the highest. We provide the relevant performance measures assuming unequal weights for the covariate categories in Table 1.

Remarks: As expected, the presence of treatment covariate interaction led to reasonable outcomes. Except for design $D4$, we observe the experience of the lowest and the highest allocation proportions by the subjects with the most favourable and the least favourable covariate category, respectively. That is, the allocation provides the safeguard to the subjects who actually needs the better therapy (i.e. patients with $Z = 0.80$). The conditional proportions (both allocation and benefited) vary sensibly with both the covariate values and treatment effectiveness. However, the precision level (i.e. statistical power) decreases for all the CARA allocation designs (i.e. $D1$, $D2$ and $D3$) with an increase in β_B. In fact, the loss in power is minimum when $\beta_A = \beta_B$ and increases with a decrease in $\beta_A - \beta_B$. It is interesting to note that among the CARA allocation designs, the loss in precision (i.e. statistical power) compared to 80% mark is minimum for $D1$. This is not surprising if we look at the corresponding CAP (or EAP) values, where lower difference in $\beta_A - \beta_B$ indicates larger allocation towards the better treatment and with higher loss in power. However, such a loss can be compensated with the significant gain in the ethics measures. Thus we have derived a class of allocations which uses all available information, even the covariate information of the entering subject.

Table 1: Performance comparison at 80% power with $\alpha_A = 119.50, \alpha_B = 121, \mathbf{p} = (1/6, 1/3, 1/2)$.

		CAP_A **(SD) with covariate profile**			EAP_A**(SD)**	**CBP (SD) with covariate profile**			**EBP (SD)**	**Power**
$(n, \beta_A, \beta_B, \sigma)$	*Design*	MF	F	LF		MF	F	LF		
(81,-3,-3,1)	D1	0.599 (0.111)	0.615 (0.091)	0.622 (0.083)	0.616 (0.056)	0.012 (0.001)	0.116 (0.063)	0.968 (0.028)	0.499 (0.052)	0.787
	D2	0.598 (0.112)	0.617 (0.092)	0.693 (0.086)	0.651 (0.057)	0.010 (0.001)	0.118 (0.063)	0.974 (0.025)	0.502 (0.051)	0.748
	D3	0.501 (0.111)	0.528 (0.091)	0.682 (0.085)	0.595 (0.056)	0.014 (0.001)	0.102 (0.059)	0.973 (0.026)	0.497 (0.052)	0.773
	D4	0.500 (0.110)	0.500 (0.090)	0.501 (0.076)	0.501 (0.052)	0.002 (0.001)	0.096 (0.056)	0.959 (0.032)	0.489 (0.052)	0.801
(88,-3,-2.4,1)	D1	0.552 (0.112)	0.609 (0.087)	0.682 (0.079)	0.633 (0.054)	0.106 (0.001)	0.122 (0.062)	0.942 (0.035)	0.491 (0.05)	0.753
	D2	0.547 (0.114)	0.612 (0.089)	0.767 (0.076)	0.674 (0.054)	0.012 (0.001)	0.121 (0.060)	0.957 (0.031)	0.498 (0.052)	0.661
	D3	0.498 (0.107)	0.528 (0.086)	0.753 (0.077)	0.629 (0.053)	0.014 (0.011)	0.106 (0.056)	0.954 (0.032)	0.492 (0.051)	0.673
	D4	0.500 (0.106)	0.499 (0.086)	0.500 (0.074)	0.500 (0.049)	0.003 (0.001)	0.101 (0.055)	0.910 (0.044)	0.468 (0.051)	0.798
(112,-3,-1.8,1)	D1	0.482 (0.109)	0.604 (0.082)	0.753 (0.069)	0.653 (0.051)	0.014 (0.002)	0.128 (0.056)	0.918 (0.035)	0.485 (0.045)	0.623
	D2	0.48 (0.111)	0.606 (0.081)	0.841 (0.06)	0.696 (0.048)	0.016 (0.002)	0.127 (0.055)	0.945 (0.030)	0.498 (0.045)	0.307
	D3	0.500 (0.099)	0.523 (0.078)	0.819 (0.063)	0.662 (0.046)	0.015 (0.002)	0.114 (0.053)	0.939 (0.033)	0.492 (0.045)	0.364
	D4	0.499 (0.099)	0.501 (0.078)	0.499 (0.066)	0.499 (0.045)	0.001 (0.002)	0.105 (0.052)	0.834 (0.051)	0.439 (0.045)	0.806
(125,-3,-3,3/2)	D1	0.595 (0.102)	0.605 (0.077)	0.608 (0.071)	0.605 (0.049)	0.022 (0.01)	0.188 (0.061)	0.915 (0.035)	0.506 (0.043)	0.802
	D2	0.597 (0.102)	0.606 (0.078)	0.663 (0.079)	0.632 (0.052)	0.024 (0.010)	0.189 (0.062)	0.923 (0.034)	0.511 (0.043)	0.784
	D3	0.503 (0.096)	0.531 (0.074)	0.648 (0.075)	0.583 (0.048)	0.028 (0.010)	0.172 (0.059)	0.922 (0.035)	0.504 (0.043)	0.769
	D4	0.499 (0.095)	0.499 (0.074)	0.501 (0.063)	0.499 (0.043)	0.002 (0.009)	0.165 (0.058)	0.899 (0.039)	0.491 (0.043)	0.804
(134,-3,-2.4,3/2)	D1	0.542 (0.103)	0.593 (0.075)	0.652 (0.069)	0.612 (0.048)	0.033 (0.011)	0.197 (0.060)	0.888 (0.038)	0.497 (0.042)	0.769
	D2	0.541 (0.104)	0.594 (0.076)	0.715 (0.075)	0.643 (0.051)	0.035 (0.011)	0.197 (0.061)	0.904 (0.037)	0.505 (0.042)	0.723
	D3	0.501 (0.093)	0.529 (0.073)	0.693 (0.072)	0.604 (0.047)	0.033 (0.010)	0.185 (0.059)	0.898 (0.038)	0.497 (0.042)	0.736
	D4	0.501 (0.094)	0.501 (0.072)	0.500 (0.061)	0.501 (0.042)	0.002 (0.012)	0.177 (0.058)	0.851 (0.044)	0.472 (0.042)	0.802
(174,-3,-1.8,3/2)	D1	0.483 (0.096)	0.58 (0.068)	0.701 (0.061)	0.621 (0.044)	0.055 (0.014)	0.209 (0.054)	0.867 (0.036)	0.493 (0.037)	0.714
	D2	0.481 (0.095)	0.583 (0.069)	0.771 (0.062)	0.656 (0.045)	0.057 (0.013)	0.208 (0.054)	0.891 (0.035)	0.505 (0.037)	0.564
	D3	0.500 (0.085)	0.526 (0.065)	0.741 (0.061)	0.626 (0.041)	0.054 (0.013)	0.199 (0.053)	0.881 (0.037)	0.498 (0.038)	0.598
	D4	0.501 (0.084)	0.501 (0.064)	0.500 (0.053)	0.500 (0.037)	0.005 (0.013)	0.195 (0.052)	0.794 (0.044)	0.454 (0.037)	0.799

Figures within the parentheses indicate the corresponding standard error

Moreover it assigns subjects according to the importance of the covariate and therefore, has the ability to perform sensibly in real situations.

4. Concluding remarks

The development of the current work is made for two treatments assuming continuous treatment outcomes. The corresponding development with several treatments for any type of responses incorporating some additional clinically relevant issues is left for future work.

References

Atkinson, AC (1982). Optimum Biased Coin Designs for Sequential Clinical Trials with Prognostic Factors. *Biometrika*, **69**, 61–67.

Azriel D, Mandel M and Rinott Y (2012). Optimal Allocation to Maximize Power of Two-sample Tests for Binary Response. *Biometrika*, **99**, 101–113.

Bandyopadhyay, U and Biswas, A (2001). Adaptive Designs for Normal Responses with Prognostic Factors. *Biometrika*, **88**, 409–419.

Begg, CB and Iglewicz, B (1980). A Treatment Allocation Procedure for Sequential Clinical Trials. *Biometrics*, **36**, 81–90.

Biswas, A, Huang, HH and Huang, WT (2006). Covariate-adjusted Adaptive Designs for Continuous Responses in a Phase III Clinical Trial: Recommendation for Practice. *Journal of Biopharmaceutical Statistics* **16**, 227–239.

Biswas, A and Mandal, S (2004). Optimal Adaptive Designs in Phase III Clinical Trials for Continuous Responses with Covariates. In *mODa 7 - Advances in Model-Oriented Design and Analysis*, (Eds. A. Di. Bucchianico, H.LÂÍauter and H.P. Wynn), 51–58. Heidelberg: Physica-Verlag.

Friedman, LM, Furberg, CD and DeMets, DL (2010). *Fundamentals of Clinical Trials, Fourth Edition.* Springer-Verlag, New York.

Garret, HE (1966). *Statistics in Psychology and Education.* Longmans, Green.

Hu, F and Rosenberger, WF (2006). *The Theory of Response-adaptive Randomization in Clinical Trials.* Wiley Interscience, New York.

Jennison, C and Turnbull, BW (2000). *Group Sequential Methods with Applications to Clinical Trials.* Boca Raton, Florida: Chapman and Hall/CRC , Chapter 17.

Seber, GAF and Lee, AJ (2003) *Linear Regression Analysis, Second Edition.* John Wiley & Sons, Inc., Hoboken, New Jersey.

Shirayev, AN (1996). *Probability, Second Edition.* (Translated by R.P. Boas) Springer, New York.

Wong, WK and Zhu, W (2008). Optimum Treatment Allocation Rules under a Variance Heterogeneity Model. *Statistics in Medicine* **27** 4581–4595.

Zhang, LX, Hu, F, Cheung, SH and Chan, WS (2007). Asymptotic Properties of Covariate-adjusted Response-adaptive Designs. *Annals of Statistics* **35** 1166–1182.

Zhang, LX and Hu, F (2009). A New Family of Covariate-adjusted Response-adaptive Designs and Their Properties. *Applied Mathematical Journal of Chinese Universities* **24** 1–13.

Appendix A

Proof of Result 2.1: We note that the ML estimator $\widehat{\boldsymbol{\theta}}_{kn}$ is a solution to the equation

$$\frac{\partial log\mathcal{L}_n(\boldsymbol{\theta}_k)}{\partial\boldsymbol{\theta}_k} = \mathbf{0}.$$

Since $\{\mathbf{Z}_j, j \geq 1\}$ are assumed uniformly bounded, under the normal response model it follows from the Corollary 3.1 of Zhang and Hu(2009) that for $\epsilon > 0$ small enough

$$log\mathcal{L}_n(\boldsymbol{\theta}_k^*) < log\mathcal{L}_n(\boldsymbol{\theta}_k)$$

with probability 1 whenever $\parallel \boldsymbol{\theta}_k^* - \boldsymbol{\theta}_k \parallel = \epsilon$ and n is large enough. This in turn implies

$$\widehat{\boldsymbol{\theta}}_{kn} \to \boldsymbol{\theta}_k$$

almost surely as $n \to \infty$. Hence the strong consistency of the parameters under the response adaptive set up is established. □

Proof of Result 2.2: It is easy to observe that $\widehat{\boldsymbol{\eta}}_{kn}$, the ML estimate of $\boldsymbol{\eta}_k$, satisfies the equation

$$\sum_{j=1}^{n} \delta_{kj}(Y_{kj} - \widehat{\boldsymbol{\eta}}_{kn}^T\boldsymbol{\xi}_j)\boldsymbol{\xi}_j = \mathbf{0},$$

which can be rewritten as

$$\sum_{j=1}^{n} \delta_{kj}(Y_{kj} - \boldsymbol{\eta}_k^T\boldsymbol{\xi}_j)\boldsymbol{\xi}_j + \sum_{j=1}^{n} \delta_{kj}(\boldsymbol{\xi}_j^T\boldsymbol{\eta}_k - \boldsymbol{\xi}_j^T\widehat{\boldsymbol{\eta}}_{kn})\boldsymbol{\xi}_j = \mathbf{0}.$$

Therefore, $\widehat{\boldsymbol{\eta}}_{kn}$ satisfies the following equation

$$\frac{1}{n}\sum_{j=1}^{n} \delta_{kj}\boldsymbol{\xi}_j\boldsymbol{\xi}_j^T(\widehat{\boldsymbol{\eta}}_{kn} - \boldsymbol{\eta}_k) = \frac{1}{n}\sum_{j=1}^{n} \delta_{kj}(Y_{kj} - \boldsymbol{\eta}_k^T\boldsymbol{\xi}_j)\boldsymbol{\xi}_j. \qquad (A.1)$$

It can be observed that the matrices

$$A_j = \sum_{i=1}^{j} \frac{1}{i}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T \left\{\delta_{ki} - E(\delta_{ki}|\mathcal{F}_{i-1})\right\}, j \geq 1$$

and the vectors

$$\mathbf{b}_j = \sum_{i=1}^{j} \frac{1}{i}[\delta_{ki}(Y_{ki} - \boldsymbol{\eta}_k^T\boldsymbol{\xi}_i)\boldsymbol{\xi}_i - E\left\{\delta_{ki}(Y_{ki} - \boldsymbol{\eta}_k^T\boldsymbol{\xi}_i)\boldsymbol{\xi}_i|\mathcal{F}_{i-1}\right\}], j \geq 1$$

are both martingales. Now noting the fact that $|\delta_{ki} - E(\delta_{ki}|\mathcal{F}_{i-1})| \leq 1$ and using the martingale property, we get

$$E[Trace(A_jA_j^T)] \leq E_{\mathbf{Z}_1}[Trace(\boldsymbol{\xi}_1\boldsymbol{\xi}_1^T)]^2 \sum_{i=1}^{\infty} \frac{1}{i^2}$$

and

$$E[Trace(\mathbf{b}_j\mathbf{b}_j^T)] \leq \sigma_k^2 E_{\mathbf{Z}_1}[Trace(\boldsymbol{\xi}_1\boldsymbol{\xi}_1^T)] \sum_{i=1}^{\infty} \frac{1}{i^2}, k = A, B$$

for every $j \geq 1$. Since $\{\mathbf{Z}_j, j \geq 1\}$ are uniformly bounded and the response variances are finite, the right hand members of the above two inequalities are both finite. Thus using L_2-martingale convergence theorem together with Kronecker's lemma (see, for example, Theorem 4 of Shirayev(1996),pp.519), we have as $n \to \infty$

$$\frac{1}{n}\sum_{j=1}^{n}\left\{\delta_{kj}\boldsymbol{\xi}_j\boldsymbol{\xi}_j^T - E(\delta_{kj}\boldsymbol{\xi}_j\boldsymbol{\xi}_j^T|\mathcal{F}_{j-1})\right\} \to 0^{p+1\times p+1}$$

and

$$\frac{1}{n}\sum_{j=1}^{n}\left[\delta_{kj}(Y_{kj} - \boldsymbol{\eta}_k^T\boldsymbol{\xi}_j)\boldsymbol{\xi}_j - E\left\{\delta_{kj}(Y_{kj} - \boldsymbol{\eta}_k^T\boldsymbol{\xi}_j)\boldsymbol{\xi}_j|\mathcal{F}_{j-1}\right\}\right] \to \mathbf{0}$$

almost surely. Since

$$E(\delta_{kj}|\mathcal{F}_{j-1}, \mathbf{Z}_j) = \pi_{kj}(\widehat{\boldsymbol{\theta}}_{j-1}, \mathbf{Z}_j),$$

we have

$$E(\delta_{kj}\boldsymbol{\xi}_j\boldsymbol{\xi}_j^T|\mathcal{F}_{j-1}) - E_{\mathbf{Z}_1}\left\{\pi_{kj}(\widehat{\boldsymbol{\theta}}_{j-1}, \mathbf{Z}_1)\boldsymbol{\xi}_1\boldsymbol{\xi}_1^T\right\} = 0.$$

Further, as $\pi_k(\boldsymbol{\theta}, \mathbf{Z})$ is continuous in $\boldsymbol{\theta}$ for fixed $\mathbf{Z}$, it follows from the strong consistency of $\widehat{\boldsymbol{\theta}}_n$ that as $n \to \infty$

$$\frac{1}{n}\sum_{j=1}^{n}E_{\mathbf{Z}_1}\left\{\pi_{kj}(\widehat{\boldsymbol{\theta}}_{j-1}, \mathbf{Z}_1)\boldsymbol{\xi}_1\boldsymbol{\xi}_1^T\right\} \to I_{\pi k}(\boldsymbol{\theta})$$

almost surely, where

$$E_{\mathbf{Z}_1}\left\{\pi_k(\boldsymbol{\theta}, \mathbf{Z}_1)\boldsymbol{\xi}_1\boldsymbol{\xi}_1^T\right\} = I_{\pi k}(\boldsymbol{\theta}), k = A, B.$$

Hence, assuming $I_{\pi k}(\boldsymbol{\theta})$ to be non-singular, equation (A.1) implies

$$(\widehat{\boldsymbol{\eta}}_{kn} - \boldsymbol{\eta}_k) - I_{\pi k}^{-1}(\boldsymbol{\theta})\frac{1}{n}\sum_{j=1}^{n}\delta_{kj}(Y_{kj} - \boldsymbol{\eta}_k^T\boldsymbol{\xi}_j)\boldsymbol{\xi}_j \to \mathbf{0}$$

almost surely,k=A,B. Thus the required conditions of Theorem 2.1 of Zhang et al.(2007) are satisfied if we take

$$h_{kj}(Y_{kj}, \boldsymbol{\xi}_j) = I_{\pi k}^{-1}(\boldsymbol{\theta})(Y_{kj} - \boldsymbol{\eta}_k^T\boldsymbol{\xi}_j)\boldsymbol{\xi}_j,$$

where $E\left\{(Y_{kj} - \boldsymbol{\eta}_k^T\boldsymbol{\xi}_j)\boldsymbol{\xi}_j|\mathbf{Z}_j\right\} = \mathbf{0}$ for any j, k=A,B. In a similar fashion, it follows that as $n \to \infty$

$$(\widehat{\sigma}_{kn} - \sigma_k) - {I^*}_{\pi k}^{-1}(\boldsymbol{\theta})\frac{1}{n}\sum_{j=1}^{n}\delta_{kj}\left\{\frac{(Y_{kj} - \boldsymbol{\eta}_k^T\boldsymbol{\xi}_j)^2}{\sigma_k^3} - \frac{1}{\sigma_k}\right\} \to 0$$

almost surely, where

$${I^*}_{\pi k}(\boldsymbol{\theta}) = \frac{2}{\sigma_k^2}E_{\mathbf{Z}_1}\left\{\pi_k(\boldsymbol{\theta}, \mathbf{Z}_1)\right\}$$

is non-singular, k=A,B. Thus the required conditions of Theorem 2.1 of Zhang et al. (2007) are satisfied with

$$h_{kj}(Y_{kj}, \boldsymbol{\xi}_j) = {I^*}_{\pi k}^{-1}(\boldsymbol{\theta})\left\{\frac{(Y_{kj} - \boldsymbol{\eta}_k^T\boldsymbol{\xi}_j)^2}{\sigma_k^3} - \frac{1}{\sigma_k}\right\}, j = 1, 2, ...$$

Hence the result follows. □

Appendix B

Proof of Result 2.3: Writing

$$\frac{1}{n}\sum_{i=1}^{n} F_i = \frac{1}{n}\sum_{k=A,B}\sum_{i=1}^{n}\{\delta_{ki}F_{ki} - E(\delta_{ki}F_{ki} \mid \mathcal{F}_{i-1})\} + \frac{1}{n}\sum_{i=1}^{n} E(F_i \mid \mathcal{F}_{i-1}), \qquad (B.1)$$

where $F_{ki} = I_{[Y_{ki} \le c]}$ represents the failure indicator for the i-th patient with treatment k,k=A,B. Since $|\delta_{ki}F_{ki} - E(\delta_{ki}F_{ki} \mid \mathcal{F}_{i-1})| \le 1$, we as in the proof of Result 2.2, apply L_2-martingale convergence theorem on the martingale

$$M_j = \sum_{i=1}^{j} \frac{1}{i}\left\{\delta_{ki}F_{ki} - E(\delta_{ki}F_{ki} \mid \mathcal{F}_{i-1})\right\}, \; j \ge 1$$

and get that

$$\frac{1}{n}\sum_{i=1}^{n}\{\delta_{ki}F_{ki} - E(\delta_{ki}F_{ki} \mid \mathcal{F}_{i-1})\} \to 0 \qquad (B.2)$$

almost surely, k=A,B. Thus the convergence will be established if we can show that, as $n \to \infty$,

$$P(F_n = 1 \mid \mathcal{F}_{n-1}) \to \psi$$

almost surely. Moreover, since the covariates are iid, we have

$$E(F_n \mid \mathcal{F}_{n-1}) = E_{\mathbf{Z}_1}\left\{\sum_{k=A,B} P(F_{kn} = 1 \mid \delta_{kn} = 1, \mathbf{Z}_1, \mathcal{F}_{n-1})P(\delta_{kn} = 1 \mid \mathbf{Z}_1, \mathcal{F}_{n-1})\right\}. \qquad (B.3)$$

It can be easily seen that that

$$P(F_{kn} = 1 \mid \delta_{kn} = 1, \mathbf{Z}_1, \mathcal{F}_{n-1}) \quad = \quad P(Y_{k1} \le c \mid \mathbf{Z}_1), \qquad (B.4)$$

and, as in Zhang et al.(2007),

$$P(\delta_{kn} = 1 \mid \mathbf{Z}_1, \mathcal{F}_{n-1}) \to \pi_k(\boldsymbol{\theta}, \mathbf{Z}_1) \qquad (B.5)$$

almost surely. Hence, by the dominated convergence theorem together with Kronecker's lemma, the required result follows from (B.1) through (B.2)-(B.5). □

Affiliation:

Rahul Bhattacharya and Uttam Bandyopadhyay
Department of Statistics
Calcutta University
35 Ballygunge Circular Road, Kolkata – 700 019, India
E-mail: rahul_bhattya@yahoo.com and ubandyopadhyay08@gmail.com

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 44 *Submitted:* 2014-09-11
December 2015 *Accepted:* 2015-03-05

Austrian Journal of Statistics
December 2015, Volume 44, 67–68.
http://www.ajs.or.at/
doi:10.17713/ajs.v44i4.148

Laudatio verfasst von Peter Filzmoser
TU Wien

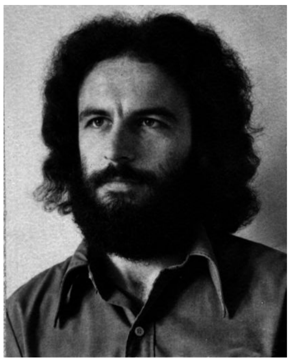
Zur Emeritierung von O.Univ.-Prof. Dipl.-Ing. Dr.techn Rudolf Dutter

Rudolf Dutter wurde in dem kleinen niederösterreichischen Dorf Grünsbach am 11.11.1946 vormittags geboren. Es wird behauptet, dass die genaue Geburtszeit 11:11 Uhr sei, was aufgrund der statistischen Schwankungsbreite dieser Information auch nicht widerlegbar ist. Sein Streben nach höherer Schulausbildung brachte ihn in die HTL für Elektrotechnik nach St. Pölten, und schließlich an die Technische Hochschule nach Wien, wo er von 1966-1970 das Studium „Technische Mathematik“ absolvierte. Seine Diplomarbeit zu einem Verkehrsflussmodell zeigte bereits das Interesse für fachübergreifende Kooperation. Diese Arbeit resultierte schließlich in seiner ersten Publikation in der damals noch jungen Zeitschrift „Computing“. Nicht zuletzt aufgrund dieser Publikation wurde er zum Doktoratsstudium an der Universität Montreal in Kanada aufgenommen.

Dieser große Schritt stieß von Seiten seines Elternhauses nicht nur auf Wohlwollen, denn er war aus ihrer Lebenssituation betrachtet sehr ungewöhnlich. In seiner Dissertation, betreut von den Professoren Anatole Joffe und Irwin Guttman, beschäftigte er sich unter anderem mit der Problematik von Ausreißern in statistischen Daten. Nach dem Abschluss seines PhD im Jahr 1973 erhielt Dr. Dutter eine Stelle als Forschungsassistent bei Prof. Peter J. Huber an der ETH Zürich. Huber war durch seine grundlegende Arbeit zu „M-Schätzern“ bereits damals sehr bekannt, und daher zog es viele heute wissenschaftlich sehr etablierte Personen zu dieser Geburtsstätte der „Robusten Statistik“. Von diesem Umfeld profitierte auch Rudolf Dutter stark, er entwickelte eine Faszination für „Computational Statistics“ und erlebte dort die vermutlich nachhaltigste Zeit für seine Forschungskarriere. Im Jahr 1976 erhielt er eine Stelle als Universitätsassistent am Institut für Statistik der TU Graz bei Prof. Ulrich Dieter. Dort konnte er sich vier Jahre später für das Fach „Statistik“ habilitieren. Als Dozent unterhielt er u.a. sehr intensiven Kontakt mit der Montanuniversität Leoben, lehrte dort das Fach „Geostatistik“, war in diverse Forschungsprojekte involviert, war recht aktiv mit statistischer Beratung und begann mit der Softwareentwicklung zur grafischen Analyse von räumlichen Daten.

1968: Diplomstudium

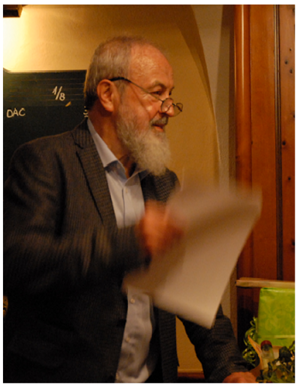
Dieses Tätigkeitsprofil entsprach dann offensichtlich den Anforderungen des ausgeschriebenen Lehrstuhls für „Technische Statistik“ an der TU Wien. Rudolf Dutter erhielt somit 1984 den Ruf und trat mit großer Begeisterung seinen Dienst am damaligen Institut für Statistik und Wahrscheinlichkeitstheorie an, also vor 31 Jahren.
Die Besetzung der „Technischen Statistik“ wurde nicht nur außerhalb, sondern auch innerhalb der TU Wien sichtbar: Statistik wurde für Studierende in Geoinformatik, Informatik, Maschinenbau und Betriebswirtschaft sowie Bauingenieurwesen in die



1973: Ph.D.-Studium

Studienpläne integriert und von Prof. Dutter abgehalten.

Ein „Statistisches Labor“ wurde ins Leben gerufen, um sowohl innerhalb als auch außerhalb der TU bei statistischen Fragestellungen zu beraten. Dies resultierte auch in zahlreichen Kooperationen mit anderen Fachgebieten, sowie in Firmenkooperationen.



Emeritierungsfeier (7.11.2015)

In seiner wissenschaftlichen Tätigkeit kann man Prof. Dutter unmittelbar mit den Themen „Computational Statistics“, „Robuste Statistik“ und „Geostatistik“ assoziieren. Mit mehreren Fachbüchern und zahlreichen Artikeln in Fachzeitschriften hat er seine Kompetenz zu diesen Themen unter Beweis gestellt. Seine Begeisterung für den computerorientierten Zugang zur Statistik und seine eigene Forschungsarbeit wirkten auch „ansteckend“, denn mehrere seiner früheren Mitarbeiterinnen und Mitarbeiter sind dem Ruf der Wissenschaft gefolgt und mittlerweile international anerkannte Forscherpersönlichkeiten geworden. Viele seiner Absolventinnen und Absolventen konnten die erworbene Statistik-Kompetenz in der Berufspraxis erfolgreich umsetzen.

Als ehemaliger Student und langjähriger Mitarbeiter von Rudi möchte ich kurz aus meiner Sicht darstellen, warum der Funken der Begeisterung für Statistik übergesprungen ist. Bereits bei der ersten Besprechung einer größeren Projektarbeit wurde klargestellt, dass man „Du“ zueinander sagt – eine ungewöhnliche Haltung eines Professors zu Studierenden. Dieses „Du“ war ein Zeichen dafür, dass Rudi die Zusammenarbeit auf gleicher Augenhöhe angestrebt hat. Automatisch wurden die Inhalte seiner Forschungsarbeit auch meine Aufgabenstellungen, an denen wir dann gemeinsam gearbeitet haben. Man wurde selbst mit noch unausgereiften Ideen ernst genommen und niemals bloß gestellt. Rudi war ein guter „Streitpartner“. Trotz unterschiedlicher Ansichten und teilweise heftiger Auseinandersetzungen darüber änderte das nichts an der persönlichen Beziehung zueinander, Ärger wurde nicht nachgetragen.

Bei Peter Filzmosers 30er

Besonders wertvoll waren die Kontakte zu anderen Forschern, die Rudi vermittelt hat. Nicht der eigene Glanz war ihm wichtig, sondern das Vorankommen der Mitarbeiterinnen und Mitarbeiter, an deren Erfolg er sich ehrlich gefreut hat. Ihm war die Sache wichtig, aber vor allem auch der Mensch, der dahintersteht, und daher hat er Gemeinschaft immer wieder gefördert durch Institutskaffee, Wanderungen, sportliche Aktivitäten, und Gespräche. Das dadurch entstandene gute Arbeitsklima brachte nicht nur wissenschaftliche Erfolge, sondern es führte auch zu tragfähigen Freundschaften. Rudis bescheidene und herzliche Art auch als Institutsvorstand hat gezeigt, wie wichtig nicht nur die wissenschaftliche Qualifikation ist, sondern wie stark die menschliche Komponente zu fruchtbringender Zusammenarbeit beiträgt.

Vorbildhaft für mich war auch sein Durchhaltevermögen, das er beim Marathonlaufen unter Beweis stellen konnte, sich aber auch in seiner ganzen Lebenshaltung zeigte. Diese Zähigkeit erwies sich als sehr vorteilhaft beim Lösen von wissenschaftlichen Problemen, aber auch beim Bewältigen von gesundheitlichen Problemen.

Natürlich war und ist ihm Familie stets ein großes Anliegen. Heute sind es wohl die Enkelkinder, die von ihm am meisten verwöhnt werden.

Nachdem die Emeritierung nur eine Entbindung von den Dienstpflichten ist, hoffen wir natürlich, dass Rudi diese Freiheit auch dazu nützt, um weiterhin die Kontakte zur Wissenschaft, zu Kolleginnen, Kollegen und Freunden zu pflegen. Die vorbildhafte Bescheidenheit, den Tatendrang, den kämpferischen Geist, die Ausdauer, und die Sanftmütigkeit möchte ich Dir auch in Zukunft wünschen, sowie bestmögliche Gesundheit im Kreise Deiner Lieben.

Address of author:

Peter Filzmoser
Institut für Stochastik und Wirtschaftsmathematik
Technische Universität Wien
E-mail: P.Filzmoser@tuwien.ac.at

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 44 *Accepted:* 2015-10-20
December 2015

ISSN 1026–597X

Contents