ISSN 1927-7032 (Print)
ISSN 1927-7040 (Online)

International Journal of Statistics and Probability

Vol. 5, No. 4 July 2016

CANADIAN CENTER OF SCIENCE AND EDUCATION

Editorial Board

Contents

International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

On the Estimation of Reliability of Weighted Weibull Distribution: A Comparative Study

Bander Al-Zahrani[1]

[1] Department of Statistics, King Abdulaziz University, Jeddah, Saudi Arabia

Correspondence: Bander Al-Zahrani, Department of Statistics, King Abdulaziz University, Jeddah, Saudi Arabia. Tel: 966-2-965-2295. E-mail: bmalzahrani@kau.edu.sa

Received: May 2, 2016 Accepted: May 25, 2016 Online Published: June 11, 2016

doi:10.5539/ijsp.v5n4p1 URL: http://dx.doi.org/10.5539/ijsp.v5n4p1

Abstract

The paper gives a description of estimation for the reliability function of weighted Weibull distribution. The maximum likelihood estimators for the unknown parameters are obtained. Nonparametric methods such as empirical method, kernel density estimator and a modified shrinkage estimator are provided. The Markov chain Monte Carlo method is used to compute the Bayes estimators assuming gamma and Jeffrey priors. The performance of the maximum likelihood, nonparametric methods and Bayesian estimators is assessed through a real data set.

Keywords: Weighted Weibull distribution, Empirical Estimator, Kernel Density Estimator, Modified Shrinkage Estimator, and Markov chain Monte Carlo.

1. Introduction

The Weibull distribution has been used very extensively as a model in reliability and survival analysis. The distribution provides much wider applications as compared with those provided by the exponential distribution. The Weibull distribution can also be used as an alternative to other distributions used in reliability engineering and life testing such as gamma and lognormal distributions. Following the method of Azzalini (1985), Gupta and Kundu (2009) proposed a weighted exponential distribution which can also be used as an alternative to gamma and Weibull distributions. Using the same idea, Shahbaz et al. (2010) proposed the weighted Weibull distribution.

Suppose two random variables X_1 and X_2 are independently identically distributed as Weibull random variable with distribution function $F(x) = 1 - e^{-\lambda x^\beta}$ and density function $f(x) = \lambda\beta x^{\beta-1}e^{-\lambda x^\beta}$, where β is the shape parameter and λ is the scale parameter. The density function of the weighted Weibull distribution is given by

$$f(x) = \frac{\alpha+1}{\alpha}\lambda\beta x^{\beta-1}e^{-\lambda x^\beta}(1 - e^{-\alpha\lambda x^\beta}), \quad x > 0,\ \alpha,\beta,\lambda > 0. \tag{1}$$

The reliability, or survival function associated with (1) is

$$\bar{F}(x) = 1 - \left\{\frac{\alpha+1}{\alpha}\left[1 - \exp\left\{-\lambda x^\beta\right\}\right] - \frac{1}{\alpha}\left[1 - \exp\left\{-(1+\alpha)\lambda x^\beta\right\}\right]\right\}. \tag{2}$$

It should be mentioned that for $\beta = 1$, the distribution $f(x)$ given by (1) is reduced to the weighted exponential distribution of Gupta and Kundu (2009). The hazard rate has decreasing trend when $\beta < 1$ and increasing trend for values of $\beta > 1$.

The model can be considered as another useful two-parameter generalization of the Weibull distribution. This lifetime distribution can model various shapes of failure rates and hence various shapes of aging criteria. In the work of Shahbaz et al. (2010) some properties such as reliability function, hazard function and moment generating function are discussed. Also, estimation of the unknown parameters of the weighted Weibull is discussed, but making comparisons basing on different methods of estimation has not been performed. The main goal of this paper is to estimate the parameters using maximum likelihood and Bayesian method and then make use of the estimated parameters to estimate the reliability function. For the same purpose, some nonparametric methods like empirical method, kernel density estimator and a modified shrinkage estimator are used.

The rest of the paper is organized as follows. In Section 2, the maximum likelihood estimators (MLEs) are obtained. In Section 4, we obtain Bayes estimators using the symmetric and asymmetric balanced loss functions. In Section 5, the MCMC methods are used to accomplish some complex calculations, and, therefore, comparisons are made between Bayesian and maximum likelihood estimators via Monte Carlo simulation study.

2. Methods of Estimation

In this section, we consider the estimation problem of the weighted Weibull distribution. We discuss the maximum likelihood estimators, empirical method, kernel density estimator and modified shrinkage estimators.

2.1 Maximum Likelihood Estimator

Suppose that X is a random variable distributed according to weighted Weibull distribution, $X \sim WW(\alpha, \beta, \lambda)$. Suppose further that $x_1 < x_2 < \cdots < x_n$ denote the observed failure times of the experimental units and $\theta = (\alpha, \beta, \lambda)$. The likelihood function is given by

$$L(x|\,\theta) = \left(\frac{\alpha+1}{\alpha}\right)^n (\beta\lambda)^n \prod_{i=1}^{n} x_i^{\beta-1} \exp\left\{-\lambda \sum_{i=1}^{n} x_i^{\beta}\right\} \prod_{i=1}^{n} \left(1 - \exp\left\{-\alpha\lambda x_i^{\beta}\right\}\right). \tag{3}$$

The log-likelihood function of the $WW(\alpha, \beta, \lambda)$ is given by

$$\begin{aligned} \log L(x;\theta) &= -n\log(\alpha) + n\log(\alpha+1) + n\log(\beta) + n\log(\lambda) + (\beta-1)\sum_{i=1}^{n}\log(x_i) \\ &\quad -\lambda\sum_{i=1}^{n} x_i^{\beta} + \sum_{i=1}^{n}\log\left(1 - \exp\left\{-\alpha\lambda x_i^{\beta}\right\}\right). \end{aligned} \tag{4}$$

The MLEs of (α, β, λ), say $(\hat{\alpha}_{ML}, \hat{\beta}_{ML}, \hat{\lambda}_{ML})$, are obtained as the solution of the Fishers score function. Setting the equations to zero and solving for α, β and λ leads to the MLEs.

$$\frac{\partial \log L}{\partial \alpha} = -\frac{n}{\alpha} + \frac{n}{\alpha+1} + \sum_{i=1}^{n} \frac{\lambda x_i^{\beta} \exp\left\{-\alpha\lambda x_i^{\beta}\right\}}{1 - \exp\left\{-\alpha\lambda x_i^{\beta}\right\}} = 0.$$

$$\frac{\partial \log L}{\partial \beta} = \frac{n}{\beta} + \sum_{i=1}^{n}\log(x_i) - \lambda\sum_{i=1}^{n} x_i^{\beta}\log(x_i) + \sum_{i=1}^{n} \frac{\alpha\lambda x_i^{\beta}\log(x_i)\exp\left\{-\alpha\lambda x_i^{\beta}\right\}}{1 - \exp\left\{-\alpha\lambda x_i^{\beta}\right\}} = 0.$$

$$\frac{\partial \log L}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i^{\beta} + \sum_{i=1}^{n} \frac{\alpha x_i \exp\left\{-\alpha\lambda x_i^{\beta}\right\}}{1 - \exp\left\{-\alpha\lambda x_i^{\beta}\right\}} = 0.$$

Usual algebraic solution for the above equations is not working due to the properties of transcendental equation. Therefore, we propose to use numerical methods to compute the MLEs. We have used `nlm()` function of R package. The corresponding "ML plug-in estimation" of $\bar{F}$, say $\hat{\bar{F}}$, is given by

$$\begin{aligned} \hat{\bar{F}}_{ML}(x) &= \frac{\hat{\alpha}_{ML}+1}{\hat{\alpha}_{ML}}\left[1 - \exp\left\{-\hat{\lambda}_{ML}x^{\hat{\beta}_{ML}}\right\}\right] \\ &\quad -\frac{1}{\hat{\alpha}_{ML}}\left[1 - \exp\left\{-(1+\hat{\alpha}_{ML})\hat{\lambda}_{ML}x^{\hat{\beta}_{ML}}\right\}\right]. \end{aligned} \tag{5}$$

2.2 Kernel Density Estimator

Here we attempt to estimate the density directly from the data without assuming a particular form for the underlying distribution. Let $X_1, X_2, \cdots, X_n$ denote a sample of size n from a random variable with density f. The kernel density estimate of f at the point x is given by

$$\hat{f}_h(x) = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{6}$$

where the kernel K satisfies $\int K(x)dx = 1$ and the smoothing parameter h is known as the bandwidth. In practice, the kernel K is generally chosen to be a unimodal probability density symmetric about zero. In this case, any function K having the following assumptions can be used as a kernel:

$$\int K(y)dy = 1, \quad \int yK(y)dy = 0, \quad \text{and} \quad \int y^2K(y)dy = \mu_2(K) < \infty.$$

The Gaussian kernel is a popular choice for K. The selection of smoothing parameter, or bandwidth h, for the kernel density is very crucial because it effects on the shape of the corresponding estimator. If the bandwidth is small, we will get an under smoothed estimator, with high variability. On the other hand, if the value of h is big, the resulting estimator will be over smooth. We will use the optimal bandwidth to estimate h, which can be given as (see e.g. Marron and Chung 2001)

$$\hat{h}_{opt} = 1.06 \min\left\{\sigma, \frac{Q}{1.349}\right\} n^{-1/5}, \tag{7}$$

where σ is the standard deviation and Q provide the interquartile range of X.

When the data are complete (uncensored), a kernel estimate (KE) for the survival function associated with (6) is given by

$$\hat{\bar{F}}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} \int_x^{\infty} K\left(\frac{t - X_i}{h}\right) dt. \tag{8}$$

2.3 Empirical Estimator

We consider nonparametric estimation method which is based on the empirical distribution function (EDF). The EDF is a step function with jumps at the order statistics $(X_1, X_2, \ldots, X_n)$ and defined as

$$F_n(x) = \frac{\text{number of observations} \leq x}{n} = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x),$$

where I is an indicator function. By Glivenko-Cantelli theorem the EDF, $F_n(x)$ converges to $F(x)$ as $x \to \infty$ almost surely, $\sup_x |F_n(x) - F(x)| \to 0$. The reliability function is estimated as the proportion of observations surviving longer than x i.e.

$$\hat{\bar{F}}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i > x). \tag{9}$$

Note that the most common way to estimate the reliability function nonparametrically is the Kaplan-Meier (K-M) method. However, if there is no censoring, as in our present case, the K-M estimate coincides with the empirical survival function.

2.4 Modified Shrinkage Estimator

Jani (1991) suggested a class of shrinkage estimators for the scale parameter of the exponential distribution. The estimator of θ, say T_p, is as follows:

$$T_p = \theta_0 \left\{1 + W\left(\frac{\theta_0}{\bar{x}}\right)^p\right\}, \tag{10}$$

where $\bar{x}$ is the sample mean, p is a non-zero real number and W is constant such that the mean square error of T_p is at minimum. The constant W can be estimated by $\hat{W}$ where

$$\hat{W} = \left(\frac{\bar{x} - \theta_0}{\bar{x}}\right)\left(\frac{\bar{x}}{\theta_0}\right)^{p+1} \frac{\sqrt{n-p}}{n^p \sqrt{n-2p}}. \tag{11}$$

This shrinkage technique can be adopted and used to present a shrinkage estimator for the reliability function.

$$\hat{\bar{F}}_p(x) = \hat{\bar{F}}_h(x) \left\{1 + W_1 \left(\frac{\hat{\bar{F}}_h(x)}{\hat{\bar{F}}_n(x)}\right)^p\right\}, \tag{12}$$

where $\hat{\bar{F}}_h(x)$ is defined as given in relation (8) and it is considered to be an initial value for the reliability function, and $\hat{\bar{F}}_n(x)$ is the empirical reliability.

$$\hat{W}_1 = \left(\frac{\hat{\bar{F}}_n(x) - \hat{\bar{F}}_h(x)}{\hat{\bar{F}}_n(x)}\right)\left(\frac{\hat{\bar{F}}_n(x)}{\hat{\bar{F}}_h(x)}\right)^{p+1} \frac{\sqrt{n-p}}{n^p \sqrt{n-2p}}. \tag{13}$$

It is obvious that for different values of p one can obtain many more shrinkage estimators. Also, it should be mentioned here that the class of shrinkage estimators given by (12) is not unique.

3. Bayesian Approaches

The posterior expectations involve integrals which sometimes can not be obtained in closed forms. To treat this problem we employ the MCMC technique to compute the Bayes estimates for the involved parameters.

3.1 Gamma Prior

Bayesian approach requires to specify the prior probability distribution of the unknown parameters. We assume that α, β and λ have independent gamma prior distributions, i.e.

$$\begin{array}{l} \pi_1(\alpha) \propto \alpha^{a_1-1} e^{-b_1\alpha}, \\ \pi_2(\beta) \propto \beta^{a_2-1} e^{-b_2\beta}, \\ \pi_3(\lambda) \propto \lambda^{a_3-1} e^{-b_3\lambda}. \end{array} \tag{14}$$

The hyper parameters a_1, a_2, a_3, b_1, b_2 and b_3 are assumed to be known and non-negative. Then the joint prior distribution of α, β and λ can be written as

$$\pi_G(\alpha, \beta, \lambda) \propto \alpha^{a_1-1}\beta^{a_2-1}\lambda^{a_3-1} e^{-(b_1\alpha+b_2\beta+b_3\lambda)}. \tag{15}$$

Based on the prior distributions given by (15), the joint density function of the sample observations and the parameters α, β and λ becomes

$$\begin{aligned} \mathcal{L}(x; \alpha, \beta, \lambda) \quad \propto \quad & \alpha^{a_1-1}\beta^{a_2-1}\lambda^{a_3-1} \left(\frac{\alpha+1}{\alpha}\right)^n (\beta\lambda)^n \prod_{i=1}^{n} x_i^{\beta-1} \\ & \times \exp\left\{-\lambda \sum_{i=1}^{n} x_i^{\beta} - (b_1\alpha + b_2\beta + b_3\lambda)\right\} \\ & \times \prod_{i=1}^{n} \left(1 - \exp\left\{-\alpha\lambda x_i^{\beta}\right\}\right). \end{aligned} \tag{16}$$

The joint posterior density function of α, β and λ, given the data can be obtained from

$$\pi_G(\alpha, \beta, \lambda | \, x) = \frac{\mathcal{L}(x | \, \alpha, \beta, \lambda)\pi_G(\alpha, \beta, \lambda | \, a, b)}{\int \mathcal{L}(x | \, \alpha, \beta, \lambda)\pi_G(\alpha, \beta, \lambda | \, a, b) \, d\alpha d\beta d\lambda}. \tag{17}$$

The Bayes estimators under the squared error loss function (SELF) is the posterior mean, $\hat{\theta} = E_{\pi(\theta|data)}(\theta)$. It is not possible to compute (17) analytically and therefore the Bayes estimates of the parameters under the SELF. For this reason, we propose to use Metropolis-Hastings algorithm, one of the MCMC methods, to obtain samples from the posterior distribution and then to compute the Bayes estimates of α, β and λ.

3.2 Jeffreys Prior

A well-known prior to represent a situation where no much information about the parameters was proposed by Jeffreys (1967). This prior, denoted by $\pi_J(\alpha, \beta, \lambda)$, is derived from the Fisher information matrix $I(\alpha, \beta, \lambda)$ given as

$$\pi_J(\alpha, \beta, \lambda) \propto \sqrt{\det I(\alpha, \beta, \lambda)}.$$

The joint posterior density function of α, β and λ, given the data can be obtained from

$$\pi_J(\alpha, \beta, \lambda | \, x) = \frac{\mathcal{L}(x | \, \alpha, \beta, \lambda)\pi_J(\alpha, \beta, \lambda | \, a, b)}{\int \mathcal{L}(x | \, \alpha, \beta, \lambda)\pi_J(\alpha, \beta, \lambda | \, a, b) \, d\alpha d\beta d\lambda}. \tag{18}$$

Also, it is not possible to compute (18) analytically and therefore the Bayes estimates of the parameters under the SELF. Thus, we use Metropolis-Hastings algorithm to obtain samples from the posterior distribution and then to compute the Bayes estimates of α, β and λ.

4. Simulation and Data Analysis

Instead of drawing direct samples from Bayesian posterior distribution, which is not often an easy task, one may use the Metropolis-Hastings algorithm, a general term of Markov chain simulation methods. This method is an extension of the usual rejection-acceptance sampling method. For a comprehensive treatment on MCMC methods, Metropolis-Hastings algorithm, one may refer to the book by Robert and Casella (2005), Hastings (1970) and Cowles and Carlin (1995). The algorithm is proposed as follows (See Al-Zahrani Gindwan 2014):

Table 1. The estimates of the parameters α, β and λ for Guinea pigs data.

	Estimates			Statistics	
Method	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	D_n	W_n^2
MLE	3.51900	1.15289	0.00578	0.1252	0.3018
BGP	3.39660	1.24232	0.00784	0.3608	3.3907
BJP	3.66899	1.06389	0.01206	0.1988	0.5404

Step 1. Draw starting points $\theta^0 = (\alpha^0, \beta^0, \lambda^0)$ at timestep $i = 0$, for which $f(\theta^0|data) > 0$, from a prior distribution $\pi(\theta)$.

Step 2. At iteration i, draw a proposal θ^* from a jumping distribution $J_i(\theta^*|\theta^{(i-1)})$, where J_i is symmetric.

Step 3. Generate a sample u from the uniform distribution $U(0, 1)$ and take $z = \log u$.

Step 4. Compute an acceptance ratio ρ, where

$$\rho = \log\left[\frac{\pi(\theta^*|data)J_i(\theta^{(i-1)}|\theta^*)}{\pi(\theta^{(i-1)}|data)J_i(\theta^*|\theta^{(i-1)})}\right].$$

Step 5. If $z < \rho$ accept θ^* as $\theta^{(i)}$ with probability $\min(\rho, 1)$, otherwise, $\theta^{(i)} = \theta^{(i-1)}$.

Step 6. Repeat steps 2-5 N times to get sequence of the parameter $\theta = (\alpha, \beta, \lambda)$ from $\pi(\theta|data)$, with optional burn-in.

Step 7. The Bayes estimates of $\theta = (\alpha, \beta, \lambda)$, say $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\lambda})$ are taken as the mean of the generated values of θ.

At the length of 40,000, we produce the Markov chain with different initial points of the involving parameters. The convergence is adjusted by drawing trace and ergodic mean plots. It is noted that the Markov chains converge after approximately 2000 observations. Therefore, burn-in of 5000 samples is quite enough to eliminate the effect of initial values. In order to minimize the auto correlation among the generated deviates, we take samples of size 3500 from the posterior with thin = 10, and starting from 3501.

The data set consists of survival times of guinea pigs injected with different amount of tubercle bacilli and was studied by Bjerkedal (1960). Guinea pigs are known to have high susceptibility of human tuberculosis, which is one of the reasons for choosing this species. We consider only the study in which animals in a single cage are under the same regimen. The data represents the survival times of Guinea pigs in days. The data are given as follows: 12, 15, 22, 24, 24, 32, 32, 33, 34, 38, 38, 43, 44, 48, 52, 53, 54, 54, 55, 56, 57, 58, 58, 59, 60, 60, 60, 60, 61, 62, 63, 65, 65, 67, 68, 70, 70, 72, 73, 75, 76, 76, 81, 83, 84, 85, 87, 91, 95, 96, 98, 99, 109, 110, 121, 127, 129, 131, 143, 146, 146, 175, 175, 211, 233, 258, 258, 263, 297, 341, 341, 376.

The estimates of the parameters, maximum likelihood estimator (MLE), Bayes with gamma prior (BGP) and Bayes with Jeffreys prior (BJP), are given in Table 1. Also, the Kolmogorov-Smirnov D_n and Cramér-von Mises W_n^2 tests statistics are computed and given in Table 1. The results show that the test statistics take the smallest values for the data set under MLEs with regard to the Baysian methods. The plug-in MLE, empirical method (EM), kernel estimator (KE), modified shrinkage estimator (MSE) and the plug-in Bayes estimators are given in Table 2 and shown in Figure 1. The findings show that all the methods of estimation considered in this example are precisely estimating the parameters and their performance are quite similar. The Bayes estimator with gamma prior reaches 0, as the value of X gets larger, faster than those obtained by the Jeffreys prior, nonparametric estimators and MLEs.

5. Concluding Remarks

In this paper we have provided the maximum likelihood plug-in estimator, empirical estimator, kernel density estimator and a modified shrinkage estimator for the weighted Weibull distribution. The Markov chain Monte Carlo method is used to compute the Bayes estimators assuming the prior distribution of the parameters are gamma and Jeffreys priors. To assess the performance of the obtained estimators we analyzed a real data set.

Table 2. The plug-in MLE, empirical method (EM), kernel estimator (KE), modified shrinkage estimator (MSE) and the plug-in Bayes estimators

	Non-parametric Estimators			Bayes (MCMC)	
MLE	EM	KE	MSE	Gamma	Jeffreys
0.980626	0.986111	0.835457	0.852625	0.951788	0.950535
0.969219	0.972222	0.826638	0.843500	0.923247	0.926407
0.934172	0.958333	0.800774	0.816551	0.839636	0.860746
0.922351	0.930556	0.791919	0.807345	0.812895	0.840514
0.922351	0.930556	0.791919	0.807345	0.812895	0.840514
0.869238	0.902778	0.749439	0.763607	0.701237	0.757327
0.869238	0.902778	0.749439	0.763607	0.701237	0.757327
0.862091	0.888889	0.743306	0.757354	0.687200	0.746899
0.854857	0.875000	0.750874	0.763172	0.673219	0.736500
0.825210	0.847222	0.723724	0.735726	0.618177	0.695357
0.825210	0.847222	0.723724	0.735726	0.618177	0.695357
0.787044	0.833333	0.685653	0.697644	0.552317	0.645422
0.779323	0.819444	0.677510	0.689551	0.539633	0.635682
0.748311	0.805556	0.643316	0.655733	0.490731	0.597665
0.717308	0.791667	0.606825	0.619892	0.444938	0.561260
0.709587	0.777778	0.597402	0.610669	0.433991	0.552422
0.701885	0.750000	0.587876	0.601360	0.423248	0.543691
0.701885	0.750000	0.587876	0.601360	0.423248	0.543691
0.694205	0.736111	0.578259	0.591971	0.412708	0.535069
0.686550	0.722222	0.582449	0.594760	0.402370	0.526555
0.678922	0.708333	0.572679	0.585244	0.392235	0.518150
0.671324	0.680556	0.562851	0.575679	0.382302	0.509855
0.671324	0.680556	0.562851	0.575679	0.382302	0.509855
0.663759	0.666667	0.552974	0.566076	0.372569	0.501668
0.656228	0.611111	0.543062	0.556445	0.363036	0.493591
0.656228	0.611111	0.543062	0.556445	0.363036	0.493591
0.656228	0.611111	0.543062	0.556445	0.363036	0.493591
0.656228	0.611111	0.543062	0.556445	0.363036	0.493591
0.648735	0.597222	0.533125	0.546798	0.353701	0.485623
0.641280	0.583333	0.523176	0.537144	0.344563	0.477764
0.633867	0.569444	0.513228	0.527496	0.335619	0.470014
0.619170	0.541667	0.493381	0.508257	0.318308	0.454837
0.619170	0.541667	0.493381	0.508257	0.318308	0.454837
0.604657	0.527778	0.473678	0.489168	0.301750	0.440087
0.597473	0.513889	0.463911	0.479706	0.293748	0.432871
0.583257	0.486111	0.444598	0.460996	0.278285	0.418753

Table 2 Continued. The plug-in MLE, empirical method (EM), kernel estimator (KE), modified shrinkage estimator (MSE) and the plug-in Bayes estimators

	Non-parametric Estimators			Bayes (MCMC)	
MLE	EM	KE	MSE	Gamma	Jeffreys
0.583257	0.486111	0.444598	0.460996	0.278285	0.418753
0.569251	0.472222	0.425651	0.442634	0.263525	0.405049
0.562328	0.458333	0.416339	0.433605	0.256403	0.398349
0.548650	0.444444	0.398083	0.415890	0.242658	0.385250
0.541896	0.416667	0.389154	0.407218	0.236030	0.378848
0.541896	0.416667	0.389154	0.407218	0.236030	0.378848
0.508994	0.402778	0.374538	0.390439	0.205223	0.348282
0.496247	0.388889	0.358720	0.374985	0.193935	0.336707
0.489964	0.375000	0.351069	0.367495	0.188502	0.331055
0.483741	0.361111	0.343592	0.360166	0.183206	0.325492
0.471476	0.347222	0.343053	0.358241	0.173011	0.314627
0.447675	0.333333	0.316303	0.331840	0.154141	0.293913
0.424844	0.319444	0.292284	0.307961	0.137156	0.274492
0.419286	0.305556	0.286687	0.302369	0.133185	0.269831
0.408352	0.291667	0.275957	0.291615	0.125556	0.260734
0.402974	0.277778	0.270818	0.286447	0.121894	0.256295
0.352404	0.263889	0.226609	0.241486	0.090304	0.215709
0.347661	0.250000	0.222805	0.237571	0.087602	0.212008
0.299064	0.236111	0.200144	0.211843	0.062446	0.175134
0.275188	0.222222	0.183280	0.194149	0.051752	0.157720
0.267619	0.208333	0.178023	0.188619	0.048587	0.152297
0.260238	0.194444	0.172932	0.183257	0.045605	0.147056
0.219711	0.180556	0.145704	0.154457	0.031035	0.119099
0.210528	0.152778	0.139790	0.148155	0.028153	0.112963
0.210528	0.152778	0.139790	0.148155	0.028153	0.112963
0.138373	0.125000	0.127518	0.128802	0.010712	0.067509
0.138373	0.125000	0.127518	0.128802	0.010712	0.067509
0.080947	0.111111	0.116906	0.112653	0.003055	0.035361
0.057915	0.097222	0.102398	0.097137	0.001382	0.023733
0.039354	0.069444	0.078970	0.074285	0.000549	0.015041
0.039354	0.069444	0.078970	0.074285	0.000549	0.015041
0.036401	0.055556	0.073904	0.069469	0.000455	0.013725
0.021295	0.041667	0.049460	0.046129	0.000124	0.007343
0.010493	0.013889	0.027447	0.025442	0.000022	0.003247
0.010493	0.013889	0.027447	0.025442	0.000022	0.003247
0.005915	0.000000	0.007929	0.007691	0.000005	0.001688

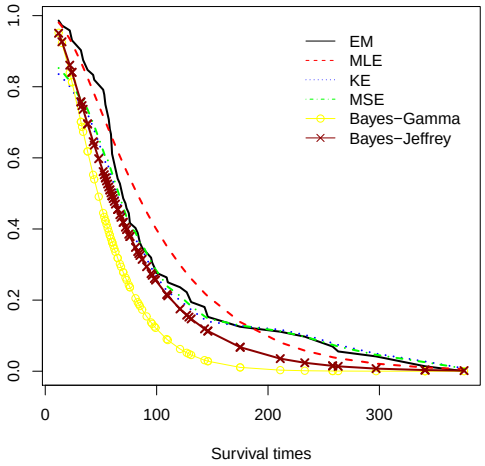


Figure 1. Empirical and fitted survival functions

Acknowledgment

The author is grateful to the editor for his valuable comments and suggestions which improved the presentation of the paper.

References

Al-Zahrani, B., & Gindwan, M. (2014). Parameter estimation of a two-parameter Lindley distribution under hybrid censoring. *Inter. J Syst. Assur. Eng. Manag., 5*, 628–636. http://dx.doi.org/10.1007/s13198-013-0213-2

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Stat., 12*, 171–178.

Bjerkedal, T (1960). Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. *Amer. J. Hyg., 72*, 130–148.

Cowles, M. K., & Carlin, B. P. (1995). Markov Chain Monte Carlo Diagnostics: A Comparative Review. *J. Amer. Stat. Assoc., 91*, 883–904. http://dx.doi.org/10.1080/01621459.1996.10476956

Gupta, R. D., & Kundu, D. (2009). A new class of weighted exponential distributions. *Statistics, 43*, 621–634. http://dx.doi.org/10.1080/02331880802605346

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*, 97–109. http://dx.doi.org/10.1093/biomet/57.1.97

Jani, P. N. (1991). A class of shrinkage estimators for the scale parameter of the exponential distribution. *IEEE T. Reliab., 40*, 68–70. http://dx.doi.org/10.1109/24.75337

Jeffreys, S. H. (1967). *Theory of Probability, third ed., Oxford University Press*, London.

Marron, J. S., & Chung, S. S. (2001). Presentation of smoothers: The family approach. *Computation. Stat., 16*, 195–207. http://dx.doi.org/10.1007/s001800100059

Robert, C. P., & Casella, G. (2005). *Monte Carlo Statistical Methods*. Springer-Verlag New York, Inc., Secaucus, New York.

Shahbaz, S. Shahbaz, M. Q., & Butt, N. S. (2010). A class of Weighted Weibull distribution. *Pak. J. Stat. Oper. Res., 6*, 53–59. http://dx.doi.org/10.18187/pjsor.v6i1.126

International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Gradient and Likelihood Ratio Tests in Cure Rate Models

Hérica P. A. Carneiro[1] & Dione M. Valença[2]

[1] Departamento de Estatística, Universidade de São Paulo, São Paulo/SP, Brazil

[2] Departamento de Estatística, Universidade Federal do Rio Grande do Norte, Natal/RN, Brazil

Correspondence: Dione M. Valença, Departamento de Estatística, UFRN, CCET, Campus Universitário, 59078-970, Natal/RN, Brazil. E-mail: dione@ccet.ufrn.br

Received: November 3, 2016 Accepted: February 4, 2016 Online Published: June 11, 2016

doi:10.5539/ijsp.v5n4p9 URL: http://dx.doi.org/10.5539/ijsp.v5n4p9

The research is financed by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

Abstract

In some survival studies part of the population may be no longer subject to the event of interest. The called *cure rate models* take this fact into account. They have been extensively studied for several authors who have proposed extensions and applications in real lifetime data. Classic large sample tests are usually considered in these applications, especially the likelihood ratio. Recently a new test called *gradient test* has been proposed. The gradient statistic shares the same asymptotic properties with the classic likelihood ratio and does not involve knowledge of the information matrix, which can be an advantage in survival models. Some simulation studies have been carried out to explore the behavior of the gradient test in finite samples and compare it with the classic tests in different models. However little is known about the properties of these large sample tests in finite sample for cure rate models. In this work we performed a simulation study based on the promotion time model with Weibull distribution, to assess the performance of likelihood ratio and gradient tests in finite samples. An application is presented to illustrate the results.

Keywords: Survival analysis, Unified model, Promotion time model, Gradient statistic.

1. Introduction

Cure rate models have been extensively studied in the literature for data sets where the event of interest may not occur for part of the population studied. That is, part of the population studied will never experience the event of interest, being that recurrence of a disease, product consumption or many other situations. An early approach developed in Boag (1949) and Berkson and Gage (1952), considers a mixture of two distributions, one representing the survival time of the individuals in risk, and a degenerated one, allowing infinite time for some fraction of population considered cured. This model is known as the *standard mixture model* and the book by Maller and Zhou (1996) presents an up to date review of the main results on the subject.

Alternatively Yakovlev et al. (1993) and Chen et al. (1999) introduce a class of models involving a competitive risk type structure. Applications to cancer clinical trials have been specially successful. They are used for modeling time-to-event data for several types of cancer, including breast cancer, leukemia, prostate cancer and many others. Such models have been discussed in the statistical literature by many authors. In Tsodikov et al. (2003) this model is refered as *bounded cumulative hazard model*. They provided an overview of the development of this cure rate model from both the frequentist and Bayesian perspectives. In Yin and Ibrahim (2005) this model is refered as *promotion time model*. A unified approach that includes standard mixture model and promotion time model as special cases, is pursued in Rodrigues et al. (2009).

In the present paper, the interest lies in testing hypotheses. The commonly used large sample tests are based on the likelihood ratio statistics (Wilks, 1938), Wald (Wald, 1943) or Rao score (Rao, 1948) tests. Particularly, the likelihood ratio is usually considered in applications to test parameters in cure rate survival models. There are hardly any works about finite-sample performance of likelihood ratio under this models. We can mention just Sposto et al. (1992), which present a small (100 samples) and restrict simulation study to compared the likelihood ratio, Wald, and score tests based on a mixture model. Although it is known the liberal tendency of the likelihood ratio test when the sample is not large, they found that the tests keep their asymptotic properties even in small samples, for some specific situations.

Recently the gradient test was proposed in Terrell (2002). As well as the usual classical statistics, the gradient statistic has asymptotically chi-square distribution. This new statistic was obtained from Rao score and Wald modified statistics

(Hayakawa and Puri, 1985). A comparison of local power properties of the gradient test with classical tests was studied in Lemonte and Ferrari (2012) and no uniform superiority was found. Some simulation studies have been conducted (Lemonte and Ferrari, 2011a, 2012; Ferrari and Pinheiro, 2014) in order to explore the characteristics of this new statistic and compare the competing tests to different models. Because statistical gradient do not need the computation of the information matrix (neither observed nor expected), it may be advantageous in problems involving censored samples, which are often observed in survival studies. Among the studies that consider the statistical gradient in survival models with censored data, we can cite Lemonte and Ferrari (2011b), that consider samples with right censoring of type II to test hypotheses about the two parameters of the Birnbaum-Saunders distribution, and Medeiros et al. (2014) that compare the performance of the gradient and likelihood ratio tests in accelerated failure time models under random censoring. The book by (Lemonte, 2016) provides a broad survey about results of gradient test in literature. There are no studies involving the gradient test with cure fraction, with or without the presence of covariates. In this paper we study the performance of the likelihood ratio and the gradient tests via simulation study, to test coefficients related to cure rate parameter in the Weibull promotion time cure model.

The paper is organized as follows. The unified approach for cure rate model is described in detail in Section 2. Section 3 briefly describes the likelihood ratio and gradient tests, and presents the resulting tests to the model based on marginal likelihood obtained after eliminating the latent variables. Section 4 presents a simulation study. In Section 5 we illustrate our results with a real data set about the time to pediatric leukemia recurrence. Some conclusions are reported in Section 6 and some basic results are presented in an Appendix.

2. Cure Fraction Model: Unified Approach

Survival models with cure fraction are models that consider cured (or immune) a fraction of the population. The occurrence of a high percentage of censoring at the end of the study in a sufficient follow-up time is an indication of cure fraction in population (Maller and Zhou, 1996). Considering the unified model, suppose we have n individuals and that for each individual ($i = 1, \ldots, n$) it is associated a (latent) random variable M_i, representing the number of causes or risk factors competing for the occurrence of the event of interest, with probability function $p_\theta(m) = P_\theta(M_i = m)$. Given $M_i = m$, suppose also that the random variables $Z_{i1}, Z_{i2}, \ldots, Z_{im}$, are independent and identically distributed (*i.i.d.*), representing (unobserved) time-to-event for the i-th individual, due to j-th cause ($j = 1, \ldots, M_i$), with common distribution function $F(z|\boldsymbol{\lambda})$ and a survival function $S(z|\boldsymbol{\lambda}) = 1 - F(z|\boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is a vector of parameters and $\lim_{t\mapsto\infty} S(t|\boldsymbol{\lambda}) = 0$. Let T_i be an observable random variable representing the time until the occurrence of the event, defined as $T_i = \min\{Z_{i0}, Z_{i1}, \ldots, Z_{iM_i}\}$ where the sequence $Z_{i1}, Z_{i2}, \ldots$ does not depend on M_i. Besides, Z_{i0} is set so that $P(Z_{i0} = \infty) = 1$. This assumption permits the occurrence of an *infinite lifetime* in the immune individuals, because when $M_i = 0$ there are no causes or risks for the occurrence of the event.

The common survival function for T_i is given by

$$\begin{aligned} S_p(t) &= P(T_i > t) = P(T_i > t, M_i = 0) + P(T_i > t, M_i \geq 1) \\ &= P(T_i > t|M_i = 0)P_\theta(M_i = 0) + P(T_i > t|M_i \geq 1)P_\theta(M_i \geq 1) \\ &= p_\theta(0) + \sum_{m=1}^{\infty} p_\theta(m) S(t|\boldsymbol{\lambda})^m, \end{aligned} \tag{1}$$

since $P(T_i > t|M_i = 0) = 1$ and $P_\theta(M_i = 0) = p_\theta(0)$.

Hence $S_p(t)$ is an improper survival function, i.e., $\lim_{t\mapsto\infty} S_p(t) > 0$. This survival function can be interpreted as an infinite linear combination of Lehmann type II distributions (Rodrigues et al., 2011; Alexandre et al., 2012). The proportion of cured individuals (cure fraction) is given by $\lim_{t\mapsto\infty} S_p(t) = p_\theta(0)$. From (1) we can obtain the sub-density function for the random variables T_i as

$$f_p(t) = f(t|\boldsymbol{\lambda}) \sum_{m=1}^{\infty} m p_\theta(m) \left[S(t|\boldsymbol{\lambda})\right]^{m-1}.$$

2.1 Likelihood for Unified Model

Furthermore, consider that for $i = 1, \ldots, n$, $Y_i = \min\{T_i, C_i\}$ is the observable lifetime for individual i, where C_i is right censoring time (random and uninformative) independent of T_i, and let δ_i be the censoring indicator, with $\delta_i = 1$ if $T_i \leq C_i$ and $\delta_i = 0$ if $T_i > C_i$. Also consider $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^\top$ the vector of associated covariates.

To simplify notation, we define the n-dimensional vectors of observations $\mathbf{y} = (y_1, y_2, \ldots, y_n)^\top$, $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_n)^\top$ and $\mathbf{m} = (m_1, m_2, \ldots, m_n)^\top$ and the covariate matrix $\mathbf{X} = (x_1, x_2, \ldots, x_n)^\top$ of dimension $n \times p$. Hence, the complete data set is denoted by $\mathcal{D}_c = (n, \mathbf{y}, \boldsymbol{\delta}, \mathbf{m}, \mathbf{X})$ and the data without the latent variables is denoted by $\mathcal{D} = (n, \boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{X})$. The covariates can be included in the model through some relation $\theta_i \equiv \theta(\boldsymbol{x}_i^\top \boldsymbol{\beta})$, where $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^\top$ is the vector of regression coefficients.

Thus, the vector of unknown parameters in the model is given by $\boldsymbol{\phi} = (\boldsymbol{\beta}^\top, \boldsymbol{\lambda}^\top)^\top$ and, to use a better notation, we consider $p_{\theta_i}(m_i) = p(m_i|\boldsymbol{\beta}, \boldsymbol{x}_i)$. After some algebraic manipulations it can be shown that the likelihood for the complete data $\mathcal{D}_c$ is given by

$$L(\boldsymbol{\phi}; \mathcal{D}_c) = \prod_{i=1}^{n} [m_i f(y_i|\boldsymbol{\lambda})]^{\delta_i} [S(y_i|\boldsymbol{\lambda})]^{m_i - \delta_i} p(m_i|\boldsymbol{\beta}, \boldsymbol{x}_i). \tag{2}$$

Note that the likelihood (2) is not observable, since it depends on the latent variables. In practice, a marginal likelihood is used. It is obtained by summing overall possible values for the variables M_i, $i = 1, \ldots, n$, and given below in (3). For details see Appendix.

$$L^*(\boldsymbol{\phi}; \mathcal{D}) = \prod_{i=1}^{n} \left[f_p(y_i|\boldsymbol{\phi})\right]^{\delta_i} \left[S_p(y_i|\boldsymbol{\phi})\right]^{1-\delta_i}. \tag{3}$$

Therefore, the logarithm of the marginal likelihood function is given for

$$\ell^*(\boldsymbol{\phi}; \mathcal{D}) = \sum_{i=1}^{n} \delta_i \log[f_p(y_i|\boldsymbol{\phi})] + (1 - \delta_i) \log[S_p(y_i|\boldsymbol{\phi})]. \tag{4}$$

3. Likelihood Ratio and Gradient Tests

Let $\ell(\boldsymbol{\phi}) = \log L(\boldsymbol{\phi})$ be a log-likelihood function of $\boldsymbol{\phi}$ a p-vector of unkown parameter, and define $U(\boldsymbol{\phi}) = \dfrac{\partial}{\partial \boldsymbol{\phi}} \ell(\boldsymbol{\phi})$ as the score function. Considering the partition $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)^\top$, where the dimensions of $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$, are q and $p - q$ respectively, we have a corresponding partition $U(\boldsymbol{\phi}) = (U_{\boldsymbol{\phi}_1}(\boldsymbol{\phi}), U_{\boldsymbol{\phi}_2}(\boldsymbol{\phi}))^\top$. The likelihood ($S_{LR}$) and gradient ($S_G$) statistics for testing the composite hypothesis

$$H_0 : \boldsymbol{\phi}_1 = \boldsymbol{\phi}_{10} \; against \; H_1 : \boldsymbol{\phi}_1 \neq \boldsymbol{\phi}_{10} \tag{5}$$

are respectively given by

$$S_{LR} = 2\left[\ell\left(\widehat{\boldsymbol{\phi}}_1, \widehat{\boldsymbol{\phi}}_2\right) - \ell\left(\boldsymbol{\phi}_{10}, \tilde{\boldsymbol{\phi}}_2\right)\right] \;\text{ and }\; S_G = U^\top_{\boldsymbol{\phi}_1}(\tilde{\boldsymbol{\phi}})(\widehat{\boldsymbol{\phi}}_1 - \boldsymbol{\phi}_{10}),$$

where $\boldsymbol{\phi}_{10}$ is a specified vector, $\widehat{\boldsymbol{\phi}} = (\widehat{\boldsymbol{\phi}}_1, \widehat{\boldsymbol{\phi}}_2)^\top$ is the (unrestricted) maximum likelihood estimators of $\boldsymbol{\phi}$ and $\tilde{\boldsymbol{\phi}} = (\tilde{\boldsymbol{\phi}}_1, \tilde{\boldsymbol{\phi}}_2)^\top$ denote the (restricted) maximum likelihood estimators of $\boldsymbol{\phi}$ under H_0 hypothesis. Asymptotically, S_{LR} and S_G have a central chi-square distribution with q degrees of freedom under H_0, and general conditions of regularity. The null hypothesis is rejected for a fixed nominal level α, if the test statistic exceeds the upper $100(1 - \alpha)\%$ quantile of the chi-square distribution.

3.1 Tests for Weibull Promotion Time Model

When each random variable M_i follows a Poisson distribution with parameter θ, the unified model comes down to promotion time model (Yakovlev et al., 1993; Chen et al., 1999). From (1) we have

$$S_p(t) = \exp\{-\theta[1 - S(t|\boldsymbol{\lambda})]\},$$

thus the cure fraction induced by this model is $p_\theta(0) = \exp(-\theta)$, and the probability density function is

$$f_p(t) = \theta f(t) \exp(-\theta F(t|\boldsymbol{\lambda})).$$

At the promotion time Weibull model, it is assumed that failure times of susceptible individuals follow a Weibull distribution. Here we use the parametrization for Weibull given in Fonseca et al. (2011), where the probability density function and survival are given by

$$f(t) = \rho t^{\rho-1} \exp(\gamma - t^\rho e^\gamma) \quad \text{and} \quad S(t) = \exp(-t^\rho e^\gamma),$$

where $\rho > 0$ and $\gamma \in \Re$. Thus the functions $S_p(t)$ and $f_p(t)$ are given by

$$S_p(t) = \exp\{-\theta[1 - \exp(-t^\rho e^\gamma)]\} \tag{6}$$

and

$$f_p(t) = \theta\rho t^{\rho-1} \exp(\gamma - t^\rho e^\gamma) \exp\{-\theta[1 - \exp(-t^\rho e^\gamma)]\}. \tag{7}$$

Now, consider the existence of heterogeneity in the population so that each random variable M_i follows a Poisson distribution with parameter θ_i. The relation often used between the parameter θ_i and the covariates in the promotion time model is given by $\theta_i = \exp\left(\boldsymbol{x}_i^\top \boldsymbol{\beta}\right)$, where $\boldsymbol{\beta}$ and $\boldsymbol{x}_i$ are defined as before. In this case the cure fraction is related to the covariates through the expression

$$p(0|\boldsymbol{\beta}, \boldsymbol{x}_i) = \exp\left[-\exp\left(\boldsymbol{x}_i^\top \boldsymbol{\beta}\right)\right]. \tag{8}$$

Considering a sample of n individuals and denoting by $\boldsymbol{\phi} = (\boldsymbol{\beta}^\intercal, \boldsymbol{\lambda}^\intercal)^\top$, the parameter vector where $\boldsymbol{\lambda}$ is the vector of parameter of Weibull model. The logarithm of the marginal likelihood function for this model is obtained substituting (6) and (7) in (4) and including covariates through the relation given above, then we have

$$\begin{aligned}\ell^*(\boldsymbol{\phi}; \mathcal{D}) = &\sum_{i=1}^{n} \delta_i \left[\boldsymbol{x}_i^\top \boldsymbol{\beta} + \gamma + \log\left(\rho y_i^{\rho-1}\right) - y_i^\rho e^\gamma\right] \\ &- \sum_{i=1}^{n} \exp\left(\boldsymbol{x}_i^\top \boldsymbol{\beta}\right)\left[1 - \exp\left(-y_i^\rho e^\gamma\right)\right].\end{aligned} \tag{9}$$

For computational reasons, we consider in (9) a reparametrization $\rho = e^{\rho^*}$, obtaining $\rho^* \in \Re$. Denoting $\boldsymbol{\lambda}^\intercal = (\gamma, \rho^*)$ and through the derivative with respect to the parameter vector $\boldsymbol{\phi}$, we get the score vector, which can be written as

$$U(\boldsymbol{\phi}) = \sum_{i=1}^{n} X_i s_i(\boldsymbol{\phi}), \tag{10}$$

where X_i is the matrix $\begin{pmatrix} \boldsymbol{x}_i & \mathbf{0} & \mathbf{0} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}_{(p+2)\times 3}$ and $s_i(\boldsymbol{\phi})$ is the vector $(s_{i1}(\boldsymbol{\phi}), s_{i2}(\boldsymbol{\phi}), s_{i3}(\boldsymbol{\phi}))^\top$ with

$$\begin{aligned} s_{i1}(\boldsymbol{\phi}) &= \delta_i - \theta_i\left[1 - \exp\left(-y_i^{e^{\rho^*}} e^\gamma\right)\right], \\ s_{i2}(\boldsymbol{\phi}) &= \delta_i\left[1 + e^{\rho^*}\log(y_i)\left(1 - y_i^{e^{\rho^*}} e^\gamma\right)\right] - \theta_i \exp\left(\rho^* + \gamma - y_i^{e^{\rho^*}} e^\gamma\right) y_i^{e^{\rho^*}} \log(y_i) \\ \text{and } s_{i3}(\boldsymbol{\phi}) &= \delta_i\left(1 - y_i^{e^{\rho^*}} e^\gamma\right) - \theta_i \exp\left(\gamma - y_i^{e^{\rho^*}} e^\gamma\right) y_i^{e^{\rho^*}}. \end{aligned}$$

Now consider we want to test only a partition $\boldsymbol{\phi}_1$ with dimension q of the vector $\boldsymbol{\phi}$. The likelihood ratio and gradient statistics for testing the composite hypothesis, as in (5) and considering $\boldsymbol{\phi}_{10} = 0$, are given by

$$S_{RV} = 2\left[\ell(\widehat{\boldsymbol{\phi}}) - \ell(\tilde{\boldsymbol{\phi}})\right] \quad \text{and} \quad S_G = U_{\boldsymbol{\phi}_1}^\top(\tilde{\boldsymbol{\phi}})\widehat{\boldsymbol{\phi}}_1,$$

where $\widehat{\boldsymbol{\phi}} = \left(\widehat{\boldsymbol{\phi}}_1, \widehat{\boldsymbol{\phi}}_2\right)^\top$ and $\tilde{\boldsymbol{\phi}} = \left(\mathbf{0}, \tilde{\boldsymbol{\phi}}_2\right)^\top$ are, respectively, the unrestricted and the restricted maximum likelihood estimator of $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)^\top$, under the null hypothesis.

4. Simulation Study

A simulation study was conducted to investigate the finite sample performance of the likelihood ratio and gradient tests to test parameters for a survival model with cure fraction. The simulation results are based on R *software* (R Development Core Team, 2010) which use the routine *optim* to maximize the likelihood function through the optimization algorithms BFGS (Broyden, Fletcher, Goldfarb and Shanno). The model considered was the Weibull promotion time. Relating the reparametrization used in (10) for $\rho = e^{\rho^*}$, with the default in the software R for parameters of Weibull distribution we found $\rho^* = \log(a)$ and $\gamma = -a\log(b)$, where $a > 0$ and $b > 0$ are shape and scale parameters, respectively.

To assess the effect of number of nuisance parameters in performance of tests, we consider cases with three, four and five covariates ($p = 3, p = 4$ and $p = 5$), generated from Bernoulli distributions. That is, for $l = 1, \ldots, p$, each covariate x_l is drawn from a Bernoulli (ν_l) where the probabilities of success $\nu_1, \ldots, \nu_5$ are set as $0.49, 0.50, 0.51, 0.52$ and 0.53, respectively. We consider samples of size $n = 30$ and 100. For each individual, values for M were generated as a random sample of Poisson distribution with mean $\theta_i = \exp\left(\boldsymbol{x}_i^\top \boldsymbol{\beta}\right)$. The values used for the vector $\boldsymbol{\beta}$ were chosen so that, when combined with the covariates, the average of cure fractions $p_{\theta_i}(0) = \exp(-\theta_i)$, $i = 1, \ldots, n$, were around 10%, 20% or 30%. The values specified for the vector $\boldsymbol{\beta}$ are given in Table 1.

Table 1. Values of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ used in the simulations, for $p = 3$, $p = 4$ and $p = 5$, according to the cure fraction proportions ($\%c.f.$), the number of tested parameters (q) and the sample size (n).

n	q	$\%c.f.$	$(\beta_1, \beta_2, \beta_3)$	$(\beta_1, \beta_2, \beta_3, \beta_4)$	$(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$
30	1	10	(0, 0.01, 0.90)	(0, 0.20, 0.40, 0.69)	(0, 0.20, 0.30, 0.50, 0.40)
		20	(0, 0.30, 0.36)	(0, 0.14, 0.30, 0.20)	(0, 0.10, 0.19, 0.31, 0.10)
		30	(0, 0.10, 0.14)	(0, 0.07, 0.09, 0.10)	(0, 0.09, 0.065, 0.05, 0.09)
	2	10	(0, 0, 0.90)	(0, 0, 0.50, 0.69)	(0, 0, 0.40, 0.50, 0.40)
		20	(0, 0, 0.50)	(0, 0, 0.44, 0.10)	(0, 0, 0.25, 0.30, 0.10)
		30	(0, 0, 0.19)	(0, 0, 0.12, 0.11)	(0, 0, 0.08, 0.09, 0.10)
100	1	10	(0, 0.70, 0.75)	(0, 0.50, 0.60, 0.51)	(0, 0.50, 0.40, 0.30, 0.30)
		20	(0, 0.35, 0.40)	(0, 0.31, 0.30, 0.21)	(0, 0.191, 0.31, 0.20, 0.10)
		30	(0, 0.135, 0.15)	(0, 0.10, 0.11, 0.10)	(0, 0.07, 0.075, 0.081, 0.091)
	2	10	(0, 0, 1.55)	(0, 0, 0.85, 0.80)	(0, 0, 0.51, 0.55, 0.60)
		20	(0, 0, 0.68)	(0, 0, 0.40, 0.45)	(0, 0, 0.16, 0.20, 0.50)
		30	(0, 0, 0.249)	(0, 0, 0.11, 0.23)	(0, 0, 0.09, 0.145, 0.09)

For the i-th immune individual, ($M_i = m_i > 0$), random samples $Z_{ik} \sim Weibull(a, b)$ of size m_i were generated with parameters $a = 2$ and $b = 4$. Hence, the failure time are denoted by $t_i = min\{Z_{ik}, k = 1, \ldots, m_i\}$, $i = 1, \ldots, n$. Random censoring were generated from independent Uniform$(0, u)$ random variable, where the value of u affects inversely the proportion of censoring in the sample. In order to evaluate separately the effect of increasing the censoring proportion among cured and not cured on the performance of tests, we consider here the censoring proportion with respect to all units under risk, that is, susceptible to the event occurrence.

We also considered three nominal levels ($\alpha = 1\%, 5\%$ and 10%). The tests were performed and the values of the statistics were compared with the respective quantiles of the chi-square distribution, i.e., 6.635, 3.841 and 2.706 to test the null hypothesis $H_{00} : \beta_1 = 0$ ($q = 1$) or 9.210, 5.991 and 4.605 to test the hypothesis $H_{01} : \beta_1 = \beta_2 = 0$ ($q = 2$).

Under each combination of parameter configuration we ran 10,000 simulations, and calculated the proportion of times that the hypotheses H_{00} and H_{01} were rejected .

4.1 Results

In Tables 2 and 3 we have estimated null rejection rates of the likelihood ratio and the gradient tests based of the null hypothesis $H_{00} : \beta_1 = 0$ ($q = 1$) and $H_{01} : \beta_1 = \beta_2 = 0$ ($q = 2$) in each considered situation, for samples of size $n = 30$ and 100, respectively. It shows that for all considered cases, the null rejection rates of the tests exceed the corresponding nominal level. This is in agreement with liberal characteristic of the likelihood ratio test in small samples and shows the same trend of the gradient test.

For $n = 30$ the tests get worse (rejection rate gets away from nominal level) when p (number of parameters) increases, and when we increase the number of tested parameters ($q = 1$ to $q = 2$). This fact is accentuated especially when we have 30% of censoring. For $n = 100$, there are not significative change in the performance of tests with increasing of p, or q, even in the presence of censoring. In general, we note that with the increase of sample size the tests become better (the null rejection rates approach the nominal level), regardless of the existence of cure fraction.

When we compare the performance of the likelihood ratio and gradient tests, we noticed they have equivalent results in almost all cases. There are just slight differences when the sample is very small ($n = 30$). Specifically, the gradient statistic presents mild advantage in some cases with uncensored samples while the likelihood ratio statistic is a little better in cases with censorship where the dimension of the vector parameters tested is lower.

We now consider a brief simulation study to investigate the finite-sample power properties of the tests. To make power comparisons, we must ensure that the test has the same (correct) size under the null hypothesis. As we have seen in our simulations that the likelihood ratio and gradient tests have different sizes, we used 100,000 Monte Carlo simulated samples, drawn under the null hypothesis, to estimate the correct critical value of each test for the fixed nominal level. Here we considered the tests for $n = 30$, $p = 3$, $q = 1$, 30% of censoring and 30% cure fraction under null hypothesis. We computed the rejection rates under the alternative hypothesis $H_{01} : \beta_1 = w$, for values of w belonging to the set $[-3, 3]$. As a result (Figure 1) we see that no test seems uniformly more powerful than the other.

Table 2. Null rejection rates of the likelihood ratio (S_{LR}) and the gradient (S_G) tests based of the null hypothesis $H_{00} : \beta_1 = 0$ ($q = 1$) and $H_{01} : \beta_1 = \beta_2 = 0$ ($q = 2$) for $n = 30$, according to the number of parameters (p), number of tested parameters (q), censoring proportions ($\%cens$) and cure fraction proportions ($\%c.f.$).

case	$\begin{pmatrix} p \\ q \end{pmatrix}$	$\%cens.$	$\%c.f.$	$n = 30$					
				$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 10\%$	
				S_{LR}	S_G	S_{LR}	S_G	S_{LR}	S_G
1			10	1.85	1.84	7.37	7.20	13.80	13.45
2		0	20	1.67	1.71	6.78	6.77	12.43	12.35
3	$\begin{pmatrix} p = 3 \\ q = 1 \end{pmatrix}$		30	1.81	1.92	7.10	7.23	12.97	13.07
4			10	2.39	2.36	8.42	8.17	14.61	14.60
5		30	20	1.99	2.26	8.10	8.14	14.58	14.63
6			30	2.02	2.58	7.87	7.97	13.48	13.81
7			10	2.29	2.21	7.77	7.48	14.06	13.75
8		0	20	1.81	1.78	7.22	7.22	13.12	12.98
9	$\begin{pmatrix} p = 4 \\ q = 1 \end{pmatrix}$		30	1.68	1.82	7.30	7.44	13.70	13.97
10			10	2.85	2.71	9.02	8.87	15.64	15.35
11		30	20	2.49	2.51	8.56	8.76	14.95	14.91
12			30	2.42	2.80	8.80	9.12	15.11	15.25
13			10	2.51	2.46	9.20	8.99	15.83	15.54
14		0	20	2.53	2.55	8.94	8.88	14.98	15.02
15	$\begin{pmatrix} p = 5 \\ q = 1 \end{pmatrix}$		30	2.45	2.56	8.38	8.63	15.05	15.17
16			10	2.98	2.79	9.80	9.87	16.55	16.50
17		30	20	3.06	3.26	10.04	10.04	17.17	17.12
18			30	3.09	3.44	10.18	10.44	16.84	16.86
19			10	1.57	1.53	7.24	7.08	13.82	13.42
20		0	20	1.70	1.84	7.29	7.50	13.24	13.22
21	$\begin{pmatrix} p = 3 \\ q = 2 \end{pmatrix}$		30	1.94	2.06	7.41	7.67	13.46	13.58
22			10	2.03	2.12	8.27	8.36	14.72	14.83
23		30	20	2.21	2.34	7.75	8.03	14.10	14.24
24			30	1.90	2.46	7.58	7.91	14.34	14.89
25			10	2.66	2.56	8.95	8.86	16.07	15.80
26		0	20	2.16	2.24	8.46	8.50	14.75	14.74
27	$\begin{pmatrix} p = 4 \\ q = 2 \end{pmatrix}$		30	2.10	2.29	7.86	8.26	14.47	14.68
28			10	2.71	2.83	9.46	9.32	16.64	16.33
29		30	20	2.62	2.83	9.31	9.66	16.12	16.36
30			30	2.65	3.23	8.99	9.66	16.07	16.70
31			10	2.88	2.87	9.58	9.54	16.34	16.43
32		0	20	2.64	2.77	9.53	9.56	16.52	16.45
33	$\begin{pmatrix} p = 5 \\ q = 2 \end{pmatrix}$		30	2.54	2.77	9.40	9.63	16.28	16.38
34			10	3.43	3.51	11.21	11.21	18.35	18.23
35		30	20	3.83	4.24	11.41	11.88	18.37	18.77
36			30	3.36	4.07	11.18	11.99	18.32	19.11

Table 3. Null rejection rates of the likelihood ratio (S_{LR}) and the gradient (S_G) tests based of the null hypothesis $H_{00} : \beta_1 = 0$ ($q = 1$) and $H_{01} : \beta_1 = \beta_2 = 0$ ($q = 2$) for $n = 100$, according to the number of parameters (p), number of tested parameters (q), censoring proportions ($\%cens$) and cure fraction proportions ($\%c.f.$).

case	$\begin{pmatrix} p \\ q \end{pmatrix}$	%cens.	$\%c.f.$	$n = 100$					
				$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 10\%$	
				S_{LR}	S_G	S_{LR}	S_G	S_{LR}	S_G
1	$\begin{pmatrix} p = 3 \\ q = 1 \end{pmatrix}$	0	10	1.32	1.30	5.63	5.65	10.88	10.87
2			20	1.20	1.21	5.31	5.33	11.02	10.98
3			30	1.33	1.36	5.84	5.85	10.57	10.66
4		30	10	1.42	1.41	5.73	5.76	10.97	10.93
5			20	1.27	1.27	5.68	5.68	10.85	10.93
6			30	1.06	1.09	5.83	5.78	11.09	11.25
7	$\begin{pmatrix} p = 4 \\ q = 1 \end{pmatrix}$	0	10	1.16	1.17	5.81	5.73	10.99	11.05
8			20	1.19	1.17	5.74	5.80	10.92	10.95
9			30	1.24	1.28	5.59	5.59	10.72	10.76
10		30	10	1.43	1.45	6.20	6.23	11.86	11.77
11			20	1.29	1.29	5.90	5.85	11.34	11.36
12			30	1.19	1.28	5.70	5.69	10.88	10.85
13	$\begin{pmatrix} p = 5 \\ q = 1 \end{pmatrix}$	0	10	1.43	1.36	5.96	5.94	11.38	11.31
14			20	1.37	1.34	6.24	6.14	11.71	11.76
15			30	1.16	1.17	5.77	5.77	11.16	11.21
16		30	10	1.53	1.51	6.46	6.44	11.97	11.98
17			20	1.27	1.27	6.05	6.07	11.90	11.95
18			30	1.39	1.40	5.94	6.04	11.15	11.19
19	$\begin{pmatrix} p = 3 \\ q = 2 \end{pmatrix}$	0	10	1.15	1.14	5.60	5.52	10.93	10.83
20			20	1.30	1.28	5.67	5.63	10.83	10.85
21			30	1.11	1.14	5.23	5.23	10.57	10.67
22		30	10	1.34	1.31	6.48	6.37	12.00	11.97
23			20	1.14	1.14	5.67	5.79	11.12	11.20
24			30	1.38	1.42	5.67	5.81	10.83	10.98
25	$\begin{pmatrix} p = 4 \\ q = 2 \end{pmatrix}$	0	10	1.19	1.19	6.06	5.99	11.36	11.23
26			20	1.27	1.33	5.81	5.79	11.37	11.29
27			30	1.23	1.20	5.38	5.43	10.58	10.61
28		30	10	1.15	1.18	5.93	5.92	11.77	11.79
29			20	1.19	1.23	5.80	5.89	11.34	11.53
30			30	1.33	1.38	6.13	6.23	11.60	11.73
31	$\begin{pmatrix} p = 5 \\ q = 2 \end{pmatrix}$	0	10	1.41	1.37	6.01	5.92	11.51	11.45
32			20	1.18	1.18	6.38	6.30	11.50	11.57
33			30	1.56	1.56	6.22	6.35	11.81	11.88
34		30	10	1.46	1.53	6.58	6.63	12.32	12.39
35			20	1.41	1.43	6.41	6.43	12.38	12.47
36			30	1.42	1.51	6.60	6.60	11.28	11.33

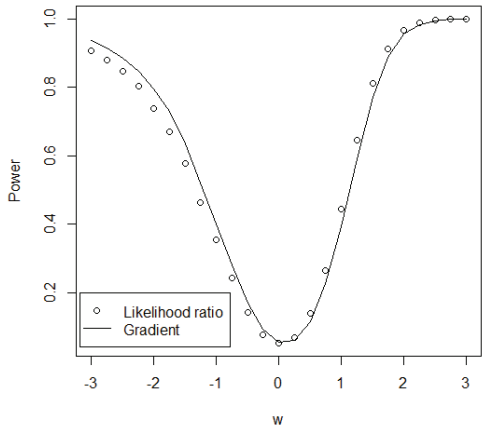


Figure 1. Power of likelihood and gradient tests for $n = 30$, $p = 3$, $q = 1$, 30% of censoring and 30% cure fraction under null hypothesis.

5. Application

To illustrate the model and the tests presented in this paper, we made an application in a data set about time of relapse free survival of 103 Brazilian children under 15 years, with acute lymphoblastic leukemia (ALL). These children were followed from 1988 to 1992 in some health institutions organized within a cooperative group for the treatment of acute leukemia in the state of Minas Gerais, Brazil.

These data are available in Colosimo and Giolo (2006). Viana et al. (1994) had described the study and analyzed the data using the Cox regression model (Cox, 1972), in order to evaluate the effect of factors on the hazard of recurrence in previously treated children.

At the end of the study, 39 children experienced the event and 64 censored (62% of censorship). Kaplan-Meier estimates of the survival function are shown in Figure 2, and show that in the last year of follow-up, the estimated survival curve apparently stabilizes at some positive value. Although follow-up seems insufficient to notice the occurrence of cured in study, there is a vast literature in the medical area about long-term survivors of pediatric leukemia (see for example Sala et al. (2004), Pui et al. (2003) and Neglia et al. (1991)).

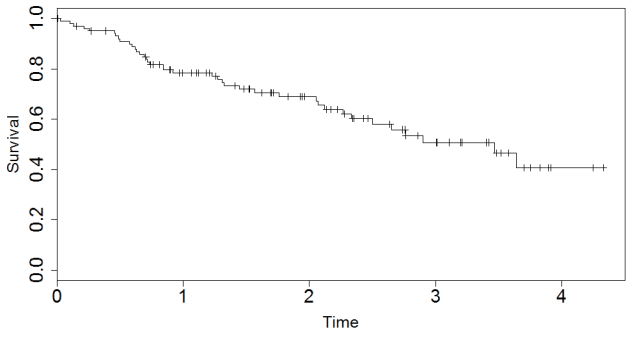


Figure 2. Kaplan-Meier estimates to of the data related to survival to acute lymphoblastic leukemia treatment.

To investigate the differences between subgroups in the data with respect to proportion of individuals who are long-term survivors, we consider a cure rate model associated with 5 factors of 2 levels: number of white cell count at diagnosis ($White = 1$ if this number is greater than $75,000$ by mm^3 and $White = 0$ otherwise); standardized age ($Age = 1$ if the index is greater than -2 and $Age = 0$ otherwise); standard weight for age and sex ($Weight = 1$ if the index is greater than -2 and $Weight = 0$ otherwise); positive periodic acid Schiff (PAS) reaction in the lymphoblasts ($Pas = 1$ for more than

5% of Pas positive marrow lymphoblasts and $Pas = 0$ otherwise) and cytoplasmic vacuolation ($Vac = 1$ if more than 10% of vacuolated blasts were present, and $Vac = 0$ otherwise).

We consider that the lifetimes for susceptible individuals Z_{il}, follow a $Weibull(\rho, \gamma)$ distributions, $i = 1, \ldots, 103$ and $l = 1, \ldots, M_i$.

Table 4. Estimates and tests for data on pediatric leukemia

Effect	Estimate(Se)	S_{LR}	p-value	S_G	p-value	df
$White$	1.211 (0.390)	8.676	0.003	8.676	0.004	1
Age	0.735 (0.371)	3.626	0.057	3.459	0.063	1
$Weight$	-0.767(0.341)	4.136	0.042	3.496	0.062	1
Pas	-1.062 (0.460)	6.491	0.011	7.073	0.008	1
Vac	1.391 (0.420)	9.072	0.003	8.087	0.004	1
ρ	0.343 (0.137)	-	-	-	-	-
γ	-1.256 (0.312)	-	-	-	-	-

We fit this model and apply the likelihood ratio and gradient tests to evaluate the effect of each factor in the cure rate. The results are shown in Table 4. At the 5% significance level, we note that for both tests the factors $White$, Pas and Vac are significant to explain the cure fraction. For the Age factor, the p-values obtained for the two tests are slightly greater than 5% so this is not significant at this level. However for the $Weight$ factor there is a divergence between the tests: the likelihood ratio test indicates that the $Weight$ factor is significant (p-value = 0.0419) while the gradient test concludes that it is not significant (p-value = 0.0616). One might argue that, according to the simulation results (see the most similar case in Table 3, line 18) the calculated p-values will always underestimate the *true p-values* (which would be calculated from the exact statistics distributions). This behavior can lead to undue rejection of hypotheses. Thus, based on this argument we do not reject the null hypothesis and we assume that the $Weight$ factor is not significant in this model.

Thus, the results of fit for the final model are given in Table 5.

Table 5. Estimates for final model - pediatric leukemia data.

Effect	Estimate	Se	S_{LR}	p-value	S_G	p-value	df
$White$	1.063	0.346	8.088	0.004	7.353	0.007	1
Vac	1.219	0.403	7.542	0.006	6.746	0.009	1
Pas	-1.052	0.455	6.583	0.010	7.196	0.007	1
ρ	0.234	0.135	-	-	-	-	-
γ	-1.691	0.295	-	-	-	-	-

From (8), we can calculate the cure rate for each combination of factors with the expression (11). The results are given in Table 6.

$$exp[-exp(1.063White_i + 1.219Vac_i - 1.052Pas_i)]. \tag{11}$$

Table 6. Cure rate for pediatric leukemia data.

White	Vac	Pas	cure rate (%)
< 75000	< 15%	< 5%	36,788
		≥ 5%	70,525
	≥ 15%	< 5%	3,389
		≥ 5%	30,671
≥ 75000	< 15%	< 5%	5,533
		≥ 5%	36,395
	≥ 15%	< 5%	< 0, 001
		≥ 5%	3,269

Thus in this study the group with the highest estimated cure rate (70.5%) is formed by children with lower white cell count at diagnosis ($White = 0$), with negative cytoplasmic vacuolation ($Vac = 0$) and PAS positive reaction ($Pas = 1$).

Note that although Viana et al. (1994) found that malnutrition (measured through the $Weight$) is the most significant adverse factor affecting time to remission, here we find that it is not significant to explain cure rate. In fact, according to Sala et al. (2004) there is no consensus about the relationship between poor nutritional status and the poor prospect for survival. Besides, covariates do not need exert the same effects on the cure fraction and the time to remission for susceptible individuals.

6. Concluding Remarks

In this work we compare via simulation, the perfomance of likelihood ratio and gradient tests to test regression coefficients related with cure fraction in Weibull promotion time model. We note that null rejection rates of the tests exceed the corresponding nominal level for small and moderate samples. This well-known liberal tendency of the likelihood ratio test, was also observed to the gradient test, which showed similar size distortions. Additionally, we note that this size distortion increases with the presence of censorship and with the increases of number of tested parameter as well as with the number of the nuisance parameters. This oversized behavior of the tests indicates that the true distributions of the likelihood ratio and gradient statistic have heavier right tail than the chi-square in small and moderate-sized samples. In applications this can lead to undue rejection of hypotheses since the calculated p-values (based in chi-square approximation) will in general underestimate the *true p-values*. The power simulation study suggest that no test seems uniformly most powerful than other when we use estimated correct critical values. Overall, we understand that the gradient statistic is equivalent to the likelihood ratio one, to test coefficients of this model.

Although the Wald and score tests shares the same asymptotic properties with the likelihood ratio and Gradient tests, they were not included in our simulation study because they require the computation of the Fisher information matrix, which cannot be obtained for the cure fractions models considered here. One could argue that the Fisher information should be replaced by the observed information matrix. We noticed, however, that in small and moderate-sized samples the observed information produced negative standard errors for a non-negligible proportion of the simulated censored samples. This is a problem to be investigated in a future study.

Due to the size distortions of the tests in small samples, an important subject of study is to obtain inferential improvements, like the Bartlett correction (Bartlett , 1937) or the Skovgaard's adjustment (Skovgaard , 1996). However the presence of censorship and cure fraction in cure rate models can make cumbersome or impossible the analytic derivation of corrections. Thus, another topic for future research will be investigate the use of a bootstrap Bartlett adjustment for the log-likelihood ratio statistic (Rocke , 1989) and bootstrap adjustment for gradient statistic. Furthermore we wish to study associated tests to models with cure rate in the presence of covariates associated with the lifetime of susceptible individuals.

7. Appendix

Complete and Marginal Likelihood

Here we present details to obtain the likelihood for the complete data (2) and the marginal likelihood function given in (3). We consider the same notations used in Section 2.1 but, for simplicity, we get $\boldsymbol{x}_i = 1$ (no covariates and $\theta_i = \theta$). Besides, for a single individual i, we denote the complete data by $\mathcal{D}_{c_i} = (y_i, \delta_i, m_i)$ and the data without the latent variables by $\mathcal{D}_i = (y_i, \delta_i)$.

The likelihood for the complete data $\mathcal{D}_c$ can be represented as follows

$$L(\boldsymbol{\phi}; \mathcal{D}_c) = \prod_{i=1}^{n} L(\boldsymbol{\phi}; \mathcal{D}_{c_i}) = \prod_{i=1}^{n} L(\boldsymbol{\lambda}; \mathcal{D}_i | M_i = m_i) p_\theta(m_i).$$

Based on classical results in survival analysis, it can be shown that the conditional likelihood function of (Y_i, δ_i) given the latent variable M_i, for a single individual i, is given by

$$L(\boldsymbol{\lambda}; \mathcal{D}_i | M_i = m_i) = \left[f_p(t; \boldsymbol{\lambda}|m_i)\right]^{\delta_i} \left[S_p(t; \boldsymbol{\lambda}|m_i)\right]^{1-\delta_i}. \quad (12)$$

Now, the conditional survival and density functions of T_i given the latent variable M_i can be obtained, respectively, as

follows

$$\begin{aligned} S_p(y_i|\boldsymbol{\lambda}, m) &= P(T_i > y_i|M_i = m_i) = P(\min\{Z_{i0}, Z_{i1}, \ldots, Z_{im_i}\} > y_i) \\ &= P(Z_{i0} > y_i, Z_{i1} > y_i, \ldots, Z_{im_i} > y_i) \\ &= \underbrace{P(Z_{i0} > y_i)}_{1} P(Z_{i1} > y_i) \ldots P(Z_{im_i} > y_i) = [S(y_i; \boldsymbol{\lambda})]^{m_i}, \end{aligned}$$

and

$$f_p(y_i|\boldsymbol{\lambda}, m_i) = -\frac{d}{dt} S(y_i; \boldsymbol{\lambda}) = m_i [S(y_i; \boldsymbol{\lambda})]^{m_i - 1} f(y_i; \boldsymbol{\lambda}).$$

Using the above results in (12), we get

$$\begin{aligned} L(\boldsymbol{\lambda}; \mathcal{D}_i|M_i = m_i) &= \left[m_i (S(y_i; \boldsymbol{\lambda}))^{m_i - 1} f(y_i; \boldsymbol{\lambda})\right]^{\delta_i} \left[(S(y_i; \boldsymbol{\lambda}))^{m_i}\right]^{1-\delta_i} \\ &= \left[m_i f(y_i; \boldsymbol{\lambda})\right]^{\delta_i} \left[S(y_i; \boldsymbol{\lambda})\right]^{m_i - \delta_i}. \end{aligned} \tag{13}$$

Thus, the likelihood for the complete data is

$$L(\boldsymbol{\phi}; \mathcal{D}_c) = \prod_{i=1}^{n} \left[m_i f(y_i|\boldsymbol{\lambda})\right]^{\delta_i} \left[S(y_i|\boldsymbol{\lambda})\right]^{m_i - \delta_i} p_\theta(m_i). \tag{14}$$

The likelihood with respect to the marginal distribution of the (Y_i, δ_i), denoted by L^*, can be obtained by summing overall possible values for the variables M_i, that is

$$L^*(\boldsymbol{\phi}; \mathcal{D}_i) = \sum_{m_i=0}^{\infty} L(\boldsymbol{\phi}; \mathcal{D}_{c_i}) = \sum_{m_i=0}^{\infty} \left[m_i f(y_i; \boldsymbol{\lambda})\right]^{\delta_i} \left[S(y_i; \boldsymbol{\lambda})\right]^{m_i - \delta_i} p_\theta(m_i). \tag{15}$$

Considering separately the cases $\delta_i = 0$ and $\delta_i = 1$, we have

$$\begin{aligned} L^*(\boldsymbol{\phi}; \mathcal{D}_i) &= \begin{cases} \sum_{m_i=0}^{\infty} [S(y_i; \boldsymbol{\lambda})]^{m_i} p_\theta(m_i) & \text{if } \delta_i = 0 \\ \sum_{m_i=0}^{\infty} m_i f(y_i; \boldsymbol{\lambda}) [S(y_i; \boldsymbol{\lambda})]^{m_i - 1} & \text{if } \delta_i = 1 \end{cases} \\ &= \begin{cases} S_p(y_i; \boldsymbol{\lambda}, \theta) & \text{if } \delta_i = 0 \\ f_p(y_i; \boldsymbol{\lambda}, \theta) & \text{if } \delta_i = 1 \end{cases} \\ &= \left[f_p(y_i; \boldsymbol{\lambda}, \theta)\right]^{\delta_i} \left[S_p(y_i; \boldsymbol{\lambda}, \theta)\right]^{1-\delta_i}. \end{aligned} \tag{16}$$

Thus, the total marginal likelihood is given by

$$L^*(\boldsymbol{\phi}; \mathcal{D}) = \prod_{i=1}^{n} L^*(\boldsymbol{\phi}; \mathcal{D}_i) = \prod_{i=1}^{n} \left[f_p(y_i; \boldsymbol{\lambda}, \theta)\right]^{\delta_i} \left[S_p(y_i; \boldsymbol{\lambda}, \theta)\right]^{1-\delta_i}. \tag{17}$$

Acknowledgements

This research was supported in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We also thank Silvia L. P. Ferrari for their constructive comments and an anonymous referee by important suggestions.

References

Alexandre, C., Cordeiro, G.M., Ortega, E.M.M and Sarabia, J.M. (2012). Generalized beta-generated distributions. *Computational Statistics & Data Analysis*, **56**, 1880–1897. http://dx.doi.org/10.1016/j.csda.2011.11.015

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Royal Society of London Proceedings Series A*, **160**(901), 268?82. http://dx.doi.org/10.1098/rspa.1937.0109

Berkson, J. and Gage, R.P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, 501–515. http://dx.doi.org/10.1080/01621459.1952.10501187

Boag, J.W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society*, Series B, **11**(1), 15–53. Retrieved from http://www.jstor.org/stable/2983694

Chen, M.H., Ibrahim, J.G. and Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, **94**, 909–919. http://dx.doi.org/10.1080/01621459.1999.10474196

Colosimo, E.A. and Giolo, S.R. (2006). *Análise de Sobrevivência Aplicada*. Edgard Blücher, Sao Paulo, SP.

Cox, D.R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society*, Series B (methodological), 34(2), 187?20. Retrieved from http://www.jstor.org/stable/2985181

Ferrari, S.L.P. and Pinheiro, E.C. (2014). Small-sample likelihood inference in extreme-value regression models. *Journal of Statistical Computation and Simulation*, **84**, 582–595. http://dx.doi.org/10.1080/00949655.2012.720686

Fonseca, S.F., Valença, D.M. and Bolfarine, H. (2011). Cure rate survival models with missing covariates: a simulation study. *Journal of Statistical Computation and Simulation*, **83**, 97–113. http://dx.doi.org/10.1080/00949655.2011.613396

Hayakawa, T. and Puri, M.L. (1985). Asymptotic expansions of the distributions of some test statistics. *Annals of the Institute of Statistical Mathematics*, **37**, 95–108. http://dx.doi.org/10.1007/BF02481083

Lemonte, A.J. (2016). *The Gradient Test: Another Likelihood-Based Test*. London: Academic Press. 1st edition.

Lemonte, A.J. and Ferrari, S.L.P. (2011a). Size and power properties of some tests in the Birnbaum-Saunders regression model. *Computational Statistics & Data Analysis*, **55**, 1109–1117. http://dx.doi.org/10.1016/j.csda.2010.09.008

Lemonte, A.J. and Ferrari, S.L.P. (2011b). Testing hypotheses in the Birnbaum-Saunders distribution under type-II censored samples. *Computational Statistics & Data Analysis*, **55**, 2388–2399. http://dx.doi.org/10.1016/j.csda.2011.02.005

Lemonte, A.J. and Ferrari, S.L.P. (2012). Local power and size proprities ofthe LR, Wald, score and gradient testes in dispersion models. *Statistical Methodology*, **9**, 537–554. http://dx.doi.org/10.1016/j.stamet.2012.03.001

Lemonte, A.J. and Ferrari, S.L.P. (2012). The local power of the gradient test. *Annals of the Institute of Statistical Mathematics*, **64**, 373–381. http://dx.doi.org/10.1007/s10463-010-0315-4

Maller, R.A. and Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*. Wiley, New York.

Medeiros, F.M.C, Silva-Júnior, A.H.M, Valença, D.M. and Ferrari, S.L.P (2014). Testing inference in accelerated failure time models. *International Journal of Statistics and Probability*, **3(2)**, 121. http://dx.doi.org/10.5539/ijsp.v3n2p121

Neglia, Joseph P and Meadows, Anna T and Robison, Leslie L and Kim, Tae H and Newton, William A and Ruymann, Frederick B and Sather, Harland N and Hammond, G Denman (1991). Second neoplasms after acute lymphoblastic leukemia in childhood. *New England Journal of Medicine*, **325(19)**, 1330-1336.

Pui, Ching-Hon and Cheng, Cheng and Leung, Wing and Rai, Shesh N and Rivera, Gaston K and Sandlund, John T and Ribeiro, Raul C and Relling, Mary V and Kun, Larry E and Evans, William E and others (2003). Extended follow-up of long-term survivors of childhood acute lymphoblastic leukemia. *New England Journal of Medicine*, **349(7)**, 640-649. http://dx.doi.org/10.1056/NEJMoa035091

R Development Core Team (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rao, C.R. (1948). Large-sample tests of statistical hypoteses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philophical Society*, **44**, 50–57. http://dx.doi.org/10.1017/S0305004100023987

Rocke, D. M. (1989). Bootstrap Bartlett adjustment in seemingly unrelated regression. *Journal of the American Statistical Association*, **84**(406), 598?01. http://dx.doi.org/10.1080/01621459.1989.10478809

Rodrigues, J. Balakrishnan, N. Cordeiro, G. N. and De Castro, M. (2011). A unified view on lifetime distributions arising from selection mechanisms. *Computational Statistics and Data Analysis*, **55**, 3311–3319. http://dx.doi.org/10.1016/j.csda.2011.06.018

Rodrigues, J., Cancho, V,G., De Castro, M. and Louzada-Neto, F. (2009). On the unification of long-term survival models. *Statistics & Probability Letters,*, **79**, 753 – 759. http://dx.doi.org/10.1016/j.spl.2008.10.029

Sala, A., Pencharz, P. and Barr, D.R. (2004). Children, cancer, and nutrition a dynamic triangle in review. *Cancer*, **100(4)**, 677 – 687. http://dx.doi.org/10.1002/cncr.11833

Skovgaard, I. (1996). An explicit large-deviation approximation to one-parameter tests. *Bernoulli*, **2**(2), 145?65.

Sposto, R., Sather, N.H. and Baker, A.S. (1992). A comparison of tests of the difference in the proportion of patients who are cured. *Biometrics*, 87 – 99. http://dx.doi.org/10.2307/2532741

Terrell, G.R. (2002). The gradient statistic. *Computing Science and Statistics*, **34**, 206–215.

Tsodikov, A.D., Ibrahim, J.G. and Yakovlev, A.Y. (2003). Estimating Cure Rates From Survival Data. *Journal of the American Statistical Association*, **98(464):**, 1063–1078. Retrieved from http://www.jstor.org/stable/30045351

Viana, M.B., Murao, M., Ramos, G., Oliveira, H.M., Carvalho, R.I., Bastos,M., Colosimo, E.A. and Silvestrini, W.S. (1994). Malnutrition as a prognostic factor in lymphoblastic leukamia: a multivariate analysis. *Brazilian Journal of Mother and Child Health*, **71**, 304–310. http://dx.doi.org/10.1136/adc.71.4.304

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, **54**, 426–482. http://dx.doi.org/10.2307/1990256

Wils, S.S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, **9**, 60–62. Retrieved from http://www.jstor.org/stable/2957648

Yakovlev, A.Y., Asselain, B., Bardou, V.J., Fourquet, A., Hoang, T., Rochefediere, A. and Tsodikov, A.D. (1993). A simple stochastic model of tumor recurrence and its application to data on premenopausal breast cancer. *Biometrie et analyse de donnees spatio-temporelles*, **12**, 66–82.

Yin, G. and Ibrahim, J.G. (2005). Cure rate models: a unified approach. *The Canadian Journal of Statistics/La revue canadienne de statistique*, 559–570. http://dx.doi.org/10.1002/cjs.5550330407

International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Alternative Second-Order N-Point Spherical Response Surface Methodology Designs and Their Efficiencies

Mary Paschal Iwundu

Department of Mathematics and Statistics, University of Port Harcourt, Nigeria. E-mail: mary.iwundu@uniport.edu.ng; mary_iwundu@yahoo.com

Received: April 11, 2016 Accepted: May 4, 2016 Online Published: June 11, 2016

doi:10.5539/ijsp.v5n4p22 URL: http://dx.doi.org/10.5539/ijsp.v5n4p22

Abstract

The equiradial designs are studied as alternative second-order N-point spherical Response Surface Methodology designs in two variables, for design radius ρ = 1.0. These designs are seen comparable with the standard second-order response surface methodology designs, namely the Central Composite Designs. The D-efficiencies of the equiradial designs are evaluated with respect to the spherical Central Composite Designs. Furthermore, D-efficiencies of the equiradial designs are evaluated with respect to the D-optimal exact designs defined on the design regions of the Circumscribed Central Composite Design, the Inscribed Central Composite Design and the Face-centered Central Composite Design. The D-efficiency values reveal that the alternative second-order N-point spherical equiradial designs are better than the Inscribed Central Composite Design though inferior to the Circumscribed Central Composite Design with efficiency values less than 50% in all cases studied. Also, D-efficiency values reveal that the alternative second-order N-point spherical equiradial designs are better than the N-point D-optimal exact designs defined on the design region supported by the design points of the Inscribed Central Composite Design. However, the N-point spherical equiradial designs are inferior to the N-point D-optimal exact designs defined on the design region supported by the design points of the Circumscribed Central Composite Design and those of the Face-centered Central Composite Design, with worse cases with respect to the design region of the Circumscribed Central Composite Design.

Keyword: Equiradial designs, Second-Order Response Surface Methodology Designs, Central Composite Designs, D-efficiency

1. Introduction

Central Composite Designs (CCDs) play a vital role in modelling second-order response functions in the presence of curvature. They are particularly useful at the second phase of process optimization. However, some spherical designs exist and can serve reasonably well when the standard Central Composite Designs are unavailable and/or cannot be employed. One such class of spherical designs is the class of equiradial designs, which according to Myer *etal.* (2009) are some special and interesting two-factor designs for modelling second-order response functions. As the name implies, equiradial designs are designs on a common sphere and are rotatable. The class of equiradial designs begins with a pentagon of equally spaced points on the sphere with design matrix expressed as

$$\begin{matrix} x_1 & x_2 \\ \{\ \rho\, Cos\ (\theta + 2\pi u/n_1) & \rho\, Sin\ (\theta + 2\pi u/n_1)\};\ u = 0, 1, 2, \dots\ , n_1 - 1 \end{matrix}$$

where x_1 and x_2 represent the two controllable variables, ρ is the radius of the design and n_1 represents the number of points on the sphere. In addition to the n_1 radial points of the design, n_c center points shall be added to the design. As indicated in Myer *etal.* (2009), the value of θ is assumed equal to zero since θ has no effect on the information matrix, X^TX, of the design. Thus the equiradial designs are such that the information matrix is invariant to design rotation.

Many works have been done using second-order response surface models and designs. They include the construction of efficient and optimal experimental designs for second-order response surface models (see for example Onukogu and Iwundu (2007)). Concerns about optimality of designs have been investigated for second-order models (see for example Dette and Grigoriev (2014)). Optimal choices of design points have been addressed by a number of researchers including Chigbu and Nduka (2006) and Iwundu (2015). Lucas (1976) compared the performances of several types of second-order response surface designs in symmetric regions on the basis of D- and G-optimality criteria. Graphical methods have been employed in studying the response variance property of second-order response surface designs as

seen in Myer *etal.*(1992), Giovannitti-Jensen and Myers (1989), Zahran *etal* (2003). Chigbu *etal.* (2009) compared the prediction variances of some Central Composite Designs in spherical regions with radius $\alpha = \sqrt{k}$ where k is the number of model controllable variables. Their results showed that Central Composite Designs, Small Composite Designs and Minimum-run resolution (MinRes) V designs are not uniformly superior under G- and I-optimality criteria as well using Variance Dispersion graphs. Iwundu and Otaru (2014) considered imposing D-Optimality criterion on the design regions supported by points of the Central Composite Designs. For the second order polynomial model used, results showed that the D-optimal designs defined over the rotatable Circumscribed Central Composite Design region had better determinant values than those defined over the Face-centered Central Composite Design region and the Inscribed Central Composite Design region.

Ukaegbu and Chigbu (2015) considered the prediction capabilities of partially replicated rotatable Central Composite Designs. Their results showed that the replicated cube designs with higher replications are more efficient and have better prediction capabilities than the replicated star designs. Iwundu (2015) studied the optimal partially replicated cube, star and center runs on design region supported by points of the Face-centered Central Composite Design, using quadratic models. With variations involving replicating the cube points while the star points and center point are held fixed, replicating the star points while the cube points and the center point are held fixed and replicating the center point while the cube points and the star points are held fixed, results showed that for the quadratic models considered, the Face-centered Central Composite Design comprising of two cube portions, one star portion and a center point performed better than other variations under D- and G-optimality criteria. When compared with the traditional method of replicating only the center point, the variation involving two cube portions, one star portion and a center point was relatively better in terms of design efficiencies. Oyejola and Nwanya (2015) studied the performance of five varieties of Central Composite Design when the axial portions are replicated and the center point increased one and three times. An excellent review of literature on some earlier works involving Central Composite Designs in spherical regions have been documented by Chigbu *etal.*(2009).

Spherical designs are useful in constructing rotatable designs in the field of combinatorics. However, it is important to obtain designs that reflect other important properties. The notions of design optimality and efficiency are paramount in assessing the quality of experimental designs. In particular, the D-optimality and D-efficiency play major roles in design optimality. They have been most studied and are also available in most statistical software. Atkinson and Donev (1992) gave various properties of the D-optimality and D-efficiency of designs under varying design conditions. It is worth noting that second-order models serve importantly in process optimization and are very reliable low-order approximating polynomials to the true unknown response functions relating a response with several controllable variables which may be natural or coded. The second-order response surface model in two controllable variables, x_1 and x_2, is given as

$$y(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \epsilon \quad 1.1$$

and written in matrix notation as

$$Y = X\beta + \epsilon \quad 1.2$$

where

Y is the Nx1 vector of observed values.

X is the Nxp design matrix

β is the px1 vector of unknown model parameters which are estimated on the basis of N uncorrelated observations.

ϵ is the random additive error associated with Y and is independently and identically distributed with zero mean and constant variance.

2. Methodology

For the importance of second-order Response Surface Methodology designs, we consider in this work the equiradial designs which are alternative second-order spherical Response Surface Methodology designs in two variables. The interest here is in comparing the efficiency of the equiradial designs with respect to the standard spherical Central Composite Designs. In particular interest is in how the equiradial designs compare with the Circumscribed and Inscribed Central Composite designs, which are both spherical and rotatable. We shall further consider the efficiencies of equiradial designs with respect to the D-optimal exact designs of Iwundu and Otaru (2014) which were defined on the design regions supported by design points of the Circumscribed Central Composite Design, the Inscribed Central Composite Design and the Face-centered Central Composite Design. The efficiency of a design provides a measure of

the optimality of the design. In comparing two designs, the relative efficiency is seen as the ratio of their separate efficiencies.

We shall employ the D-efficiency criterion as the test criterion. This criterion has been extensively used as a single numerical measure of the efficiency of designs. The D-efficiency criterion aims at minimizing the variance-covariance matrix associated with the parameter estimates of the model used. By definition, the D-efficiency of a design $\xi^{(1)}$ is given as

$$D_{eff} = 100\, x \left(\det M\left(\xi^{(1)}\right)\right)^{\frac{1}{p}} \quad 2.1$$

and the D-efficiency of a design $\xi^{(1)}$ relative to the design $\xi^{(2)}$ is given as

$$D_{eff} = \left(\frac{\det M(\xi^{(1)})}{\det M(\xi^{(2)})}\right)^{\frac{1}{p}} \quad 2.2$$

where $M(.)$ is the information matrix of the design and p is the number of model parameters. For an N-point design, say ξ_N, the information matrix of the design ξ_N is X^TX and normalized as $\frac{X^TX}{N}$ to remove the effect of changing design sizes. The (Nxp) matrix, X, is the design matrix whose columns are built from the model and the design ξ_N and $(.)^T$ represents transpose. Among other things, D-efficiency values depend on the number of points in the design and the number of controllable variables in the model. In comparing designs, the best design is one with the largest D-efficiency value. In terms of relative efficiency, the ratio in equation 2.2 exceeds unity if the design $\xi^{(1)}$ is better than the design $\xi^{(2)}$.

In comparing the N-point equiradial designs with the Circumscribed Central Composite designs, the 9-point Circumscribed Central Composite design comprising of the factorial points { (1,1), (1,-1), (-1,1), (-1,-1)}, the axial points {(1.414,0), (-1.414,0), (0,1.414), (0,-1.414)} and the center point {(0,0)} shall be employed. Similarly, in comparing the N-point equiradial designs with the Inscribed Central Composite Designs, the 9-point Inscribed Central Composite design comprising of the factorial points { (0.7,0.7), (0.7,-0.7), (-0.7,0.7), (-0.7,-0.7)}, the axial points {(1,0), (-1,0), (0,1), (0,-1)} and the center point {(0,0)} shall be employed. For comparisons with the D-optimal exact designs, the N-point designs generated by Iwundu and Otaru (2014) shall be employed correspondingly with the N-point equiradial designs.

3. Results

The design measures associated with the N-point equiradial designs for ρ = 1, n_1= 5, 6, ⋯ , 11 and n_c = 1 are as follows;

$$\xi_6 = \begin{pmatrix} 1 & 0 \\ 0.31 & 0.95 \\ -0.81 & 0.59 \\ -0.81 & -0.59 \\ 0.31 & -0.95 \\ 0 & 0 \end{pmatrix}$$

$$\xi_7 = \begin{pmatrix} 1 & 0 \\ 0.5 & 0.87 \\ -0.5 & 0.87 \\ -1 & 0 \\ -0.5 & -0.87 \\ 0.5 & -0.87 \\ 0 & 0 \end{pmatrix}$$

$$\xi_8 = \begin{pmatrix} 1 & 0 \\ 0.62 & 0.78 \\ -0.22 & 0.97 \\ -0.90 & 0.43 \\ -0.90 & -0.43 \\ -0.22 & -0.97 \\ 0.62 & -0.78 \\ 0 & 0 \end{pmatrix}$$

$$\xi_9 = \begin{pmatrix} 1 & 0 \\ 0.71 & 0.71 \\ 0 & 1 \\ -0.71 & 0.71 \\ -1 & 0 \\ -0.71 & -0.71 \\ 0 & -1 \\ 0.71 & -0.71 \\ 0 & 0 \end{pmatrix}$$

$$\xi_{10} = \begin{pmatrix} 1 & 0 \\ 0.77 & 0.64 \\ 0.17 & 0.98 \\ -0.5 & 0.87 \\ -0.94 & 0.34 \\ -0.94 & -0.34 \\ -0.5 & -0.87 \\ 0.17 & -0.98 \\ 0.77 & -0.64 \\ 0 & 0 \end{pmatrix}$$

$$\xi_{11} = \begin{pmatrix} 1 & 0 \\ 0.81 & 0.59 \\ 0.31 & 0.95 \\ -0.31 & 0.95 \\ -0.81 & 0.59 \\ -1 & 0 \\ -0.81 & -0.59 \\ -0.31 & -0.95 \\ 0.31 & -0.95 \\ 0.81 & -0.59 \\ 0 & 0 \end{pmatrix}$$

$$\xi_{12} = \begin{pmatrix} 1 & 0 \\ 0.84 & 0.54 \\ 0.41 & 0.91 \\ -0.14 & 0.99 \\ -0.65 & 0.76 \\ -0.96 & 0.28 \\ -0.96 & -0.28 \\ -0.65 & -0.76 \\ -0.14 & -0.99 \\ 0.41 & -0.91 \\ 0.84 & -0.54 \\ 0 & 0 \end{pmatrix}$$

For the bivariate quadratic model in equation 1.1, the normalized information matrices and the associated determinant values corresponding to the equiradial designs are, respectively, as follows;

$$M(\xi_6) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0.4174 & 0.4168 \\ 0 & 0.4174 & 0 & 0 & -0.0005 & -0.0007 \\ 0 & 0 & 0.4168 & -0.0007 & 0 & 0 \\ 0 & 0 & -0.0007 & 0.105 & 0 & 0 \\ 0.4174 & -0.0005 & 0 & 0 & 0.3132 & 0.105 \\ 0.4168 & -0.0007 & 0 & 0 & 0.105 & 0.3118 \end{bmatrix}$$

$\det M(\xi_6) = 2.639818966\text{x}10^{-4}$

$M(\xi_7)$ =

1	0	0	0	0.4285	0.4325
0	0.4285	0	0	0	0
0	0	0.4285	0	0	0
0	0	0	0.1081	0	0
0.4285	0	0	0	0.3214	0.1081
0.4325	0	0	0	0.1081	0.3273

det $M(\xi_7)$ = 2.67816305x10^{-4}

$M(\xi_8)$ =

1	0	0	0	0.4357	0.4357
0	0.4357	0	0	-0.0003	0.00095
0	0	0.4335	0.00095	0	0
0	0	0.00095	0.1072	0	0
0.4357	-0.0003	0	0	0.3265	0.1072
0.4335	0.00095	0	0	0.1072	0.3224

det $M(\xi_8)$ = 2.37557395x10^{-4}

$M(\xi_9)$ =

1	0	0	0	0.4462	0.4462
0	0.4462	0	0	0	0
0	0	0.4462	0	0	0
0	0	0	0.1129	0	0
0.4462	0	0	0	0.3351	0.1129
0.4462	0	0	0	0.1129	0.3351

det $M(\xi_9)$ = 2.4889799568x10^{-4}

$M(\xi_{10})$ =

1	0	0	0	0.451	0.4485
0	0.451	0	0	0.0011	-0.001
0	0	0.4485	-0.001	0	0
0	0	-0.001	0.1123	0	0
0.451	0.0011	0	0	0.3391	0.1123
0.4485	-0.001	0	0	0.1123	0.3352

det $M(\xi_{10})$ = 2.298611217x10^{-4}

$$M(\xi_{11}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0.4553 & 0.4547 \\ 0 & 0.4553 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.4547 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1145 & 0 & 0 \\ 0.4553 & 0 & 0 & 0 & 0.3417 & 0.1145 \\ 0.4547 & 0 & 0 & 0 & 0.1145 & 0.3402 \end{bmatrix}$$

det $M(\xi_{11})$ = 2.224863024x10^{-4}

$$M(\xi_{12}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0.4594 & 0.4761 \\ 0 & 0.4562 & 0 & 0 & -0.003 & -0.016 \\ 0 & 0 & 0.4593 & -0.0005 & 0.00089 & 0.0047 \\ 0 & 0 & -0.0005 & 0.1134 & -0.0008 & -0.004 \\ 0.4594 & -0.003 & 0.00089 & -0.0008 & 0.3484 & 0.1297 \\ 0.4761 & -0.016 & 0.0047 & -0.004 & 0.1297 & 0.3512 \end{bmatrix}$$

det $M(\xi_{12})$ = 2.149505806x10^{-4}.

These designs are compared with the Circumscribed Central Composite design and the Inscribed Central Composite design whose design points have been listed in Section 2. The respective normalized information matrices M_1 and M_2 together with the determinant values are as listed below, where M_1 represents the normalized information matrix associated with the Circumscribed Central Composite design and M_2 represents the normalized information matrix associated with the Inscribed Central Composite design.

$$M_1 = \begin{bmatrix} 0.9999 & 0 & 0 & 0 & 0.8887 & 0.8887 \\ 0 & 0.8887 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.8887 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.4444 & 0 & 0 \\ 0.8887 & 0 & 0 & 0 & 1.3327 & 0.4444 \\ 0.8887 & 0 & 0 & 0 & 0.4444 & 1.3327 \end{bmatrix}$$

det M_1 = 6.158433838x10^{-4}

$$M_2 = \begin{bmatrix} 0.9999 & 0 & 0 & 0 & 0.44 & 0.44 \\ 0 & 0.44 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.44 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1067 & 0 & 0 \\ 0.44 & 0 & 0 & 0 & 0.3289 & 0.1067 \\ 0.44 & 0 & 0 & 0 & 0.1067 & 0.3289 \end{bmatrix}$$

det M_2 = 2.224059802x10^{-4}.

The D-efficiency values of the N-point equiradial designs relative to the Circumscribed Central Composite design and the Inscribed Central Composite design are as in Table 1.

Table 1. D-Efficiency values of equiradial designs relative to the Circumscribed and the Inscribed Central Composite designs

Design size N	D-Efficiency values of equiradial designs relative to the Circumscribed Central Composite design	D-Efficiency values of equiradial designs relative to the Inscribed Central Composite design
6	0.4030426836	1.028974504
7	0.4040125476	1.031450583
8	0.3960196875	1.011044681
9	0.3991096601	1.018933432
10	0.3938518937	1.005510269
11	0.3917171302	1.001832841
12	0.3984739855	0.994333398

The equiradial designs are further compared with D-optimal exact designs whose design points are as in Iwundu and Otaru (2014). The D-efficiency values of the N-point equiradial designs relative to the N-point D-optimal exact designs defined on the design regions supported by design points of the Circumscribed Central Composite design, the Inscribed Central Composite design and the Face-centered Central Composite design are as in Tables 2-4. Each table comprises the design size N, the determinant values of the normalized information matrices associated with the equiradial designs, the determinant values of the normalized information matrices associated with the D-optimal exact designs as well as the D-efficiency values.

Table 2. D-Efficiency values of equiradial designs relative to the D-optimal exact designs defined on the design regions supported by points of the Circumscribed Central Composite Design

Design size N	Determinant value (equiradial design)	Determinant value (D-optimal exact design)	D-Efficiency value
6	$2.639818966x10^{-4}$	$3.1947x10^{-2}$	0.4496321364
7	$2.67816305x10^{-4}$	$3.837429233x10^{-2}$	0.4371523047
8	$2.37557395x10^{-4}$	$4.6828x10^{-2}$	0.4145184104
9	$2.488979568x10^{-4}$	$6.1584x10^{-2}$	0.3991096601
10	$2.298611217x10^{-4}$	$6.545687882x10^{-2}$	0.3898687014
11	$2.224863024x10^{-4}$	$6.004443063x10^{-2}$	0.3933734922
12	$2.149505806x10^{-4}$	$5.782736734x10^{-2}$	0.3935810684

Table 3. D-Efficiency values of equiradial designs relative to the D-optimal exact designs defined on the design regions supported by points of the Inscribed Central Composite Design

Design size N	Determinant value (equiradial design)	Determinant value (D-optimal exact design)	D-Efficiency value
6	$2.639818966x10^{-4}$	$1.166000031x10^{-4}$	1.145897918
7	$2.67816305x10^{-4}$	$1.384704002x10^{-4}$	1.116211976
8	$2.37557395x10^{-4}$	$2.224059802x10^{-4}$	1.018933432
9	$2.488979568x10^{-4}$	$1.713104121x10^{-4}$	1.056000521
10	$2.298611217x10^{-4}$	$2.362949253x10^{-4}$	0.9954096679
11	$2.224863024x10^{-4}$	$2.174265558x10^{-4}$	1.003841429
12	$2.149505806x10^{-4}$	$2.090927535x10^{-4}$	1.004615652

Table 4. D-Efficiency values of equiradial designs relative to the D-optimal exact designs defined on the design region supported by points of the Face-centered Central Composite Design

Design size N	Determinant value (equiradial design)	Determinant value (D-optimal exact design)	D-Efficiency value
6	$2.639818966x10^{-4}$	$5.486968437x10^{-3}$	0.6030781776
7	$2.67816305x10^{-4}$	$8.159865377x10^{-3}$	0.5658382953
8	$2.37557395x10^{-4}$	$8.7890625x10^{-3}$	0.5478197267
9	$2.488979568x10^{-4}$	$9.754610572x10^{-3}$	0.5425859777
10	$2.298611217x10^{-4}$	$9.360x10^{-3}$	0.5391359338
11	$2.224863024x10^{-4}$	$9.5374x10^{-3}$	0.5345383641
12	$2.149505806x10^{-4}$	$1.0154x10^{-2}$	0.5259570053

3. Discussion of Results

The equiradial designs have been examined as alternative spherical designs to the rotatable Central Composite Designs (CCDs) and the D-optimal exact designs in modelling second-order response functions. These designs are seen comparable with the standard second-order Response Surface Methodology designs. The equiradial designs which are simple to construct seem to show some appealing optimality properties. A careful look at the D-efficiency values makes it interesting to note that equiradial designs are not generally inferior designs. In fact, they appear more optimal than some frequently used second-order Response Surface Methodology designs. In particular, the study revealed that equiradial designs perform generally better than the Inscribed Central Composite designs and the D-optimal exact designs defined on the design region supported by the design points of the Inscribed Central Composite design for the design sizes considered. Besides N = 12, the equiradial designs were better than the Inscribed Central Composite design under the D-efficiency criterion. Additionally, each N-point equiradial design was better than the corresponding N-Point D-optimal exact design defined on the design region supported by the design points of the Inscribed Central Composite design, except for N=10 which gave relative efficiency value of 0.9954096679. However, it is clear from the relative efficiency value that the 10-point equiradial design is as good the 10-point D-optimal exact design.

It is further observed that the equiradial designs are not as credible as the Circumscribed Central Composite design in terms of D-efficiency. This was seen in the relative efficiency values being less than 50% in all cases considered. The observation is not different for N-Point D-optimal exact designs defined on the design region supported by the design points of the Circumscribed Central Composite design. However, when compared with the D-optimal exact designs defined on the design region supported by the design points of the Face-centered Central Composite design, the equiradial designs were not too inferior as the relative efficiency values exceeded 50% in all cases considered.

Although the equiradial designs could serve as alternatives to the standard Response Surface Methodology designs, they should be used with caution especially when design optimality is paramount.

References

Atkinson, A. C., & Donev, A. N. (1992). Optimum Experimental Designs, *Oxford: Oxford University Press*.

Chigbu, P. E., & Nduka, U.C. (2006). On The Optimal Choice Of The Cube And Star Replications In Restricted Second-Order Designs. United Nations Educational, Scientific and Cultural Organization and International Atomic Energy Agency: The Abdus Salam International Centre For Theoretical Physics, Trieste, Italy. Available at: http://www.ictp.it/~pub−off

Chigbu, P. E., Ukaegbu, E. C., & Nwanya, J. C. (2009). On comparing the prediction variances of some Central Composite Designs in Spherical regions: A Review. STATISTICA, anno LXIX, n, 4.

Dette, H., & Grigoriev, Y. (2014). Construction of efficient and optimal experimentl design for response surface models. *The annals of statistics, 42*(4), 1635-1656. http://dx.doi.org/10.1214/14-AOS1241

Giovannitti-Jensen, A., & Myers, R. H. (1989). Graphical Assessment of the Prediction Capability of Response Surface Designs, *Technometrics, 31*(2), 159-171.

Iwundu, M. P. (2015). Optimal Partially Replicated Cube, Star and Center Runs in Face-centered Central Composite Designs. *International Journal of Statistics and Probability, 4*(4). http://dx.doi.org/10.5539/ijsp.v4n4p1

Iwundu, M. P., & Otaru, O. A. P. (2014). Imposing D-optimality criterion on the design regions of the Central

Composite Designs (CCD). *Scientia Africana, 13*(1), 109-119.

Lucas, J. M. (1976). Response surface design is the best: A performance comparison of several types of quadratic response surface designs in symmetric regions. *Technometrics, 18*, 411-417.

Myers, R. H., Vinning, G. G., Giovannitti-Jensen, A. & Myers, S. L. (1992). Variance Dispersion properties of Second-order response surface designs. *Journal of Quality technology, 24*, 1-11.

Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M. (2009). Response Surface Methodology: Process and Product Optimization using designed experiments. 3rd Edition. John Wiley & Sons, Inc. New Jersey.

Onukogu, I. B., & Iwundu, M. P. (2007). A Combinatorial Procedure for Constructing D-Optimal Designs. *Statistica, 67*(4), 415-423.

Oyejola, B. A., & Nwanya, J. C. (2015). Selecting the right central composite design. *Journal of Statistics and Applications, 5*(1), 21-30.

Ukaegbu, E. C., & Chigbu, P. E. (2015). Graphical Evaluation of the Prediction Capabilities of Partially Replicated Orthogonal Central Composite Designs. *Quality and Reliability Engineering International, 31*, 707 – 717.

Zahran, A., Anderson-Cook, C. M., & Myers, R. H. (2003). Fraction of Design Space to Assess the Prediction Capability of Response Surface Designs. *Journal of Quality Technology, 35*, 377-386.

International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Collapsed Double Symmetry Model and Its Decomposition for Square Contingency Tables

Kouji Yamamoto[1], Yuya Matsuda[2] & Sadao Tomizawa[2]

[1] Department of Clinical Epidemiology and Biostatistics, Graduate School of Medicine, Osaka University

[2] Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science

Correspondence: Kouji Yamamoto, Department of Clinical Epidemiology and Biostatistics, Graduate School of Medicine, Osaka University, 2-2, Yamadaoka, Suita, Osaka, Japan. Tel: 81-6-6210-8373. E-mail: yamamoto-k@stat.med.osaka-u.ac.jp

Received: February 16, 2016 Accepted: March 3, 2016 Online Published: June 11, 2016

doi:10.5539/ijsp.v5n4p31 URL: http://dx.doi.org/10.5539/ijsp.v5n4p31

Abstract

For a square contingency table with ordinal categories, there may be a case that one wants to analyze several collapsed tables obtained by combining some adjacent categories of the original table. This paper proposes some new models which indicate double symmetry, quasi double symmetry and marginal double symmetry for the collapsed square tables. It also gives a decomposition of the double symmetry model for collapsed tables. Two kinds of occupational mobility data are analyzed using new models.

Keywords: Collapsed table, decomposition, double symmetry, marginal double symmetry, quasi double symmetry

1. Introduction

Consider square contingency tables having same ordinal row and column classifications, for instance, as Tables 1a and 2a. For these data, a lot of observations tend to fall in (or near) the main diagonal cells. Thus for such data, in many cases, the independence between the row and column variables does not hold. So, we are interested in symmetry instead of independence. Bowker (1948) proposed the symmetry (S) model, which indicates that the probabilities are symmetric with respect to the main diagonal of the table. Wall and Lienert (1976) proposed the point-symmetry (PS) model, which indicates that the probabilities are symmetric with respect to the center point in the table. Tomizawa (1985) proposed the double symmetry (DS) model such that both S and PS models hold. For other models of symmetry, see, for example, Caussinus (1965), Bishop, Fienberg and Holland (1975, Chap.8), Agresti (2013, Chap.11), and Tahata and Tomizawa (2014).

Table 1. Occupational status for Japanese father-daughter pairs; from Hashimoto (2003, p.144).

(a) Original table

	Daughter's class					
Father's class	(1)	(2)	(3)	(4)	(5)	Total
(1)	10	5	38	14	1	68
(2)	2	22	74	31	6	135
(3)	5	8	112	20	9	154
(4)	5	11	94	44	11	165
(5)	9	15	147	26	70	267
Total	31	61	465	135	97	789

(1) Capitalist, (2) New middle, (3) Working, (4) Self-employed and (5) Farming.

(b) Collapsed table ($T^{(1,4)}$ table)

	Daughter's class			
Father's class	High	Middle	Low	Total
High	10	57	1	68
Middle	12	416	26	454
Low	9	188	70	267
Total	31	661	97	789

(c) Collapsed table ($T^{(2,3)}$ table)

	Daughter's class			
Father's class	High	Middle	Low	Total
High	39	112	52	203
Middle	13	112	29	154
Low	40	241	151	432
Total	92	465	232	789

Table 2. The social status of occupations for the husband's father and the wife's father; from Katz (1978). (The upper and lower parenthesized values are the MLEs of expected frequencies under the CoDS and CoMDS models, respectively.)

(a) Original table

	Husband's father's status				
Wife's father's status	(1)	(2)	(3)	(4)	Total
(1)	44	17	4	12	77
	(38.00)	(22.87)	(5.38)	(12.50)	(78.75)
	(38.47)	(20.15)	(4.74)	(15.65)	(79.01)
(2)	10	3	6	2	21
	(7.24)	(3.00)	(6.00)	(2.35)	(18.59)
	(7.69)	(3.00)	(6.00)	(2.17)	(18.86)
(3)	29	7	22	22	80
	(21.01)	(7.00)	(22.00)	(25.90)	(75.91)
	(22.30)	(7.00)	(22.00)	(23.82)	(75.12)
(4)	13	8	21	32	74
	(12.50)	(7.79)	(20.46)	(38.00)	(78.75)
	(10.55)	(8.58)	(22.51)	(37.37)	(79.01)
Total	96	35	53	68	252
	(78.75)	(40.66)	(53.84)	(78.75)	(252.00)
	(79.01)	(38.73)	(55.25)	(79.01)	(252.00)

(1) Professional, technical, kindred white-collar managers, officials, (2) Clerical and sales white collar workers, (3) Craftsmen (blue collar workers), (4) Operatives, service workers, laborers excepting farm workers (blue collar workers)

(b) Collapsed table ($T^{(1,3)}$ table)

	Husband's father's status			
Wife's father's status	High	Middle	Low	Total
High	44	21	12	77
	(38.00)	(28.25)	(12.50)	(78.75)
	(38.47)	(24.89)	(15.65)	(79.01)
Middle	39	38	24	101
	(28.25)	(38.00)	(28.25)	(94.50)
	(29.99)	(38.00)	(25.99)	(93.98)
Low	13	29	32	74
	(12.50)	(28.25)	(38.00)	(78.75)
	(10.55)	(31.09)	(37.37)	(79.01)
Total	96	88	68	252
	(78.75)	(94.50)	(78.75)	(252.00)
	(79.01)	(93.98)	(79.01)	(252.00)

For analyzing the data in Tables 1a and 2a, in some cases, we would like to divide the occupational status into a simple categories, for example, "High", "Middle" and "Low". Table 1b is the collapsed table with "High" category which is

"(1) capitalist" in Table 1a, "Middle" category obtained by combining "(2) new middle", "(3) working" and "(4) self-employed" categories in Table 1a, and "Low" category being "(5) farming" in Table 1a. In other words, Table 1b is obtained by collapsing Table 1a into the 3×3 table by using cut points after the first and fourth rows and after the first and fourth columns. Yamamoto, Tahata and Tomizawa (2012) proposed some S models for collapsed square contingency tables, for example, the collapsed quasi-symmetry model. Yamamoto, Murakami and Tomizawa (2013) proposed some PS models for collapsed square contingency tables, for example, the collapsed quasi point-symmetry model.

Tomizawa (1985) gave the theorem that the DS model holds if and only if both the quasi double symmetry (QDS) and marginal double symmetry (MDS) models hold. See Appendix for the DS, QDS and MDS models.

In the present paper, Section 2 proposes some new models for the collapsed tables and gives the decomposition for the new model. These proposed models are totally different from existing models. Section 3 describes the goodness-of-fit test, and analyzes the data in Tables 1a and 2a using new models. Section 4 provides remarks.

2. New Models and Decomposition

Consider the $r \times r$ contingency table having ordered categories. Let p_{ij} denote the probability that an observation will fall in the (i, j)th cell of the $r \times r$ table $(i = 1, \ldots, r; j = 1, \ldots, r)$.

We consider the $[(r-1)/2]$ ways of collapsing an $r \times r$ table into a 3×3 table by using cut points after the hth and $h'(= r - h)$th rows and columns for $h = 1, ..., [(r-1)/2]$, where

$$\left[\frac{r-1}{2}\right] = \begin{cases} \dfrac{r-2}{2} & (\text{when } r \text{ is even}), \\ \dfrac{r-1}{2} & (\text{when } r \text{ is odd}). \end{cases}$$

We would like to refer to such a collapsed 3×3 table as the $T^{(h,h')}$ table $(h = 1, ..., [(r-1)/2])$. The $G_{kl}^{(h,h')}$ denotes the cumulative probability in the collapsed $T^{(h,h')}$ table $(k = 1, 2, 3; l = 1, 2, 3; h = 1, ..., [(r-1)/2])$; i.e.,

$$G_{11}^{(h,h')} = \sum_{i=1}^{h}\sum_{j=1}^{h} p_{ij}, \quad G_{12}^{(h,h')} = \sum_{i=1}^{h}\sum_{j=h+1}^{h'} p_{ij}, \quad G_{13}^{(h,h')} = \sum_{i=1}^{h}\sum_{j=h'+1}^{r} p_{ij},$$

$$G_{21}^{(h,h')} = \sum_{i=h+1}^{h'}\sum_{j=1}^{h} p_{ij}, \quad G_{22}^{(h,h')} = \sum_{i=h+1}^{h'}\sum_{j=h+1}^{h'} p_{ij}, \quad G_{23}^{(h,h')} = \sum_{i=h+1}^{h'}\sum_{j=h'+1}^{r} p_{ij},$$

$$G_{31}^{(h,h')} = \sum_{i=h'+1}^{r}\sum_{j=1}^{h} p_{ij}, \quad G_{32}^{(h,h')} = \sum_{i=h'+1}^{r}\sum_{j=h+1}^{h'} p_{ij}, \quad G_{33}^{(h,h')} = \sum_{i=h'+1}^{r}\sum_{j=h'+1}^{r} p_{ij}.$$

We propose the collapsed double symmetry (CoDS), collapsed quasi double symmetry (CoQDS) and collapsed marginal double symmetry (CoMDS) models.

First, we propose the CoDS model defined by

$$G_{kl}^{(h,h')} = G_{lk}^{(h,h')} = G_{k^*l^*}^{(h,h')} = G_{l^*k^*}^{(h,h')}$$

for $k = 1, 2, 3; l = 1, 2, 3$, and $h = 1, \ldots, [(r-1)/2]$, where $k^* = 4 - k$ and $l^* = 4 - l$. This model indicates both S and PS models hold in each collapsed $T^{(h,h')}$ table $(h = 1, ..., [(r-1)/2])$, namely, a structure of DS in each collapsed $T^{(h,h')}$ table. We point out that the CoDS model is totally different from the DS model, because the CoDS model indicates that there are DS structures in *all* collapsed tables. We note that the CoDS model holds if the DS model holds, but the converse does not necessarily hold.

Consider collapsing the r categories of an original table with ordered categories into 3 categories (say, groups A, B and C), by using cut points h and h'. The CoDS model describes that for each collapsed $T^{(h,h')}$ table $(h = 1, ..., [(r-1)/2])$, (i) the probability that both of the row and column values of an observation are in group A equals to the probability that both of the row and column values of it are in group C, (ii) the probability that the row and column values are in groups A and B, respectively, equals to the probability that the row and column values are in groups B and A, respectively, and it also equals to the probability that the row and column values are in groups C and B, respectively, and moreover it equals to the probability that the row and column values are in groups B and C, respectively, and (iii) the probability that the row and column values are in groups A and C, respectively, equals to the probability that the row and column values are in groups C and A, respectively.

Next, we propose the CoQDS model defined by

$$G_{kl}^{(h,h')} = \mu^{(h)}\alpha_k^{(h)}\beta_l^{(h)}\psi_{kl}^{(h)},$$

for $k = 1, 2, 3; l = 1, 2, 3$, and $h = 1, \ldots, [(r-1)/2]$, where

$$\psi_{kl}^{(h)} = \psi_{lk}^{(h)} = \psi_{k^*l^*}^{(h)} = \psi_{l^*k^*}^{(h)}.$$

Note that a special case of the CoQDS model obtained by putting $\{\alpha_k^{(h)} = \beta_k^{(h)}\}$ and $\{\alpha_k^{(h)} = \alpha_{k^*}^{(h)}\}$ is the CoDS model. Also noting that the CoQDS model is identical to the collapsed quasi point-symmetry model in Yamamoto et al. (2013) because the QDS model is identical to the quasi point-symmetry model only when the square contingency table is the 3×3 table. For the quasi point-symmetry model, see Yamamoto et al. (2013).

Finally, we propose the CoMDS model defined by

$$G_{k\cdot}^{(h,h')} = G_{k^*\cdot}^{(h,h')} = G_{\cdot k}^{(h,h')} = G_{\cdot k^*}^{(h,h')},$$

for $k = 1, 2, 3$, and $h = 1, \ldots, [(r-1)/2]$, where

$$G_{k\cdot}^{(h,h')} = \sum_{t=1}^{3} G_{kt}^{(h,h')}, \quad G_{\cdot k}^{(h,h')} = \sum_{s=1}^{3} G_{sk}^{(h,h')}.$$

This model states that there is a structure of MDS in each collapsed $T^{(h,h')}$ table ($h = 1, ..., [(r-1)/2]$). So, the CoMDS model is different from the MDS model. The MDS indicates that the row and column marginal distribution are symmetric and point-symmetric with respect to the midpoint of the row and column categories. Note that the CoMDS model holds if the MDS model holds, but the converse holds only when r is odd.

In a similar manner to the explanation of the CoDS model, the CoMDS model indicates that for each collapsed $T^{(h,h')}$ table ($h = 1, ..., [(r-1)/2]$), (i) the probability that the row value is in group A equals to the probability that the row value is in group C, and it is also equal to the probability that the column value is in group A, and moreover it is equal to the probability that the column value is in group C, and (ii) the probability that the row value is in group B equals to the probability that the column value is in group B; namely, for the original table, the probability that the row variable is h or below is equal to the probability that it is h' or above ($h = 1, ..., [(r-1)/2]$), and it is also equal to the probability that the column variable is h or below, and moreover it is equal to the probability that the column variable is h' or above ($h = 1, ..., [(r-1)/2]$).

Applying the decomposition theorem of the DS model in Tomizawa (1985) (described in Section 1) for collapsed 3×3 table, we can obtain the following theorem:

Theorem 1 *The CoDS model holds if and only if both the CoQDS and CoMDS models hold.*

3. Analysis of Data

3.1 Goodness-of-fit Test

Assume that a random sample of fixed size n is cross-classified according to the categorical variables. The distribution of the cell counts $\{n_{ij}\}$ is then the multinomial distribution specified by the sample size n and the population cell probabilities $\{p_{ij}\}$. We can obtain the maximum likelihood estimates (MLEs) of expected frequencies under the models by using the Newton-Raphson method in the log-likelihood equation. The likelihood-ratio approach to testing models leads to the test statistic

$$G^2 = 2\sum_{i=1}^{r}\sum_{j=1}^{r} n_{ij}\log\left(\frac{n_{ij}}{\widehat{m}_{ij}}\right),$$

where $\widehat{m}_{ij}$ is the MLE of expected frequency m_{ij} under the model. The number of degrees of freedom for the CoDS model is $5(r-2)/2$ when r is even and $5(r-1)/2$ when r is odd, that for the CoQDS model is $r-2$ when r is even and $r-1$ when r is odd, and that for the CoMDS model is $3(r-2)/2$ when r is even and $3(r-1)/2$ when r is odd. Note that the number of degrees of freedom for the CoDS model is equal to the sum of that for the CoQDS model and that of the CoMDS model.

3.2 Analysis of Data in Table 1

Consider the data in Table 1a, taken from Hashimoto (2003, p.144), which describes the cross-classification for father's and his daughter's occupational status categories in Japan. Table 1b is obtained by collapsing Table 1a into the 3×3 table by using cut points after the first and fourth rows and columns. Table 1c is obtained by collapsing Table 1a into the 3×3

table by using cut points after the second and third rows and columns. Table 3 gives the values of the goodness-of-fit test statistic G^2 for the models applied to the data in Table 1a. We see from Table 3 that each of CoQDS and QDS models fits the data in Table 1a well although each of CoDS, CoMDS, DS and MDS models fits those data poorly.

Since the CoDS model fits the data in Table 1a very poorly, we shall explore a reason of the poor fit of the CoDS model by considering the decomposition of the CoDS model. From Theorem 1, we can see that the lack of structure of the CoDS model is caused by the influence of the lack of structure of CoMDS model rather than the CoQDS model.

3.3 Analysis of Data in Table 2

Consider the data in Table 2a, taken from Katz (1978), which describes the cross-classification of the social status of occupations for the husband's father and the wife's father. Table 2b is obtained by collapsing Table 2a into the 3×3 table by using cut points after the first and third rows and columns. We see from Table 3 that each of CoDS, CoQDS, and CoMDS models fits the data in Table 2a well although each of DS, QDS, and MDS models fits those data poorly.

Table 3. Likelihood ratio chi-squared statistic G^2 values for models applied to Tables 1a and 2a.

	Table 1a		Table 2a	
Models	Degree of freedom	G^2	Degree of freedom	G^2
CoDS	10	463.34*	5	8.33
CoQDS	4	6.77	2	1.78
CoMDS	6	445.18*	3	6.44
DS	16	476.87*	10	62.07*
QDS	10	11.04	5	12.75*
MDS	6	445.18*	5	47.14*

* means significant at the 0.05 level.

We denote the MLE of $nG_{kl}^{(h,h')}$ by $\widehat{M}_{kl}^{(h,h')}$ where $n = 252$ (sample size). Under the CoDS model, we obtain

$$\widehat{M}_{11}^{(1,3)} = \widehat{M}_{33}^{(1,3)} = 38.00,\ \widehat{M}_{12}^{(1,3)} = \widehat{M}_{21}^{(1,3)} = \widehat{M}_{32}^{(1,3)} = \widehat{M}_{23}^{(1,3)} = 28.25,$$

and

$$\widehat{M}_{13}^{(1,3)} = \widehat{M}_{31}^{(1,3)} = 12.50.$$

Namely, (i) the probability that the status for the wife's father and the husband's father are both "High" is estimated to be equal to the probability that those are both "Low", (ii) the probability that wife's father's status is "High" and husband's father's status is "Middle" is estimated to be equal to the probability that wife's father's status is "Middle" and husband's father's status is "High", to the probability that wife's father's status is "Low" and husband's father's status is "Middle", and to the probability that wife's father's status is "Middle" and husband's father's status is "Low", and (iii) the probability that wife's father's status is "High" and husband's father's status is "Low" is estimated to be equal to the probability that wife's father's status is "Low" and husband's father's status is "High".

Under the CoMDS model, we obtain

$$\widehat{M}_{1\cdot}^{(1,3)} = \widehat{M}_{\cdot 1}^{(1,3)} = \widehat{M}_{3\cdot}^{(1,3)} = \widehat{M}_{\cdot 3}^{(1,3)} = 79.01,\ \widehat{M}_{2\cdot}^{(1,3)} = \widehat{M}_{\cdot 2}^{(1,3)} = 93.98,$$

where

$$\widehat{M}_{k\cdot}^{(h,h')} = \sum_{t=1}^{3} \widehat{M}_{kt}^{(h,h')},\ \widehat{M}_{\cdot l}^{(h,h')} = \sum_{s=1}^{3} \widehat{M}_{sl}^{(h,h')}.$$

Namely, (i) the probability that wife's father's status is "High" is estimated to be equal to the probability that wife's father's status is "Low", to the probability that husband's father's status is "High", and to the probability that husband's father's status is "Low", and (ii) the probability that wife's father's status is "Middle" is estimated to be equal to the probability that husband's father's status is "Middle".

4. Remarks

For the analysis of square contingency tables having same ordered row and column classifications, there may be a case that one wants to divide the categories having more than 3 categories into the simpler 3 categories. Then the CoDS, CoQDS and CoMDS models would be useful for seeing the structures of various DS for the $[(r-1)/2]$ ways of collapsed 3×3 tables.

Each of the DS and MDS models is invariant under the combining of adjacent categories as $T^{(h,h')}$ tables ($h = 1, \ldots, [(r-1)/2]$). Thus, if the DS model (or the MDS model) holds for an original $r \times r$ table, then the DS model (or the MDS model) holds for *all* the collapsed square tables, i.e., the CoDS model (or the CoMDS model) holds; but the converse does not always hold.

However, the QDS model is not invariant under combining of adjacent categories as $T^{(h,h')}$ tables. Therefore, when QDS model holds for an original $r \times r$ table, it is not guaranteed that then the QDS model holds for *all* the collapsed square tables.

Each of proposed models should be applied for the data on an ordinal scale because each model is not invariant under the arbitrary same permutation of the categories of rows and columns.

Acknowledgements

The authors are greatly thankful to two reviewers and the editor for the comments and suggestions to improve this paper.

Appendix

For the $r \times r$ contingency table with the cell probabilities $\{p_{ij}\}$, the DS, QDS, and MDS models in Tomizawa (1985) are defined as follows. The DS model is

$$p_{ij} = p_{ji} = p_{i^*j^*} = p_{j^*i^*} \qquad (i = 1, \ldots, r; j = 1, \ldots, r),$$

where $i^* = r + 1 - i$ and $j^* = r + 1 - j$. The QDS model is

$$p_{ij} = \mu\alpha_i\beta_j\psi_{ij} \qquad (i = 1, \ldots, r; j = 1, \ldots, r),$$

where

$$\psi_{ij} = \psi_{ji} = \psi_{i^*j^*} = \psi_{j^*i^*}.$$

The MDS model is

$$p_{i\cdot} = p_{\cdot i} = p_{i^*\cdot} = p_{\cdot i^*} \qquad (i = 1, \ldots, r),$$

where

$$p_{i\cdot} = \sum_{t=1}^{r} p_{it}, \ p_{\cdot i} = \sum_{s=1}^{r} p_{si}.$$

References

Agresti, A. (2013). *Categorical Data Analysis* (3rd ed.). New Jersey: Wiley.

Bishop, Y. M. M., Fienberg, S. E. & Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge: The MIT Press. http://doi.org/10.1007/978-0-387-72806-3

Bowker, A. H. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association, 43*, 572-574. http://doi.org/10.1080/01621459.1948.10483284

Caussinus, H. (1965). Contribution à l'analyse statistique des tableaux de corrélation. *Annales de la Faculté des Sciences de l'Université de Toulouse, 29*, 77-182.

Hashimoto, K. (2003). *Class Structure in Contemporary Japan*. Melbourne: Trans Pacific Press. http://doi.org/10.4057/jsr.50.164

Katz, B. M. (1978). Tests for equality of correlated proportions in a polychotomous response design. *Journal of Educational Statistics, 3*, 145-178. http://doi.org/10.2307/1164882

Tahata, K. and Tomizawa, S. (2014). Symmetry and asymmetry models and decompositions of models for contingency tables. *SUT Journal of Mathematics, 50*, 131-165.

Tomizawa, S. (1985). Double symmetry model and its decomposition in a square contingency table. *Journal of the Japan Statistical Society, 15*, 17-23. http://doi.org/10.11329/jjss1970.15.17

Wall, K. D. and Lienert, G. A. (1976). A test for point-symmetry in J-dimensional contingency-cubes. *Biometrical Journal, 18*, 259-264.

Yamamoto, K., Murakami, S. and Tomizawa, S. (2013). Point-symmetry models and decomposition for collapsed square contingency tables. *Journal of Applied Statistics, 40*, 1446-1452. http://doi.org/10.1080/02664763.2013.786028

Yamamoto, K., Tahata, K. and Tomizawa, S. (2012). Some symmetry models for the analysis of collapsed square contingency tables with ordered categories. *Calcutta Statistical Association Bulletin, 64*, 21-36.

International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Diagonal Exponent Conditional Symmetry Model for Square Contingency Tables with Ordered Categories

Kiyotaka Iki[1], Akira Shibuya[1] & Sadao Tomizawa[1]

[1] Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science, Chiba, Japan

Correspondence: Kiyotaka Iki, Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science, Noda City, Chiba, 278-8510, Japan. Tel: 81-4-7124-1501. E-mail: iki@is.noda.tus.ac.jp

Received: February 24, 2016 Accepted: April 1, 2016 Online Published: June 11, 2016

doi:10.5539/ijsp.v5n4p38 URL: http://dx.doi.org/10.5539/ijsp.v5n4p38

Abstract

For square contingency tables with ordered categories, this article proposes new models which indicate that in addition to the structure of asymmetry of the probabilities with respect to the main diagonal of the table, the expected frequency has an exponential form along every subdiagonal of the table. Also it gives the new three kinds of decompositions using the proposed model and proves the orthogonality of the test statistics.

Keywords: conditional symmetry, diagonal exponent symmetry, ordered category, orthogonal decomposition, square contingency table

1. Introduction

Consider an $R \times R$ square contingency table with the same row and column classifications. We may be interested in symmetry or asymmetry about the main diagonal of the table rather than independence. Let p_{ij} denote the probability that an observation will fall in the ith row and jth column of the table ($i = 1, \ldots, R$; $j = 1, \ldots, R$). Caussinus (1965) considered the quasi-symmetry (QS) model defined by

$$p_{ij} = \alpha_i \beta_j \psi_{ij} \quad (i = 1, \ldots, R; j = 1, \ldots, R),$$

where $\psi_{ij} = \psi_{ji}$. The QS model with $\{\alpha_i = \beta_i\}$ is the symmetry (S) model (see, Bowker, 1948; Martin & Pardo, 2010; Kolassa & Bhagavatula, 2012). The conditional symmetry (CS) model is defined by

$$p_{ij} = \begin{cases} \gamma \psi_{ij} & (i < j), \\ \psi_{ij} & (i \geq j), \end{cases}$$

where $\psi_{ij} = \psi_{ji}$; see McCullagh (1978). The CS model states that p_{ij} ($i < j$) is γ times higher than p_{ji}. The CS model with $\gamma = 1$ is the S model.

The global symmetry (GS) model is defined by

$$\sum_{i<j} \sum p_{ij} = \sum_{i<j} \sum p_{ji};$$

see Read (1977). The GS model states that the probability that an observation will fall in one of the upper-right triangle cells above the main diagonal of the table is equal to the probability that it falls in one of the lower-left triangle cells below the main diagonal. Read (1977) gave the theorem that the S model holds if and only if both the CS and GS models hold.

Tomizawa (1992) considered the diagonal exponent symmetry (DES) model defined by

$$p_{ij} = \begin{cases} \delta^{i+j} d_{|j-i|} & (i \neq j), \\ \psi_{ii} & (i = j). \end{cases}$$

The DES model states that in addition to the structure of the S model, $p_{i+1,j+1}$ ($i \neq j$) is δ^2 times higher than p_{ij}; in other words, for fixed distance k ($k = 1, \ldots, R-2$) from the main diagonal of the table, $p_{i,i+k}$ increase (decrease) exponentially along every subdiagonal of the table as the value i increase ($i = 1, \ldots, R-k$).

Iki, Yamamoto & Tomizawa (2014) considered the quasi-diagonal exponent symmetry (QDES) model defined by

$$p_{ij} = \begin{cases} \alpha^i \beta^j d_{|j-i|} & (i \neq j), \\ \psi_{ii} & (i = j). \end{cases}$$

The QDES model with $\alpha = \beta$ is the DES model. The QDES model states that in addition to the structure of the QS model (instead of the S model), the expected frequency has an exponential form along every subdiagonal of the table. Under the QDES model, we see the structure of $p_{ij}/p_{ji} = (\beta/\alpha)^{j-i}$ $(i < j)$.

Let X and Y denote the row and column variables, respectively. We define the mean equality (ME) model as $\mathrm{E}(X) = \mathrm{E}(Y)$. Iki et al. (2014) gave the theorem that the DES model holds if and only if both the QDES and ME models hold. Other symmetry and asymmetry models have been described in Tahata & Tomizawa (2014).

We are interested in considering new models which indicate that in addition to the structure of the CS model (instead of the S model), the expected frequency has an exponential form along every subdiagonal of the table. The present paper proposes two new models and gives the new three kinds of decompositions of the DES model.

2. New Models

Consider a model defined by

$$p_{ij} = \begin{cases} \delta^{i+j} d_{j-i} & (i \neq j), \\ \psi_{ii} & (i = j), \end{cases}$$

where $d_{j-i} = \gamma d_{i-j}$ $(i < j)$. This model states that in addition to the structure of the CS model, $p_{i+1,j+1}$ $(i \neq j)$ is δ^2 times higher than p_{ij}. Thus we shall refer to this model as the diagonal exponent conditional symmetry (DECS) model. Under the DECS model, we see the structure of $p_{ij}/p_{ji} = \gamma$ $(i < j)$. The DECS model with $\gamma = 1$ is the DES model.

Next, consider a model defined by

$$p_{ij} = \begin{cases} \alpha^i \beta^j d_{j-i} & (i \neq j), \\ \psi_{ii} & (i = j), \end{cases}$$

where $d_{j-i} = \gamma d_{i-j}$ $(i < j)$. We shall refer to this model as the quasi-diagonal exponent conditional symmetry (QDECS) model. Under the QDECS model, we see the structure of $p_{ij}/p_{ji} = \gamma(\beta/\alpha)^{j-i}$ $(i < j)$. The QDECS model with $\gamma = 1$ is the QDES model. Also, QDECS model with $\alpha = \beta$ is the DECS model.

Figure 1 shows the relationships among the models. In figure, $A \to B$ indicates that model A implies model B.

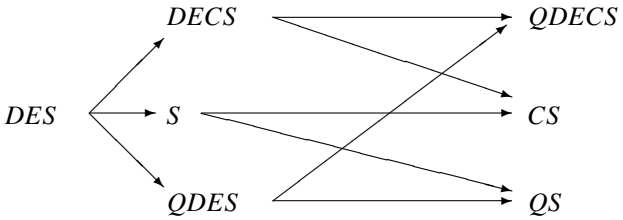


Figure 1. Relationships among models.

3. Decompositions and Orthogonality of Test Statistics

We obtain the new three kinds of decompositions of the DES model as follows:

Theorem 1 *The DES model holds if and only if all the QDECS, GS and ME models hold.*

Theorem 2 *The DES model holds if and only if both the DECS and GS models hold.*

Theorem 3 *The DES model holds if and only if both the DECS and ME models hold.*

The proofs of these theorems are given in Appendix 1.

Consider the model that has the structure of both the GS and ME models. We shall refer to this model as the GSME model. From Theorem 1, we can obtain the following the corollary:

Corollary 1 *The DES model holds if and only if the QDECS and GSME models hold.*

Let n_{ij} denote the observed frequency in the (i, j)th cell of the table $(i = 1, \ldots, R; j = 1, \ldots, R)$ with $n = \sum\sum n_{ij}$, and let m_{ij} denote the corresponding expected frequency. Assume that $\{n_{ij}\}$ have a multinomial distribution. The maximum likelihood estimates (MLEs) of $\{m_{ij}\}$ under the DECS and QDECS models could be obtained using iterative procedures; for example, see Darroch & Ratcliff (1972). The MLEs of $\{m_{ij}\}$ under the GSME model could be obtained using Newton-Raphson method to the log-likelihood equations.

Let $G^2(M)$ denote the likelihood ratio chi-squared statistic for testing goodness-of-fit of model M. The numbers of degrees of freedom (df) for the DECS and QDECS models are $R^2 - 2R - 1$ and $R^2 - 2R - 2$, respectively.

The orthogonality (asymptotic separability or independence) of the test statistics for goodness-of-fit of two models is discussed by, e.g., Darroch & Silvey (1963) and Read (1977). We obtain the following theorems for the orthogonality.

Theorem 4 *The test statistic $G^2(DES)$ is asymptotically equivalent to the sum of $G^2(QDECS)$ and $G^2(GS\,ME)$.*

Theorem 5 *The test statistic $G^2(DES)$ is asymptotically equivalent to the sum of $G^2(DECS)$ and $G^2(GS)$.*

The proof of Theorem 4 is given in Appendix 2. We shall omit the proof of Theorem 5 because it is obtained in a similar way to the proof of Theorem 4.

4. An Example

Consider the data in Table 1 taken from Agresti (2002, p. 462). These data are insomniac patient's reported time (in minutes) to fall asleep after going to bed. The response is the patient's reported time at baseline (before treatment) and following two weeks of treatment (hypnotic drug).

Table 1. Insomniac patient's reported time (in minutes) to fall asleep after going to bed; from Agresti (2002, p. 462). (The upper and lower parenthesized values are MLEs of expected frequencies under the DECS model and the special DECS model with $\delta = 1$, respectively.)

	Follow-up				
Initial	< 20	20-30	30-60	> 60	Total
< 20	7	4	1	0	12
	(7.00)	(1.81)	(1.69)	(0.94)	
	(7.00)	(1.88)	(1.72)	(0.94)	
20-30	11	5	2	2	20
	(15.58)	(5.00)	(1.87)	(1.75)	
	(16.12)	(5.00)	(1.88)	(1.72)	
30-60	13	23	3	1	40
	(14.53)	(16.12)	(3.00)	(1.94)	
	(14.78)	(16.12)	(3.00)	(1.88)	
> 60	9	17	13	8	47
	(8.06)	(15.03)	(16.67)	(8.00)	
	(8.06)	(14.78)	(16.12)	(8.00)	
Total	40	49	19	11	119

We see from Table 2 that the CS, DECS and QDECS models fit these data well, although the other models fit poorly. Since the DECS model is a special case of the QDECS model, we shall test the hypothesis that the DECS model holds (i.e., the hypothesis of $\alpha = \beta$) assuming that the QDECS model holds. Since $G^2(DECS|QDECS) = G^2(DECS) - G^2(QDECS) = 1.54$ with 1 df being the difference between the numbers of df for the DECS and the QDECS models, this hypothesis is accepted at the 0.05 significance level. Similarly, the hypothesis that the DECS model holds assuming that the CS model holds is accepted for these data. Therefore, the DECS model would be preferable to the CS and QDECS models.

Under the DECS model, the MLEs of parameters of γ and δ are $\widehat{\gamma} = 0.116$ and $\widehat{\delta} = 1.017$. Since $\widehat{\delta}$ is close to 1, we are now interested in a special DECS model obtained by putting $\delta = 1$. For this model we obtain the likelihood ratio statistic $G^2(DECS \text{ with } \delta = 1) = 10.30$ with 8 df. Thus the special DECS model with $\delta = 1$ also fits these data well. Moreover, we shall test the hypothesis that the special DECS model with $\delta = 1$ holds (i.e., the hypothesis of $\delta = 1$) assuming that the DECS model holds for these data. Since $G^2(DECS \text{ with } \delta = 1|DECS) = G^2(DECS \text{ with } \delta = 1) - G^2(DECS) = 0.05$ with 1 df, this hypothesis is accepted at the 0.05 significance level. Therefore the special DECS model with $\delta = 1$ may be preferable to the DECS model.

Under the special DECS model with $\delta = 1$, the MLE of parameter γ is $\widehat{\gamma} = 0.116$. Thus, under the special DECS model with $\delta = 1$, the probability that a patient's reported time at baseline and his or her reported time at following two weeks of treatment are i and j $(i > j)$, respectively, is estimated to be $\widehat{\gamma}^{-1} = 8.621$ times higher than the probability that those are j and i, respectively. Thus, since $\widehat{\gamma}^{-1} > 1$, the patient's reported time at following two weeks of treatment is faster than the patient's reported time at baseline. Also under this model, the probability that a patient's reported time at baseline and his or her reported time at following two weeks of treatment are $i + 1$ and $j + 1$, respectively, is estimated to be equal to the probability that those are i and j, respectively.

We see from Table 2 and Theorem 1 that the poor fit of the DES model is caused by the influence of the lack of structure of the GS and ME models rather than the QDECS model. Similarly, from Theorem 2, the poor fit of the DES model is caused by the influence of the lack of structure of the GS model rather than the DECS model. Also, from Theorem 3, the poor fit of the DES model is caused by the influence of the lack of structure of the ME model rather than the DECS model.

Table 2. Likelihood ratio chi-squared values G^2 for models applied to Table 1.

Applied models	df	G^2
S	6	74.55*
CS	5	5.62
DES	8	79.18*
DECS	7	10.25
QDECS	6	8.71
GS	1	68.93*
ME	1	66.64*
GSME	2	71.19*

$*$ means significant at the 0.05 level.

5. Concluding Remarks

We have proposed the DECS and QDECS models, and given the three kinds of decompositions of the DES model. These decompositions may be useful for seeing the reason for the poor fit of the DES model.

The $G^2(DES)$ is asymptotically equivalent to the sum of values $G^2(QDECS)$ and $G^2(GSME)$ as described by Theorem 4. However, we point out that for the decomposition in Theorem 1, the $G^2(DES)$ is not asymptotically equivalent to the sum of values $G^2(QDECS)$, $G^2(GS)$ and $G^2(ME)$ because the sum of values $G^2(GS)$ and $G^2(ME)$ is not asymptotically equivalent to the $G^2(GSME)$.

Acknowledgements

The authors would like to thank the editor and referees for their helpful comments.

Appendix 1

Proof of Theorem 1. If the DES model holds, then the QDECS, GS and ME models hold. Assuming that all the QDECS, GS and ME models hold, then we shall show that the DES model holds. Since the QDECS model holds, we have

$$p_{st} - p_{ts} = (\gamma\theta^{t-s} - 1)p_{ts} \quad (s < t), \tag{A.1}$$

where $\theta = \beta/\alpha$. Since (A.1) and the GS model hold, we see

$$\sum_{s<t}\sum(\gamma\theta^{t-s} - 1)p_{ts} = 0,$$

namely,

$$\sum_{k=1}^{R-1}\sum_{s=1}^{R-k}(\gamma\theta^{k} - 1)p_{s+k,s} = 0. \tag{A.2}$$

From (A.2), we see

$$\gamma = \frac{\sum_{\ell=1}^{R-1}\sum_{t=1}^{R-\ell} p_{t+\ell,t}}{\sum_{k=1}^{R-1}\sum_{s=1}^{R-k} p_{s+k,s}\theta^{k}}. \tag{A.3}$$

The ME model can be expressed as

$$\sum_{i=1}^{R-1} G_{1(i)} = \sum_{i=1}^{R-1} G_{2(i)}, \tag{A.4}$$

where

$$G_{1(i)} = \sum_{s=1}^{i}\sum_{t=i+1}^{R} p_{st}, \quad G_{2(i)} = \sum_{s=1}^{i}\sum_{t=i+1}^{R} p_{ts}.$$

From (A.1) and (A.4), we see

$$\sum_{i=1}^{R-1}\sum_{s=1}^{i}\sum_{t=i+1}^{R}(\gamma\theta^{t-s}-1)p_{ts}=0,$$

namely,

$$\sum_{k=1}^{R-1}\sum_{s=1}^{R-k}k(\gamma\theta^{k}-1)p_{s+k,s}=0. \tag{A.5}$$

From (A.5), we see

$$\gamma=\frac{\sum_{\ell=1}^{R-1}\sum_{t=1}^{R-\ell}\ell p_{t+\ell,t}}{\sum_{k=1}^{R-1}\sum_{s=1}^{R-k}kp_{s+k,s}\theta^{k}}. \tag{A.6}$$

From (A.3) and (A.6), we obtain

$$\left(\sum_{\ell=1}^{R-1}\sum_{t=1}^{R-\ell}p_{t+\ell,t}\right)\left(\sum_{k=1}^{R-1}\sum_{s=1}^{R-k}kp_{s+k,s}\theta^{k}\right)-\left(\sum_{\ell=1}^{R-1}\sum_{t=1}^{R-\ell}\ell p_{t+\ell,t}\right)\left(\sum_{k=1}^{R-1}\sum_{s=1}^{R-k}p_{s+k,s}\theta^{k}\right)=0,$$

namely,

$$\sum_{k=1}^{R-1}\sum_{\ell=1}^{R-1}\sum_{s=1}^{R-k}\sum_{t=1}^{R-\ell}(k-\ell)p_{s+k,s}p_{t+\ell,t}\theta^{k}=0. \tag{A.7}$$

The equation (A.7) is also expressed as

$$(\theta-1)\sum_{m=2}^{R-1}\left(\theta^{m-1}\sum_{g=m}^{R-1}\sum_{s=1}^{R-g}\sum_{\ell=1}^{R-1}\sum_{t=1}^{R-\ell}(g-\ell)p_{s+g,s}p_{t+\ell,t}\right)=0. \tag{A.8}$$

In addition,

$$\sum_{g=m}^{R-1}\sum_{s=1}^{R-g}\sum_{\ell=1}^{R-1}\sum_{t=1}^{R-\ell}(g-\ell)p_{s+g,s}p_{t+\ell,t}=\sum_{g=m}^{R-1}\sum_{\ell=1}^{m-1}\sum_{s=1}^{R-g}\sum_{t=1}^{R-\ell}(g-\ell)p_{s+g,s}p_{t+\ell,t}+\sum_{g=m}^{R-1}\sum_{\ell=m}^{R-1}\sum_{s=1}^{R-g}\sum_{t=1}^{R-\ell}(g-\ell)p_{s+g,s}p_{t+\ell,t}. \tag{A.9}$$

The first term on the right-hand side of (A.9) is positive and the second term equals zero. Thus (A.9) is positive. Therefore, noting that $\theta>0$, from (A.8) we obtain $\theta=1$, i.e., $\alpha=\beta$. Thus, from (A.2) we obtain $\gamma=1$. Namely, the DES model holds. The proof is complicated.

Proof of Theorem 2. If the DES model holds, then the DECS and GS models hold. Assuming that both the DECS and GS models hold, then we shall show that the DES model holds. Since the DECS and GS models hold, we see

$$\begin{aligned}\sum\sum_{s<t}p_{st}-\sum\sum_{s<t}p_{ts} &= \sum\sum_{s<t}\delta^{s+t}d_{t-s}-\sum\sum_{s<t}\delta^{s+t}d_{s-t}\\ &= \sum\sum_{s<t}\delta^{s+t}\gamma d_{s-t}-\sum\sum_{s<t}\delta^{s+t}d_{s-t}\\ &= (\gamma-1)\sum\sum_{s<t}\delta^{s+t}d_{s-t}\\ &= 0.\end{aligned}$$

Thus, we obtain $\gamma=1$. Namely, the DES model holds. The proof is complicated.

Proof of Theorem 3. If the DES model holds, then the DECS and ME models hold. Assuming that both the DECS and ME models hold, then we shall show that the DES model holds. Since the DECS and ME models hold, we see

$$\begin{aligned}\sum_{i=1}^{R-1}\sum_{s=1}^{i}\sum_{t=i+1}^{R}p_{st}-\sum_{i=1}^{R-1}\sum_{s=1}^{i}\sum_{t=i+1}^{R}p_{ts} &= \sum_{i=1}^{R-1}\sum_{s=1}^{i}\sum_{t=i+1}^{R}\delta^{s+t}d_{t-s}-\sum_{i=1}^{R-1}\sum_{s=1}^{i}\sum_{t=i+1}^{R}\delta^{s+t}d_{s-t}\\ &= \sum_{i=1}^{R-1}\sum_{s=1}^{i}\sum_{t=i+1}^{R}\delta^{s+t}\gamma d_{s-t}-\sum_{i=1}^{R-1}\sum_{s=1}^{i}\sum_{t=i+1}^{R}\delta^{s+t}d_{s-t}\\ &= (\gamma-1)\sum_{i=1}^{R-1}\sum_{s=1}^{i}\sum_{t=i+1}^{R}\delta^{s+t}d_{s-t}\\ &= 0.\end{aligned}$$

Thus, we obtain $\gamma = 1$. Namely, the DES model holds. The proof is complicated.

Appendix 2

Proof of Theorem 4. The QDECS model is expressed as

$$\log p_{ij} = \begin{cases} \gamma^* + i\alpha^* + j\beta^* + d^*_{i-j} & (i < j), \\ i\alpha^* + j\beta^* + d^*_{j-i} & (i > j), \\ \psi^*_{ii} & (i = j). \end{cases} \tag{A.10}$$

Let

$$p = (p_{11}, \ldots, p_{1R}, p_{21}, \ldots, p_{2R}, \ldots, p_{R1}, \ldots, p_{RR})^t,$$
$$\beta = (\gamma^*, \alpha^*, \beta^*, \phi)^t,$$

where "t" denotes the transpose, and

$$\phi = (d^*_{-1}, d^*_{-2}, \ldots, d^*_{-(R-1)}, \psi^*_{11}, \psi^*_{22}, \ldots, \psi^*_{RR}),$$

is the $1 \times (2R-1)$ vector. The QDECS model is expressed as

$$\log p = X\beta = (X_0, X_1, X_2, X_3)\beta,$$

where X is the $R^2 \times L$ matrix with $L = 2R + 2$, $X_0 = (v_1, \ldots, v_R)^t$ (the $R^2 \times 1$ vector), $X_1 = J_R \otimes 1_R$ (the $R^2 \times 1$ vector), $X_2 = 1_R \otimes J_R$ (the $R^2 \times 1$ vector), and X_3 is the $R^2 \times (2R-1)$ matrix of 1 or 0 elements determined from (A.10); and where v_p is the $1 \times R$ vector of 0 for the first p elements or 1 for the others, 1_s is the $s \times 1$ vector of 1 elements, $J_R = (1, \ldots, R)^t$ and $\otimes$ denotes the Kronecker product. The matrix X is full column rank which is L. In a similar manner to Haber (1985), we denote the linear space spanned by the columns of the matrix X by $S(X)$ with the dimension L.

Let U be an $R^2 \times l_1$, where $l_1 = R^2 - L = R^2 - 2R - 2$, full column rank matrix such that the linear space spanned by the columns of U, i.e., $S(U)$, is the orthogonal complement of $S(X)$. Thus, $U^tX = O_{l_1,L}$, where $O_{s,t}$ is the $s \times t$ zero matrix. Therefore the QDECS model is expressed as

$$h_1(p) = 0_{l_1},$$

where 0_s is the $s \times 1$ zero vector, and $h_1(p) = U^t \log p$. The GSME model is expressed as

$$h_2(p) = 0_{l_2},$$

where $l_2 = 2$ and $h_2(p) = Wp$ with

$$W = \begin{pmatrix} (2X_0 - 1_{R^2} + w_1 + w_2 + \cdots + w_R)^t \\ (X_2 - X_1)^t \end{pmatrix}; \text{ the } 2 \times R^2 \text{ matrix},$$

where w_i $(i = 1, \ldots, R)$ is the $R^2 \times 1$ vector, being one of column vectors in X_3 shouldering ψ^*_{ii}. Note that $X_3 1_{2R-1} = 1_{R^2}$. Thus W^t belongs to $S(X)$, i.e., $S(W^t) \subset S(X)$. Hence $WU = O_{l_2,l_1}$. From Corollary 1, the DES model is expressed as

$$h_3(p) = 0_{l_3},$$

where $l_3 = l_1 + l_2 = R^2 - 2R$, and $h_3 = (h_1^t, h_2^t)^t$.

Let $H_s(p)$ $(s = 1, 2, 3)$ denote the $l_s \times R^2$ matrix of partial derivative of $h_s(p)$ with respect to p, i.e., $H_s(p) = \partial h_s(p)/\partial p^t$. Let $\Sigma(p) = diag(p) - pp^t$, where $diag(p)$ denotes a diagonal matrix with ith component of p as ith diagonal component. Let $\widehat{p}$ denote p with $\{p_{ij}\}$ replaced by $\{\widehat{p}_{ij} = n_{ij}/n\}$. Then $\sqrt{n}(\widehat{p} - p)$ has asymptotically a normal distribution with mean 0_{R^2} and covariance matrix $\Sigma(p)$. Using the delta method, $\sqrt{n}(h_3(\widehat{p}) - h_3(p))$ has asymptotically a normal distribution with mean 0_{l_3} and covariance matrix

$$H_3(p)\Sigma(p)H_3(p)^t = \begin{bmatrix} H_1(p)\Sigma(p)H_1(p)^t & H_1(p)\Sigma(p)H_2(p)^t \\ H_2(p)\Sigma(p)H_1(p)^t & H_2(p)\Sigma(p)H_2(p)^t \end{bmatrix}.$$

Since $H_1(p)p = U^t 1_{R^2} = 0_{l_1}$, $H_1(p)diag(p) = U^t$ and $H_2(p) = W$, we see

$$H_1(p)\Sigma(p)H_2(p)^t = U^tW^t = O_{l_1,l_2}.$$

Thus, we obtain $\Delta_3(p) = \Delta_1(p) + \Delta_2(p)$, where

$$\Delta_s(p) = h_s(p)^t [H_s(p)\Sigma(p)H_s(p)^t]^{-1} h_s(p). \tag{A.11}$$

Under each $h_s(p) = 0_{l_s}$ $(s = 1, 2, 3)$, the Wald statistic $W_s = n\Delta_s(\widehat{p})$ has asymptotically a chi-squared distribution with l_s degrees of freedom. From (A.11), we see that $W_3 = W_1 + W_2$. From the asymptotic equivalence of the Wald statistic and likelihood ratio statistic, we obtain Theorem 4.

References

Agresti, A. (2002). *Categorical Data Analysis, 2nd ed.* Wiley, New York.

Bowker, A. H. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association, 43*, 572-574. http://dx.doi.org/10.1080/01621459.1948.10483284

Caussinus, H. (1965). Contribution à l'analyse statistique des tableaux de corrélation. *Annales de la Faculté des Sciences de l'Université de Toulouse, 29*, 77-182.

Darroch, J. N. & Silvey, S. D. (1963). On testing more than one hypothesis. *Annals of Mathematical Statistics, 34*, 555-567. http://dx.doi.org/10.1214/aoms/1177704168

Darroch, J. N. & Ratcliff, D. (1972). Generalized iterative scaling for log-linear models. *Annals of Mathematical Statistics, 43*, 1470-1480. http://dx.doi.org/10.1214/aoms/1177692379

Haber, M. (1985). Maximum likelihood methods for linear and log-linear models in categorical data. *Computational Statistics and Data Analysis, 3*, 1-10. http://dx.doi.org/10.1016/0167-9473(85)90053-2

Iki, K., Yamamoto, K. & Tomizawa, S. (2014). Quasi-diagonal exponent symmetry model for square contingency tables with ordered categories. *Statistics and Probability Letters, 92*, 33-38. http://dx.doi.org/10.1016/j.spl.2014.04.029

Kolassa, J. E. & Bhagavatula, H. G. (2012). Accurate approximations to the distribution of a statistic testing symmetry in contingency tables. *Institute of Mathematical Statistics, 8*, 181-189. http://dx.doi.org/10.1214/11-IMSCOLL812

Martin, N. & Pardo, L. (2010). A new measure of leverage cells in multinomial loglinear models. *Communications in Statistics-Theory and Methods, 39*, 517-530. http://dx.doi.org/10.1080/03610920903139991

McCullagh, P. (1978). A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika, 65*, 413-418. http://dx.doi.org/10.1093/biomet/65.2.413

Read, C. B. (1977). Partitioning chi-square in contingency table: A teaching approach. *Communications in Statistics-Theory and Methods, 6*, 553-562. http://dx.doi.org/10.1080/03610927708827513

Tahata, K. & Tomizawa, S. (2014). Symmetry and asymmetry models and decompositions of models for contingency tables. *SUT journal of Mathematics, 50*, 131-165.

Tomizawa, S. (1992). A model of symmetry with exponents along every subdiagonal and its application to data on unaided vision of pupils at Japanese elementary schools. *Journal of Applied Statistics, 19*, 509-512. http://dx.doi.org/10.1080/02664769200000046

Copyrights

International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Dynamic Reliability of a Cluster Server

Andrzej Korzeniowski[1], Rachel Traylor[1]

[1] Department of Mathematics, University of Texas at Arlington

Correspondence: Andrzej Korzeniowski, Department of Mathematics, University of Texas at Arlington, Arlington, TX 76019, USA. Tel: 1-817-272-3595. E-mail: korzeniowski@uta.edu

Received: April 8, 2016 Accepted: April 18, 2016 Online Published: June 11, 2016

doi:10.5539/ijsp.v5n4p45 URL: http://dx.doi.org/10.5539/ijsp.v5n4p45

Abstract

Suppose a single server has K channels, each of which performs a different task. Customers arrive to the server via a nonhomogenous Poisson process with intensity $\lambda(t)$ and select 0 to K tasks for the server to perform. Each channel services the tasks in its queue independently, and the customer's job is complete when the last task selected is complete. The stress to the server is a constant multiple η of the number of tasks selected by each customer, and thus the stress added to the server by each customer is random. Under this model, we provide the survival function for such a server in both the case of independently selected channels and correlated channels. A numerical comparison of expected lifetimes for various arrival rates is given, and the relationship between the dependency of channel selection and expected server lifetime is presented.

Keywords: cluster server, correlated Bernoulli variables, reliability, dynamic server lifetime

1. Introduction

Many single-server queuing models assume each arrival brings one task that must be completed by the server (Gross & Shortle, 2008; Tang, 1997). Common examples include a cashier's register, a hostess stand at a restaurant, or a basic web server. Of particular interest is the reliability of such queuing servers under workload or stress. In particular, Cha and Lee (2011) formulated a dynamic reliability model for a web server wherein customers arrive via a nonhomogenous Poisson process $\{N(t), t \geq 0\}$, the service times are i.i.d. with general distribution $G(w)$, and each task brought by a customer increases the stochastic hazard function by a constant η for the duration of its time in the system. (Cha and Lee, 2011).

Traylor (2015) provided a generalization to the model of Cha and Lee that relaxed the constant stress assumption and allowed for the job stresses to be i.i.d. with probability distribution $\mathcal{H}$. It is assumed that the set of arrival times $\mathfrak{T} = \{T_j\}_{j=1}^{N(t)}$ are mutually independent. The service times W_j, $j = 1, ..., N(t)$ remain i.i.d. and mutually independent of the set of arrival times. It is also assumed that the job stresses $\mathcal{H}_j$, $j = 1, ..., N(t)$ are mutually independent of all arrival times and service times. Under these conditions, Traylor provides the survival function for such a server in the following theorem.

Theorem 1. *Suppose that jobs arrive to a server according to a nonhomogenous Poisson process $\{N(t), t \geq 0\}$ with intensity function $\lambda(t) \geq 0$ Let the arrival times $\{T_j\}_{j=1}^{N(t)}$ be independent, and let the service times $\{W_j\}_{j=1}^{N(t)} \overset{i.i.d.}{\sim} g_W(w)$ be mutually independent of all arrival times. Assume the random job stresses $\mathcal{H}_j \overset{i.i.d.}{\sim} \mathcal{H}$, $j = 1, ..., N(t)$. Then*

$$S_Y(t) = \bar{F}_0(t) \exp\left(-E_{\mathcal{H}}\left[\mathcal{H} \int_0^t e^{-\mathcal{H}w} m(t-w) \bar{G}_W(w) dw\right]\right) \tag{1}$$

where $\bar{F}_0(t) = \exp\left(-\int_0^t r_0(s)ds\right)$, $\bar{G}_w(w) = 1 - G(w)$, $m(t) \equiv E[N(t)] = \int_0^t \lambda(x)dx$. and $r_0(s)$ is the hazard function of the idle server at time s.

Suppose now that this single server is partitioned into K channels, with each channel assigned to a different possible task a customer may choose. Customers may choose 0 to K tasks for the server to complete, and the customer is fully serviced when the last task is complete. The choice of such tasks may be correlated or uncorrelated. We propose a generalization to the random stress model of Traylor (2015) that gives the survival function for such a "clustered-task" server in both the correlated and independent cases. This provides a highly general reliability model for a variety of server types in many different industries. It also allows for a correlated multi-server system to be analyzed in a simple manner.

The paper is organized as follows. Section 2 gives a general description of the server model and its assumptions. Section 3 examines the reliability of the server when the channel selection is independent; correlated channels are discussed in

section 4. Section 5 provides some numerical illustrations of the effect of channel selection dependency on the expected lifetime of the server, and Section 6 gives the conclusion.

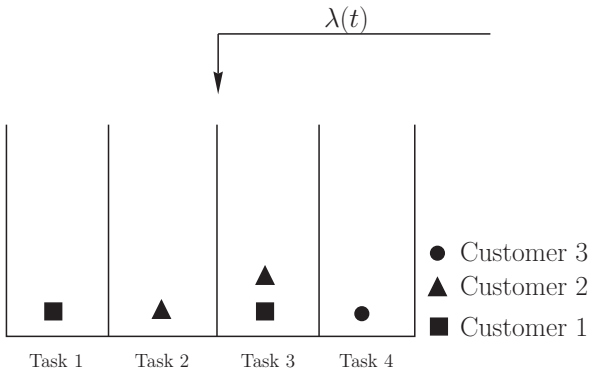


Figure 1. Illustration of Clustered Tasks in a Multichannel Server

2. Model Description

Suppose a server has multiple channels $Q_1, ..., Q_K$, but each channel serves a different type of task. A customer arrives to the server and may select any number from 0 to K tasks for the server to perform. Said customer will select each possible task j with probability p_j. Figure 1 illustrates an example of such a situation in which three customers visit the server and each customer picks a different number and set of tasks at random. A customer is considered fully serviced (i.e. the job is complete) upon completion of the last task belonging to that particular customer.

2.1 Model Assumptions

The following mathematical assumptions are made for the multichannel server with clustered tasks:

(i) Customers arrive to the server with K channels via a nonhomogenous Poisson process (NHPP) with intensity $\lambda(t)$.

(ii) The breakdown rate of the idle server is given by $r_0(t)$.

(iii) Each channel corresponds to a different task the server can perform.

(iv) The selection of each task is a Bernoulli random variable with probability p_k. Thus the number of tasks selected by each customer is a binomial random variable.

(v) The workload stress to the server is a constant multiple η of the number of tasks requested by the customer, i.e. the additional stress is given by ηN, where N is the number of tasks requested.

(vi) The distribution of each channel's service time is given by $G_i(w), i = 1, ..., K$. Since the customer's service is not complete until all requested tasks have finished, the service life distribution for the customers is therefore given by $\max_i G_i(w)$.

Under these assumptions, this model is a special interpretation of the random stress environment developed in Traylor (2015). In this case, the random workload stress is ηN, where N is a binomial random variable, and the service life distribution $G_W(w) = \max_i G_i(w)$, which may be easily obtained through the mathematical properties of order statistics. Two variations are considered in this section: independent channels and correlated channels.

3. Independent Channels

Suppose the selection probabilities for each task in the server are identical, that is, $p_1 = p_2 = \ldots = p_K = p$. Then $N \sim \text{Bin}(K, p)$. Using Theorem 1, the survival function of the multichannel server is given in the following theorem:

Theorem 2 (Survival Function of Multichannel Server with Clustered Tasks and Independent Channels)**.** *Suppose conditions (i)-(vi) above are satisfied. In addition, assume $p_1 = p_2 = \ldots = p_K = p$. Then the survival function of the server is given by*

$$S_Y(t) = \bar{F}_0(t) \exp\left(-K\eta\left[e^{-\eta t}\left(1 - p + pe^{-\eta t}\right)^{K-1} - p(1-p)^{K-1}\right]\int_0^t m(t-w)\bar{G}_W(w)dw\right)$$

where $m(x) = \int_0^x \lambda(s)ds$, $\bar{F}_0(t) = e^{-\int_0^t r_0(s)ds}$, $\bar{G}_W(w) = 1 - G_W(w)$, and $G_W(w) = \max_i G_i(w)$.

Proof. Since $p_1 = \ldots = p_K = p$, the number of tasks selected by any particular customer $N \sim \text{Bin}(K, p)$. Then the random stress distribution $\mathcal{H}$ is given by $\mathcal{H} = \eta N$. Thus

$$S_Y(t) = \bar{F}_0(t) \exp\left(-E_{\mathcal{H}}\left[\mathcal{H} \int_0^t e^{-\mathcal{H}w} m(t-w)\bar{G}_W(w)dw\right]\right)$$

In this case,

$$\begin{aligned}
&E\left[\mathcal{H} \int_0^t e^{-\mathcal{H}w} m(t-w)\bar{G}_W(w)dw\right] \\
&\qquad = E\left[\eta N \int_0^t e^{-\eta N w} m(t-w)\bar{G}_W(w)dw\right] \\
&\qquad = \sum_{n=0}^{K}\left[\eta n \int_0^t e^{-\eta n w} m(t-w)\bar{G}_W(w)dw\right] \cdot P(N=n) \\
&\qquad = \sum_{n=0}^{K}\left[\eta n \int_0^t e^{-\eta n w} m(t-w)\bar{G}_W(w)dw\right]\binom{K}{n} p^n (1-p)^{K-n} \\
&\qquad = \eta \int_0^t m(t-w)\bar{G}_W(w)\left(\sum_{n=0}^{K} n e^{-\eta n w}\binom{K}{n} p^n (1-p)^{K-n}\right)dw
\end{aligned}$$

Now,

$$\begin{aligned}
\sum_{n=0}^{K} n e^{-\eta n w}\binom{K}{n} p^n (1-p)^{K-n} &= \sum_{n=0}^{K} \frac{K!}{(K-n)!n!} n e^{-\eta n w} p^n (1-p)^{K-n} \\
&= \sum_{n=0}^{K} \frac{K(K-1)!}{(n-1)!(K-1-(n-1))!} e^{-\eta n w} p^n (1-p)^{K-n} \\
&= \sum_{n=0}^{K} K\binom{K-1}{n-1} e^{-\eta n w} p^n (1-p)^{K-n}
\end{aligned}$$

Making a change of indices, let $j = n - 1$. Then

$$\sum_{n=0}^{K} K\binom{K-1}{n-1} e^{-\eta n w} p^n (1-p)^{K-n} = K \sum_{j=0}^{K-1}\binom{K-1}{j} p^{j+1} (1-p)^{K-(j+1)} e^{-\eta(j+1)w}$$

Note the above resembles a scaled and shifted moment generating function of a binomial random variable. Let $X \sim \text{Bin}(K-1, p)$. Then

$$\begin{aligned}
K \sum_{j=0}^{K-1}\binom{K-1}{j} p^{j+1} (1-p)^{K-(j+1)} e^{-\eta(j+1)w} &= K\left(E\left[e^{-\eta(X+1)t}\right] - P(X=0)\right) \\
&= K\left(e^{-\eta t} E\left[e^{-\eta X t} - p(1-p)^{K-1}\right]\right) \\
&= K\left(e^{-\eta t}\left[1 - p + p e^{-\eta t}\right]^{K-1} - p(1-p)^{K-1}\right)
\end{aligned}$$

Thus,

$$S_Y(t) = \bar{F}_0(t) \exp\left(-K\eta\left[e^{-\eta t}\left(1 - p + p e^{-\eta t}\right)^{K-1} - p(1-p)^{K-1}\right] \int_0^t m(t-w)\bar{G}_W(w)dw\right)$$

□

4. Correlated Channels

Now suppose the server tasks are correlated, in that the selection of one particular task may affect the selection of any or all of the other tasks. Thus the channels are a sequence of dependent Bernoulli random variables. Based on ideas from random graphs of Korzeniowski (2003), the construction of dependent Bernoulli random variables was given in Korzeniowski (2013), and is briefly summarized below.

4.1 Dependent Bernoulli Random Variables and the Generalized Binomial Distribution

Korzeniowski (2013) constructed a sequence of dependent Bernoulli random variables using a binary tree that distributes probability mass over dyadic partitions of [0,1]. Let $0 \leq \delta \leq 1$, $0 < p < 1$, and $q = 1 - p$. Then define the following quantities:

$$\begin{aligned} q^+ &:= q + \delta p & p^+ &:= p + \delta q \\ q^- &:= q(1-\delta) & p^- &:= p(1-\delta) \end{aligned} \tag{2}$$

The quantities in (2) satisfy the following conditions:

$$\begin{aligned} q^+ + p^- = q^- + p^+ = q + p = 1 \\ qq^+ + pq^- = q, \quad qp^- + pp^+ = 1 \end{aligned} \tag{3}$$

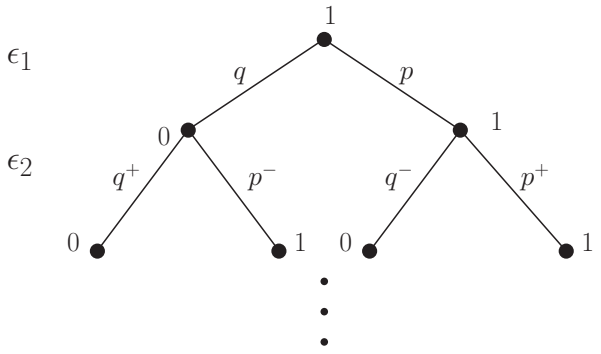


Figure 2. Construction of Dependent Bernoulli Random Variables

Figure 2 shows the construction of the dependent Bernoulli random variables. The following examples using coin flips illustrate the effect of the dependency coefficient δ:

Example 1 ($\delta = 1$)**.** *For $\delta = 1$, $q^+ = q + p = 1, q^- = 0$, $p^+ = p + q = 1$, and $p^- = 0$. Supposing the first coin flip $\varepsilon_1 = 1$. Then every successive ε_i will also be 1. Similarly if $\varepsilon_1 = 0$. Thus the result of the first coin flip completely determines the outcomes of all the rest.*

Example 2 ($\delta = 0$)**.** *For $\delta = 0$, $q^+ = q^- = q$, and $p^+ = p^- = p$. Thus, the first coin flip (and all subsequent ones) have no effect on the ones that follow.*

Example 3 ($\delta = \frac{1}{4}$)**.** *Suppose $p = q = \frac{1}{2}$. Then $p^+ = q^+ = \frac{5}{8}$, and $p^- = q^- = \frac{3}{8}$. Then the subsequent outcomes $\varepsilon_i, i \geq 2$ are more likely to match the outcomes of ε_1 than not.*

Now suppose $p = \frac{1}{4}, q = \frac{3}{4}$. Then $p^+ = \frac{7}{16}, p^- = \frac{3}{16}, q^+ = \frac{13}{16}$, and $q^- = \frac{9}{16}$. In this example of an unfair coin, the dependency coefficient δ still attempts to skew the results following the first coin flip in favor of the outcome of ε_1. However, the dependency here heightens the effect of $\varepsilon_1 = 0$ on subsequent flips, and cannot overcome the discrepancy between the probability of success and failure to skew $\varepsilon_i, i \geq 2$ in favor of a 1 following the outcome of $\varepsilon_1 = 1$.

Using these dependent Bernoulli random variables, Korzeniowski (2013) derived a Generalized Binomial Distribution for identically distributed but dependent Bernoulli random variables.

Generalized Binomial Distribution

Let $X = \sum_{i=1}^{n} \varepsilon_i$, where $\varepsilon_i, i = 1, ..., n$ are identically distributed Bernoulli random variables with probability of success p and dependency coefficient δ. Then

$$P(X = k) = q\binom{n-1}{k}(p^-)^k(q^+)^{n-1-k} + p\binom{n-1}{k-1}(p^+)^{k-1}(q^-)^{n-1-(k-1)}, \quad k = 0, 1, ..., n \tag{4}$$

4.2 Survival Function of Correlated Channels in a Cluster Server

Suppose the selection of tasks may be modeled by the dependent Bernoulli random variables given in the previous section. That is, suppose the customer selects Tasks 1,...,K in sequence, and the selection or rejection of Task 1 affects all subsequent tasks by a dependency coefficient δ. From Korzeniowski (2013), the correlation between task selections $\varepsilon_i, \varepsilon_j$ is given by

$$\rho = \text{Cor}(\varepsilon_i, \varepsilon_j) = \begin{cases} \delta, & i = 1; j = 2, ..., K \\ \delta^2, & i \neq j; i, j \geq 2 \end{cases} \tag{5}$$

This illustrates the dependency of Tasks 2,...,K on the outcome of Task 1, and notes that while Tasks 2,...,K are still correlated with each other, the dependency is much lower. In a similar fashion to the independent channel server, the survival function is derived.

Theorem 3. *Suppose conditions (i)-(vi) above are satisfied. In addition, suppose the selection of channels* $1, \ldots, K$ *are determined by identically distributed Bernoulli random variables with dependency coefficient* δ *as defined in Korzeniowski (2013). Then the survival function of the server is given by*

$$S_Y(t) = \bar{F}_0(t) \exp\left(-\eta \int_0^t m(t-w)\bar{G}_W(w)\mathscr{S}(w)dw\right) \tag{6}$$

where $m(x) = \int_0^x \lambda(s)ds$, *and*

$$\begin{aligned} \mathscr{S}(w) = &\sum_{n=0}^{K} e^{-\eta n w} \sum_{j=0}^{K-n-1} \binom{K-1}{n-1, j, K-1-n-j} p^{K-1-j}(1-p)^{j+1}\delta^{K-1-n-j}(1-\delta)^n \\ &+ \sum_{n=0}^{K} n e^{-\eta n w} \sum_{i=0}^{n-1} \binom{K-1}{K-1-n, i, n-1-i} p^{i+1}(1-p)^{K-n}\delta^{n-1-j}(1-\delta)^{K-n-j} \end{aligned}$$

Proof. By Theorem 1,

$$S_Y(t) = \bar{F}_0(t) \exp\left(-E\left[\mathcal{H} \int_0^t e^{-\mathcal{H}w} m(t-w)\bar{G}_W(w)dw\right]\right)$$

Similar to the proof of Theorem 2, $\mathcal{H} = \eta X$, where this time X has the generalized binomial distribution given in (4). Then

$$\begin{aligned} &E\left[\mathcal{H} \int_0^t e^{-\mathcal{H}w} m(t-w)\bar{G}_W(w)dw\right] \\ &\quad = \sum_{x=0}^{K} \left[\eta x \int_0^t e^{-\eta x w} m(t-w)\bar{G}_W(w)dw\right] P(X = x) \\ &\quad = \sum_{x=0}^{K} \eta x \left[\int_0^t e^{-\eta x w} m(t-w)\bar{G}_W(w)dw\right]\left[q\binom{K-1}{x}(p^-)^x(q^+)^{K-1-x}\right] \\ &\qquad + \sum_{x=0}^{K} \eta x \left[\int_0^t e^{-\eta x w} m(t-w)\bar{G}_W(w)dw\right]\left[p\binom{K-1}{x-1}(p^+)^{x-1}(q^-)^{K-x}\right] \\ &\quad = \eta \int_0^t m(t-w)\bar{G}_W(w)(\mathscr{S}_1(w) + \mathscr{S}_2(w))dw \end{aligned}$$

where $\mathscr{S}_1(w) = \sum_{x=0}^{K} x e^{-\eta x w} q\binom{K-1}{x}(p^-)^x(q^+)^{K-1-x}$
and $\mathscr{S}_2(w) = \sum_{x=0}^{K} x e^{-\eta x w} p\binom{K-1}{x-1}(p^+)^{x-1}(q^-)^{K-x}$. Using the definitions given in (2),

$$\begin{aligned} \mathscr{S}_1(w) &= \sum_{x=0}^{K} x e^{-\eta x w}(1-p)\binom{K-1}{x}(p-\delta p)^x(1-p+\delta p)^{K-1-n} \\ &= \sum_{x=0}^{K} x e^{-\eta x w}(1-p)\binom{K-1}{x}p^x(1-\delta)^x \sum_{j=0}^{K-1-x}\binom{K-1-x}{j}(1-p)^j(\delta p)^{K-1-x-j} \end{aligned}$$

Now, $x\binom{K-1}{x}\binom{K-1-x}{j} = \frac{(K-1)!}{(x-1)!j!(K-1-x-j)!} = \binom{K-1}{x-1,j,K-1-x-j}$. Then

$$\mathscr{S}_1(w) = \sum_{x=0}^{K} e^{-\eta x w} \sum_{j=0}^{K-x-1} \binom{K-1}{x-1, j, K-1-x-j} (1-p)^{j+1}(1-\delta)^x \delta^{K-1-x-j} p^{K-1-j}$$

Similarly,

$$\begin{aligned}
\mathscr{S}_2(w) &= \sum_{x=0}^{K} x e^{-\eta x w} p \binom{K-1}{x-1} (p + \delta(1-p))^{x-1} ((1-p)(1-\delta))^{K-x} \\
&= \sum_{x=0}^{K} x e^{-\eta x w} p (1-\delta)^{K-x} (1-p)^{K-x} \sum_{i=0}^{x-1} \binom{x-1}{i} p^i (1-\delta)^i \delta^{x-1-i} \\
&= \sum_{x=0}^{K} x e^{-\eta x w} \sum_{i=0}^{x-1} x \binom{K-1}{K-1-x, i, x-1-i} p^{i+1} \delta^{x-1-i} (1-\delta)^{K-x+i} (1-p)^{K-x}
\end{aligned}$$

Clearly $\mathscr{S}(w) = \mathscr{S}_1(w) + \mathscr{S}_2(w)$ □

5. Numerical Illustrations

5.1 Expected Server Lifetime

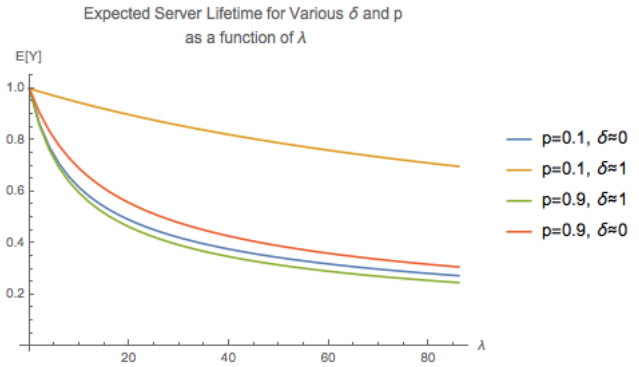


Figure 3. Expected Server Lifetime for Various Selection Probabilities and Dependency Coefficients

To measure the effects of the probability of task selection p and the dependency coefficient δ, we look at the expected server lifetime, given by $E[Y] = \int_0^\infty S_Y(t)$ as a function of the arrival rate λ. Figure 3 shows the expected server lifetime for $K = 3$ channels under two different selection probabilities ($p = 0.1, 0.9$) and dependency coefficients ($\delta = 0, 1$). In addition, $\eta = r_0 = 1$. Mathematica was used for calculations.

For $\delta = 0$, $p = p^+ = p^-$ and $q = q^+ = q^-$, and hence the channels are uncorrelated and the selection of tasks $1, ..., K$ are independent Bernoulli random variables. When $\delta = 1$, the selection decision made at channel 1 completely determines the subsequent task selections. Thus, when $\delta = 1$ and $p = 0.1$, it is highly likely that the first task will not be selected and thus no others will be selected. This results in a server that is expected to remain fairly idle, even as the arrival intensity increases. Thus, the expected lifetime changes very little compared to a completely idle server ($\lambda = 0$).

On the opposite end, for $p = 0.9, \delta = 1$, we again have a perfectly correlated set of channels, but the selection probability for Task 1 is very high. Thus, 0.9λ customers per unit time will select all tasks, and 0.1λ customers per unit time will select no tasks. This will result in a high server workload, and the expected lifetime decreases sharply with increasing λ.

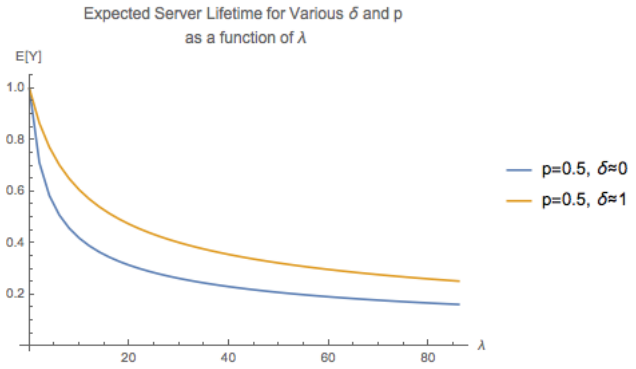


Figure 4. Expected Server Lifetime for $p = 0.5$ and Various Dependency Coefficients

Figure 4 compares the effect of the dependency coefficient δ when the selection probability of Task 1 is $p = 0.5$. A completely correlated system with $p = 0.5$ has a significantly longer expected lifetime for a given arrival rate λ than independent channels.

6. Conclusion

This paper provides a reliability model for a multichannel clustered-task server under very general assumptions. Requests arrive via a nonhomogenous Poisson process and consist of the random selection of 0 to K tasks in sequence. A request is considered serviced when the last task in that request is completed. Since each channel has a generic service distribution $G_i(w)$, the service distribution of the requests is given by $\max_i G_i(w)$. The stress to the server brought by each request is a constant multiple η of the number of tasks selected, and thus is random with either a binomial distribution or a generalized binomial distribution. The survival function in both cases was formulated, and a numerical comparison of the correlated and uncorrelated cases was done using the expected server lifetime as a metric.

References

Cha, K. H., & Lee, E. Y. (2011). A Stochastic Breakdown Model for an Unreliable Web Server System and an Optimal Admission Control Policy. *J. Appl. Probab., 48*(2), 453-466. http://dx.doi.org/10.1239/jap/1308662638.

Gross, D., & Shortle, J. (2008). *Fundamentals of Queuing Theory*. John Wiley and Sons, Hoboken, New Jersey.

Korzeniowski, A. (2013). On Correlated Random Graphs. *Journal of Probability and Statistical Science, 11*. 43-58.

Korzeniowski, A. (2003). On Universal Representations of Random Graphs. *Annals of Combinatorics, 7*. 299-313.

Tang, Y. H. (1997).A single-server M/G/1 queueing system subject to breakdownssome reliability and queueing problems. *Microelectronics Reliability, 37*. http://dx.doi.org/10.1016/S0026-2714(96)00018-2

Traylor, R. (2015). *A Stochastic Reliability Model of a Server under a Random Workload*. Preprint. Available at http://arxiv.org/abs/1511.04130

Traylor, R. (2015). *Reliability of Series and Parallel Systems with Correlated Nodes*. Preprint. Available at http://arxiv.org/abs/1601.03630

Copyrights

International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

The Probability of Pathogenicity in Clinical Genetic Testing: A Solution for the Variant of Uncertain Significance

Michael Gollob[1], Jeffrey S. Rosenthal[2] & Kevin Thorpe[2]

[1] Peter Munk Cardiac Centre

[2] University of Toronto

Correspondence: Jeffrey S. Rosenthal. E-mail: jeff@math.toronto.edu

Received: April 4, 2016 Accepted: April 18, 2016 Online Published: June 11, 2016

doi:10.5539/ijsp.v5n4p52 URL: http://dx.doi.org/10.5539/ijsp.v5n4p52

Abstract

We present a direct calculation for determining the probability that a rare genetic variant is the cause of an observed disease, under appropriate assumptions, in terms of the joint prevalence of the disease and of rare variants. Our calculation provides a resolution of the so-called "variant of unknown (or uncertain) significance" problem, which has plagued medical genetics researchers.

Keywords: genetic variant, rare variant, pathogenicity, variant of unknown significance, genetic probability

1. Introduction

It frequently arises in medical genetics that a patient has a particular disease, and also has a genetic variant which may or may not be the cause of that disease. The variant is said to be of *unknown* or *uncertain significance* if its disease-causing probability cannot be determined, and this is a common challenge (see e.g. Richards et al. (2008), Cheon, Mozersky, and Cook-Deegan (2014), Domcheck and Weber (2008), and the references therein). The problem is put succinctly in the patient guide by Ambry Genetics (2015), which states, "Variants of unknown significance are DNA changes with too little information known to classify as either pathogenic or benign, and it is unknown whether they contribute to a medical condition."

Assessing whether or not the genetic variant did indeed cause the disease is important not only for future medical research and prevention, but also as a practical guide to whether or not the patient's family members are also at risk. However, in the case of a rare variant with little or no previous information available, it is unclear how this assessment should be made.

In this paper, we present a direct calculation for determining the probability that a rare genetic variant is indeed the cause of an observed disease. Our calculation requires one assumption, namely the natural-seeming "Variant Fraction Assumption" that in the absence of any other evidence, the probability that a newly observed rare genetic variant of a gene causes a specified disease is equal to the fraction of all previously-observed rare variants of that gene which did cause that disease (see Section 4.3 for details). With this one assumption, we are able to compute the desired probability purely in terms of the disease's prevalence in the population, and the fraction of rare variants in the general population and in the diseased population. Our results are described below.

2. Formal Set-Up

To set up our probability model, we make the following assumptions and notations:

- A certain disease D has a known prevalence p in the general population.
- Among the healthy population, a certain known fraction q have some distinct rare variant of a certain gene G.
- Among patients with the disease D, a certain fraction $r \geq q$ have some distinct rare variant of the gene G.
- Some subset S of the rare variants of G always cause the disease D, i.e. the probability of disease given a rare variant in S is 1.
- Rare variants of G which are *not* in S have no effect on D, i.e. the probability of disease given a rare variant *not* in S is the same as for patients without a rare variant.
- A new patient X is found to have a never-before-seen variant V of the gene G.

The question is, given all of the above, if X gets the disease D, then what is the probability that the genetic variant V was actually a *cause* of the disease D in X? That is, we wish to compute the conditional probability

$$P(V \in S \mid X \text{ has the disease } D)\,,$$

i.e. the probability that X's variant V is in fact a *cause* of the disease D in X, given that X has the variant V and also has the disease D. We describe our calculation of this and related probabilities below.

We first note that since the genetic variant V has never been seen before, it is impossible compute its probabilities without making *some* additional assumption. Thus, we also make the *Variant Fraction Assumption* that in the absence of any other evidence, the probability that a newly observed genetic variant of a gene G is a cause of D, is equal to the fraction of *all* of the previously-observed rare variants of G which were indeed found to cause D. For a more precise statement of this assumption, see Section 4.3 below.

Remark. In fact, the assumption that $r \geq q$ follows directly from the other assumptions; see Appendix A2.

3. Main Result

Our main results are as follows. (Below, "unconditional probability" means the probability without conditioning on whether or not X has the disease D. Also, we write "V is the sole cause of D" to mean that V caused D, and furthermore patient X would not have gotten D in the hypothetical case that they instead did *not* have the rare variant V.)

Theorem 1. *For the above set-up, for a patient X having a rare variant V of gene G, under the Variant Fraction Assumption, we have the following probabilities, where* $y = (r-q)/(1-q)$, $z = p(1-y)/(1-py)$, $w = py\,/\,[pr + (1-p)q]$, *and* $u = (1-z)w$.
(a) The unconditional probability that the variant V is a cause of the disease D is given by:

$$\mathbf{P}(V \in S) \;=\; w\,.$$

(b) The unconditional probability that patient X will get disease D is given by:

$$\mathbf{P}(X \text{ gets } D) \;=\; z \,+\, u\,.$$

(c) Conditional on patient X getting the disease D, the conditional probability that the variant V was the sole *cause of D is given by:*

$$\mathbf{P}(\text{X's disease D was caused solely by V} \mid \text{X has D}) \;=\; \frac{\text{u}}{\text{z} + \text{u}}\,.$$

(d) Conditional on patient X getting the disease D, the conditional probability that the variant V is a cause of D is given by:

$$\mathbf{P}(V \in S \mid X \text{ gets } D) \;=\; \frac{w}{z+u} \;=\; \frac{r-q}{r(1-q)}\,.$$

Theorem 1 thus gives precise probabilities for the relevant possibilities related to patient X and disease D and rare genetic variant V. In particular, Theorem 1(d) gives a precise estimate of the probability, given that patient X has disease D and has the rare genetic variant V, that V is in fact a cause of the disease D.

Theorem 1 is proved in the next section. We first consider some numerical examples.

For example, if $p = 1/4,000$ is the prevalence of the disease in the general population, and $q = 2\% = 0.02$ is prevalence of *some* rare variant of G in the healthy population, and $r = 40\% = 0.4$ is the prevalence of *some* rare variant of G in patients with the disease, then conditional on X getting the disease, the probability that the variant V is a cause of the disease works out to: $\mathbf{P}(V \in S \mid X \text{ gets } D) = 0.9699815 \doteq 97.0\%$, or about 97 percent (i.e., nearly certain).

Or, if $p = 1/400$ and $q = 0.01$ and $r = 0.4$, then $\mathbf{P}(V \in S \mid X \text{ gets } D) \doteq 98.5\%$.

By contrast, if $p = 1/50$ and $q = 0.1$ and $r = 0.2$, then $\mathbf{P}(V \in S \mid X \text{ gets } D) \doteq 58.1\%$, which is much smaller.

Or, if $p = 1/400$ and $q = 0.1$ and $r = 0.15$, then $\mathbf{P}(V \in S \mid X \text{ gets } D) \doteq 39.5\%$.

A plot of other values of $\mathbf{P}(V \in S \mid X \text{ gets } D)$ is presented in Figure 1 as a function of r, with $q = 0.1$ fixed.

Probabilities for other parameter values can be computed using the above formulae, or using our simple javascript online calculator available at: `www.probability.ca/pathprob`

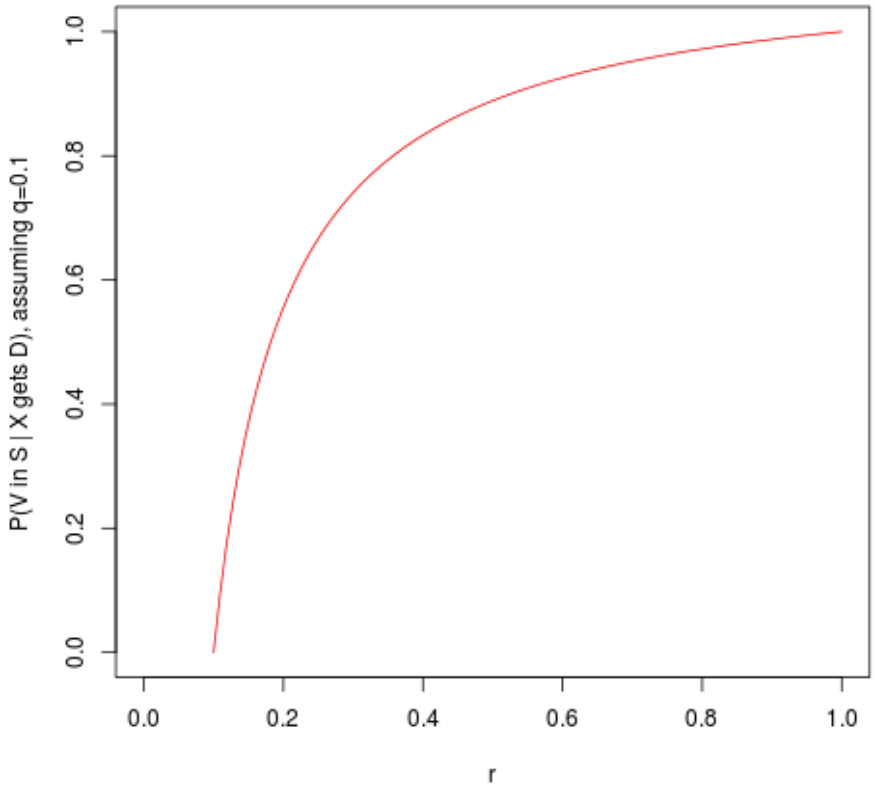


Figure 1. Disease cause probabilities as a function of r, with $q = 0.1$ fixed.

As a further check, we note that the formula in Theorem 1(d) gives answers which make sense even for certain extreme parameter values. For example, if $r = 1$ (i.e., 100%), then it is computed that the formula gives a value of 1. This makes sense, since if $r = 1$, then by the definition of r, *every* diseased patient has a rare variant of G, which means that the disease can *only* be caused by a rare variant of G, so that V must indeed cause D.

Or, if $q = 0$, then again it is computed that the formula gives a value of 1. This also makes sense, since if $q = 0$, then by the definition of q, *no* healthy patients have rare variants of G, which means that rare variants of G *always* cause D, so again V must indeed cause D.

By contrast, if $r = q$, then it is computed that the formula gives a value of 0. This again makes sense, since if the rate of rare variants of G is the same for diseased and healthy patients, then the rare variants of G do *not* cause any additional disease at all, so none of them cause D.

As a final comment, we note that since $u = (1 - z)w$, the formula in Theorem 1(c) is smaller than that in Theorem 1(d) by a factor of $(1 - z)$. This is due to the possibility that $V \in S$ but X still "would" have gotten D by chance alone, i.e. that V does indeed cause D, but X would have gotten D even in the absence of V (e.g. from a variant of some other gene besides G). Now, usually z will be very small, so the answers in (c) and (d) will be very similar, though not identical.

4. Proof of Theorem 1

In this section, we prove Theorem 1 using a sequence of probability calculations. We break up our argument into several steps.

4.1 Preliminary Population Prevalence Calculations

We first note that our assumption above that rare variants of G which are *not* in S have no effect on D, can be written more formally as

$$P(X \text{ has } D \mid X \text{ has a rare variant } V \notin S) \;=\; P(X \text{ has } D \mid X \text{ has no rare variant}).$$

From this it follows (see Appendix A1) that

$$P(X \text{ has rare variant } V \notin S \mid X \text{ has } D, \text{ and no variant in } S) = P(X \text{ has some rare variant} \mid X \text{ does not have } D).$$

We next calculate two population fractions that will be important in our solution.

First, we write y for the fraction of diseased patients who have a rare variant of G which is in S, i.e. which does cause D. (Thus, y is close to r, but slightly less since even among diseased patients without a variant in S, a fraction q of them will still happen to have a rare variant not in S by chance alone, just like for the healthy population.)

In terms of y, the set of all diseased patients can be divided into three groups: a fraction y with a rare variant of G which is in S, a fraction $(1-y)q$ with a rare variant of G which is *not* in S, and a fraction $(1-y)(1-q)$ with no rare variant of G at all. Hence, the fraction of diseased patients with *some* rare variant of G is equal to $y + (1-y)q$.

On the other hand, we know that the overall fraction of diseased patients with some rare variant of G is equal to r. For this to hold, we must have $y + (1-y)q = r$. Solving for y, we obtain that $y = (r-q)/(1-q)$.

Then, since a fraction p of the population is diseased, and a fraction y of diseased patients have a variant in S, it follows that the fraction of the total population who have a rare variant in S is equal to the product py. And, the fraction who do *not* have a rare variant in S is equal to $1 - py$.

Next, we write z for the prevalence of the disease D among all people who specifically do *not* have a variant in S. Then the fraction of people who have D but do *not* have a variant in S is equal to $(1-py)z$. And, the fraction of people with a variant in S (who therefore have D) is equal to py. So, the total fraction of the population who have the disease D is equal to $(1-py)z + py$.

On the other hand, we know that the overall prevalence of the disease is equal to p. For this to hold, we must have $p = py + (1-py)z$. Solving for z, we compute that $z = p(1-y)/(1-py)$.

4.2 The Prior Probability of Disease

Prior to diagnosis, what was the prior probability that a given patient X, who has some rare variant V, would get the disease D? That is, what is the conditional probability that X gets D, conditioning (throughout) on the fact that X has a rare variant V?

To answer this question, let I be the indicator function of the subset S, so I(V)=1 if V $\in$ S, otherwise I(V)=0. We know that the prior probability of X getting the disease D is equal to 1 (i.e., 100%) if I(V)=1, or is equal to z if I(V)=0. So, if we *knew* I(V), i.e. if we *knew* whether or not V causes D, then could write this as:

$$\mathbf{P}(\text{X gets D} \,|\, \text{I(V)}) \;=\; z \,+\, (1-z) \times I(V)\,.$$

In fact we do not know I(V), i.e. it could equal either 1 or 0. So, instead, we use the Law of Total Expectation (that the expected value of a conditional probability is the unconditional probability). This shows that:

$$\begin{aligned}\mathbf{P}(X \text{ gets } D) \;&=\; \mathbf{E}\Big[\mathbf{P}(\text{X gets D} \,|\, \text{I(V)})\Big] \;=\; z \,+\, (1-z) \times \mathbf{E}[I(V)] \\ &=\; z \,+\, (1-z) \times \mathbf{P}[I(V)=1] \;=\; z \,+\, (1-z) \times \mathbf{P}(V \in S)\,.\end{aligned}$$

This gives a formula for the prior probability that X would get D, in terms of the probability $\mathbf{P}(V \in S)$ that V is in S. However, we do not know $\mathbf{P}(V \in S)$, so it must be estimated.

4.3 The Variant Fraction Assumption

To continue, we need to obtain an estimate for $\mathbf{P}(V \in S)$, the unconditional probability (in the absence of any other evidence) that the newly observed variant V of G is in fact a cause of the disease D. Now, since V was never before seen, there is no way to directly calculate this probability. Instead, as mentioned above, we use the *Variant Fraction Assumption* that in the absence of any other evidence, the probability that a newly observed genetic variant of a gene G is a cause of D, is equal to the fraction of *all* of the previously-observed rare variants of G which were indeed found to cause D. That is, we assume that

$$\mathbf{P}(V \in S) \;=\; \frac{\text{fraction of the population with a disease-causing rare variant of G}}{\text{fraction of the population with any rare variant of G}}\,.$$

This assumption appears to be quite reasonable, in the absence of any other prior information about the new variant V. In any case, some such assumption must be made, otherwise no probabilities associated with V can possibly be computed. But under this one assumption, all of the remaining probability calculations can be completed.

4.4 Estimating the Variant Probability

Using the above Variant Fraction Assumption, we are able to compute the desired probability $\mathbf{P}(V \in S)$. To do this, we need to compute the fraction of *all* of G's rare variants which are in S. We proceed as follows.

Since a fraction p of the population is diseased, and since a fraction r of them have some rare variant of G, it follows that the fraction of the population who are diseased *and* have some rare variant of G is pr. Similarly, since y is the fraction of diseased patients who have a variant in S, it follows that the fraction of the population who are diseased and have a rare variant in S is py. Also, the fraction of the population who are *healthy* and have *some* rare variant of G is $(1-p)q$.

That is, a fraction $pr+(1-p)q$ of the population has a rare variant, and a fraction py of the population has a rare variant in S. So, assuming these rare variants are all *distinct*, the fraction of all the rare variants of G in the entire population which are in S is equal to $py/[pr+(1-p)q]$.

We then estimate the probability $\mathbf{P}(V \in S)$ by the above fraction of all the rare variants of G which are in S, i.e. by

$$\mathbf{P}(V \in S) \;\approx\; py\,/\,[pr+(1-p)q] \;=:\; w$$

where $w = py\,/\,[pr+(1-p)q]$, as claimed in Theorem 1(a).

Now, since we earlier derived the prior-to-diagnosis probability $\mathbf{P}(\text{X gets D}) = z+(1-z)\times\mathbf{P}(V \in S)$, it then follows from the above estimate that

$$\mathbf{P}(X \text{ gets } D) \;=\; z \;+\; (1-z)\times w \;=:\; z \;+\; u$$

where $u = (1-z)w$, as claimed in Theorem 1(b).

4.5 The Cause Probability

The above formula for $\mathbf{P}(\text{X gets D})$ can be interpreted as follows: Patient X can get the disease D either without any influence at all from the gene G (with probability z), or caused by the variant V of G (with probability f).

Under this interpretation, given that X *does* in fact have the disease D, the conditional probability that the disease D in X was *caused* by the genetic variant V, and *would not otherwise have arisen*, is given by the second probability divided by the sum of the two probabilities, i.e. by:

$$P(\text{X's disease } D \text{ was caused solely by } V \mid X \text{ has } D) \;=\; \frac{u}{z+u}\,,$$

as claimed in Theorem 1(c).

This formula thus gives an estimate of the probability, given that patient X has disease D, that the disease was caused solely by their rare genetic variant V of the gene G. This is similar to, but not quite the same as, our desired conditional probability, as we now explain.

4.6 Computing the Conditional Probability

Putting the previous equations together, we can compute the required conditional probability, as follows:

$$\mathbf{P}(V \in S \mid X \text{ gets } D) \;=\; \frac{\mathbf{P}(V \in S,\ \text{and } X \text{ gets } D)}{\mathbf{P}(X \text{ gets } D)}$$

$$=\; \frac{\mathbf{P}(V \in S)}{\mathbf{P}(X \text{ gets } D)} \;=\; \frac{w}{z+u}\,.$$

This gives our first formula claimed in Theorem 1(d).

Finally, through careful algebraic simplification, it is verified directly that in fact $\frac{w}{z+u} = \frac{r-q}{r(1-q)}$ (which, surprisingly, does not depend on p), thus giving the second formula claimed in Theorem 1(d).

Remark. The above equations can be combined as follows. We have that

$$\frac{u}{z+u} \;= \mathbf{P}(\text{ D was caused solely by V} \mid \text{X gets D})$$

$$=\; \mathbf{P}(\text{D was caused solely by V})\,/\,\text{P(X gets D)}$$

$$= \mathbf{P}(\text{D was caused solely by V} \mid \text{V in } S)\; P(\text{V in S} \mid X \text{ gets D})$$

$$= \mathbf{P}(\text{D was caused solely by V} \mid \text{V in } S)\;\frac{w}{z+u}\,,$$

whence

$$\mathbf{P}(\text{D was caused solely by V} \mid \text{V in S}) \;=\; \frac{u}{w} \;=\; \frac{(1-z)w}{w} \;=\; 1-z\,.$$

Taking the complementary event,

$$\mathbf{P}(\text{X would have gotten D even without having V} \mid \text{V in S}) \;=\; z\,,$$

which makes sense since z is the prevalence of the disease in people who do *not* have a rare variant, corresponding to the appropriate hypothetical.

Final Remark. After originally completing this research, we became aware of the related recent paper by Ruklisa, Ware, Walsh, Balding, and Cook (2015). That paper mostly focuses on specific probability estimates for specific diseases and specific genetic variants. However, it does begin with what they call the "prior odds of pathogenicity", which appears to correspond to odds for the event $\mathbf{P}(V \in S \mid X \text{ gets } D)$ considered in Theorem 1(d) above. For that case, they assert that the odds "might be assumed to be" given by

$$\frac{\text{Burden of rare variants in cases} - \text{Burden of rare variants in controls}}{\text{Burden of rare variants in controls}}\,.$$

In our notation, this appears to correspond to asserting that

$$\frac{\mathbf{P}(V \in S \mid X \text{ gets } D)}{1 - \mathbf{P}(V \in S \mid X \text{ gets } D)} \;=\; \frac{r-q}{q}\,,$$

or equivalently that

$$\mathbf{P}(V \in S \mid X \text{ gets } D) \;=\; \frac{\frac{r-q}{q}}{1+\frac{r-q}{q}} \;=\; \frac{r-q}{r}\,.$$

This last expression is fairly similar to the final formula in our Theorem 1(d), but it differs by a factor of $1-q$. It appears that the reason for this discrepency is their assertion that "it is reasonable to assume that the burden of benign rare variants in cases is equal to the burden of rare variants in controls", which apparently corresponds to assuming that

$$\mathbf{P}(\text{Rare variant not in S} \mid \text{Diseased}) \;=\; \mathbf{P}(\text{Rare variant not in S} \mid \text{Not diseased})\,,$$

which is different from what we believe (for more see Appendix A3).

Appendix: Additional Probability Calculations

We here provide a few additional probability calculations, to further clarify some of the material in the main text. For these calculations, define a through f to be the proportions of the total population in each of the six categories implied by the following Table:

	Diseased	Not diseased
Rare variant in S	a	d
Rare variant not in S	b	e
No rare variant	c	f

A1. We first show that the assumption

$$\mathbf{P}(\text{X has D} \mid \text{X has a rare variant V} \notin \text{S}) \;=\; \mathbf{P}(\text{X has D} \mid \text{X has no rare variant}) \qquad (*)\,.$$

implies the condition that

$$\mathbf{P}(\text{X has rare variant V} \notin \text{S} \mid \text{X has D, and no variant in S}) = \mathbf{P}(\text{X has some rare variant} \mid \text{X does not have D})\,. \qquad (**)$$

In terms of the above Table, condition $(*)$ is equivalent to saying that $b/(b+e) = c/(c+f)$, i.e. that $b/e = c/f$. Meanwhile, condition $(**)$ is equivalent to saying that $b/(b+c) = (d+e)/(d+e+f)$. On the other hand, our assumption that variants in S always cause the disease D implies that $d = 0$. Hence, if $b/e = c/f$, then $(d+e)/(d+e+f) = e/(e+f) = b/(b+c)$, thus establishing $(**)$.

A2. We here show that the assumption $r \geq q$ actually follows from our other assumptions. Indeed, in terms of the above Table, using as above that $d = 0$ and $e/(e + f) = b/(b + c)$, we have that

$$q = \mathbf{P}(\text{Rare variant} \mid \text{Not Diseased}) = \frac{d+e}{d+e+f} = \frac{e}{e+f} = \frac{b}{b+c}.$$

Also

$$r = \mathbf{P}(\text{Rare variant} \mid \text{Diseased}) = \frac{a+b}{a+b+c}.$$

The result then follows since

$$r - q = \frac{a+b}{a+b+c} - \frac{b}{b+c} = \frac{(ab+ac+b^2+bc)-(ab+b^2+bc)}{(a+b+c)(b+c)} = \frac{ac}{(a+b+c)(b+c)} \geq 0.$$

A3. Finally, we consider the quantities which arise in the reference Ruklisa et al. (2015), as discussed in our Final Remark above, namely

$$\mathbf{P}(\text{Rare variant not in S} \mid \text{Diseased}) \qquad (\&)$$

and

$$\mathbf{P}(\text{Rare variant not in S} \mid \text{Not diseased}). \qquad (\&\&)$$

In our notation, $(\&\&) = \mathbf{P}(\text{Rare variant not in S} \mid \text{Not diseased}) = q$ which, in terms of the above Table, is equal to $e/(e + f)$. By contrast, $(\&) = \mathbf{P}(\text{Rare variant not in S} \mid \text{Diseased}) = b/(a + b + c)$. Hence, assuming $(*)$, we have $(\&) = b/(a + b + c) \leq b/(b + c) = e/(e + f) = (\&\&)$. Furthermore, we have non-zero equality only when $a = 0$, which corresponds to no variants in S, i.e. to the gene G having no effect whatsoever on the disease D. Otherwise, $a > 0$, and hence $(\&) < (\&\&)$ (assuming $b > 0$), leading to a different result from that of Ruklisa et al.

References

Ambry, G. (2015). Your genetic test result. Available at: http://www.ambrygen.com/sites/default/files/General%20VUS%20patient%20brochure%20Updates_0.pdf

Cheon, J. Y., Mozersky, J., & Cook-Deegan, R. (2014). Variants of uncertain significance in BRCA: a harbinger of ethical and policy issues to come? *Genome Medicine, 6*(121), 1–10. http://dx.doi.org/10.1186/s13073-014-0121-3

Domchek, S., & Weber, B. L. (2008). Genetic variants of uncertain significance: flies in the ointment. *Journal of Clinical Oncology, 26*(1), 16–17. http://dx.doi.org/10.1200/JCO.2007.14.4154

Richards, C. S., Bale, S., Bellissimo, D. B., Das, S., Grody, W. W., Hegde, M. R., Lyon, E., & Ward, B. E. (2008). ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genetics in Medicine, 10*(4), 294–300. http://dx.doi.org/10.1097/GIM.0b013e31816b5cae

Ruklisa, D., Ware, J. S., Walsh, R., Balding, D. J., & Cook, S. A. (2015). Bayesian models for syndrome- and gene-specific probabilities of novel variant pathogenicity. *Genome Medicine, 7*(5), 1–16. http://dx.doi.org/10.1186/s13073-014-0120-4

Copyrights

International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

On the Dichotomy between the Psychological or Empirical Aspect of Subjective Probability and the Logical or Geometrical One

Pierpaolo Angelini[1] & Antonio Maturo[2]

[1] ITA "Emilio Sereni", Roma.

[2] Università "Gabriele D'Annunzio", Chieti-Pescara.

Correspondence: Prof. Pierpaolo Angelini, ITA "E. Sereni", via Prenestina n. 1395, Roma. Tel: 39-340-9099-992.

E-mail: pp.angelini@virgilio.it

Received: April 12, 2016 Accepted: May 11, 2016 Online Published: June 11, 2016

doi:10.5539/ijsp.v5n4p59 URL: http://dx.doi.org/10.5539/ijsp.v5n4p59

Abstract

In the domain of the logic of certainty we examine the objective notions of the subjective probability with the clear aim of identifying their fundamental characteristics before the assignment, by the individual, of the probabilistic evaluation. Probability is an additional and subjective notion that one applies within the range of possibility, thus giving rise to those gradations, more or less probable, that are meaningless in the logic of certainty. Each criterion for evaluations under conditions of uncertainty is a device or instrument for obtaining a measurement; it furnishes an operational definition of probability or prevision **P** and together with the corresponding conditions of coherence can be taken as a foundation for the entire theory of probability. When we examine these criteria and their corresponding conditions of coherence we show the inevitable dichotomy between the subjective or psychological or empirical aspect of probability and the objective or logical or geometrical one.

Keywords: random entity, coherence, convex hull, barycenter, vector space, metric.

1. Introduction

According to the subjectivistic conception of probability, the concept of probability and the foundations of probability theory have a psychological value: such a theory is rigidly deduced and reconstructed on the sole basis of a psychological interpretation and formulation because the mathematical principles are always the same from whatever point of view one starts (de Finetti, 1931b). In analyzing the objective meaning of the notion of coherence, it is necessary to point out in which way some probability evaluations may be incoherent or intrinsically contradictory and the rules of probabilistic logic, as those of formal logic in the field of propositions, are essential in order to teach us how to reason in the field of probability evaluations. Since all probability evaluations have and can have only an essentially and exclusively psychological value, it is necessary to separate what in a problem is logical from what is essentially of a merely empirical value and nature. So, one will be able to say whether every other problem is logically determined or undetermined (de Finetti, 1930a, 1930b). Evidently, this separation is fundamental in order to be able to deepen the criticism of principles of any mathematical theory and, in particular, of probability theory (de Finetti, 1931a). Probability, as an individual's psychological perception, is subject to certain laws. If an event has an objective probability, all the individuals who will conform their psychological position to it, can be said to be judging correctly, while the others to be wrong. Apart from this, the laws are the same for everybody and, in particular, hold for objective probabilities, so it is not true that if everything is subjective, everything must be arbitrary and no law can be valid. Essentially, it needs to characterize the whole of the formally admissible opinions, without bothering if reasons exist of any other type which might cause someone to consider any one of them more or less right. In fact, such reasons are beyond the merely logical or objective aspect of the problem which only mathematics can and must deal with: thus a clear separation of the two phases, the formal phase and the practical or empirical phase, appears appropriate and inevitable. The formal phase, that is to say, the characterization of the not incoherent opinions, is to be dealt with mathematically; the practical or empirical phase, that is to say, the choice of one among such possible opinions, has to be left to good sense and judgement of every single individual. The only difference between those who follow the subjectivistic conception and those who follow the objectivistic one is that while such a choice is free and arbitrary for the former, it can be right in only one way for the latter. Therefore, the subjectivistic approach takes into consideration, along with the objectively right evaluation of probability, all those evaluations which are not contradictory by themselves, although wrong according to the objectivistic viewpoint. A person who does not share the subjectivistic

viewpoint believes in the existence of an objective value of probability which cannot be maintained except for a certain more or less limited field; it would only be to events of a certain type, more or less schematic and artificial, that he will assign an objective probability, while in practical life he would be incessantly guided to think or say that a certain event appears more or less easily, is more or less probable or verisimilar and on such judgements he will found his decisions, also in areas which, according to his way of thinking, would be precluded from probability theory. Clearly, in order to justify conceptually such judgements, it needs to conform to the subjectivistic viewpoint whose validity field is not subject to any restriction (de Finetti & Minisola, 1961; de Finetti & Emanuelli, 1967a; de Finetti, 1955, 1963, 1969, 1970).

2. Logic of Certainty

When a given individual, according to his state of information, defines a set more or less large of possible alternatives, of which one and only one is necessarily true, he finds himself into the domain of the logic of certainty. We denote by $\mathcal{S}$ the abstract space of alternatives and by $\mathcal{Q}$, subset of $\mathcal{S}$, the space of the only alternatives possible for a certain individual; in fact, it may be convenient to think of $\mathcal{Q}$ as embedded in a larger and more manageable space $\mathcal{S}$. However, his information as well as his knowledge could also allow him to eliminate a part of the alternatives that can be imagined because he believes that they are impossible; vice versa, all the others will be possible. After all, a rather crude analysis can be made if all the possible alternatives are collected in order to obtain an unique and certain alternative. The possibility, unlike probability, has no gradations, thus the domain of the logic of certainty is objective; it is equally possible, for a given individual at a certain time, that the next FIFA world cup is won by a very weak national football team, that the next President of the Italian Republic is a woman, that the unemployment rate falls by three percentage points at the end of next year in Italy. Into the domain of the logic of certainty, only true and false exist as final and certain answers and certain and impossible and possible as options with regard to the temporary knowledge of any individual; into this domain we study the objective notions of subjective probability with the clear aim of identifying their fundamental characteristics before the assignment, by the individual, of the probabilistic evaluation. Probability is an additional and subjective notion that one applies within the range of possibility, thus giving rise to those gradations, more or less probable, that are meaningless in the logic of certainty. The field of the logic of certainty is objective because the elements of $\mathcal{Q}$ do not depend on the individual's opinions but only on his degree of ignorance (de Finetti, 1967b, 1970).

3. Events and Random Numbers

An event E is a statement which we do not know yet to be true or false; the event which is certain and the one which is impossible can be taken as a limit case. The statements of which we can say if they are true or false on the basis of an ascertainment well determined and always possible, at least conceptually, have objective meaning. Such objective statements are said propositions if one is thinking more in terms of the expressions in which they are formulated or, equally, events if one is thinking more in terms of the situations and circumstances to which their being true or false corresponds (de Finetti 1954). For any individual who does not know with certainty the value of a number X, which is random in a non-redundant usage for him, there are two or more than two, a finite or infinite number, possible values for X, where the set of these values is I(X): in any case, only one is the true value of each random number (de Finetti, 1970).

Remark 1 Events are also questions whose wordings, unambiguous and exhaustive, have the aim of removing any opportunity to complaining in case that a bet is based upon them: they admit two answers, yes = 1 or no = 0, true = 1 or false = 0, where such answers are always alternative. Also the random numbers can be identified by questions whose wordings are indisputably clear and complete; unlike events, they contain two or more than two answers which consist only of numbers, only one of which is the one that actually occurs.

Remark 2 For the representation of random numbers it is useful to think of a set $\mathcal{S}$, whose subset $\mathcal{Q}$ is constituted by the only possible alternatives for a certain individual at a given time. Sometimes, $\mathcal{S}$ can coincide with a manifold less extensive of the linear ambit or linear space $\mathcal{A}$ in which $\mathcal{S}$ is contained: in the case of two random numbers, $\mathcal{S}$ can coincide with a curve of the Cartesian plane $\mathcal{A}$, otherwise, if the numbers are three, $\mathcal{S}$ can coincide with a surface of the three-dimensional space $\mathcal{A}$. Then, the possible points of $\mathcal{Q}$ would be positioned on the curve of the Cartesian plane or on the surface of the three-dimensional space and such points may be all the points or a part or a few points of $\mathcal{S}$ according to the individual's knowledge at a given time and the existence of other restrictions and conditions. We could have $\mathcal{A} = \mathbb{R}^2$ or $\mathcal{A} = \mathbb{R}^3$ under one-to-one correspondence between the points of the two-dimensional or three-dimensional space and the ordered lists of two or three real numbers. If $\mathbb{R}^2$ and $\mathbb{R}^3$ are equipped with a scalar product positive-definite, they would be Euclidean spaces or metric spaces. However, since every vector space may be considered as an affine space over itself, $\mathcal{A}$ could also be an affine space and this, theoretically, would be the best thing by virtue of the fact that the affine properties are more general than the metric ones. The affine properties are the basis of essential concepts of probability theory and only they make sense, being independent of the choice of a coordinate system; however, the

importance of the metric properties appears in order to represent analytical conditions of coherence (de Finetti, 1931b, 1954, 1970).

Remark 3 The space of alternatives $\mathcal{S}$ of a random number X coincides with the real line x on which it is possible to consider $\mathcal{Q}$, subset of $\mathcal{S}$, which consists of the only possible values or points for a certain individual. Every point of the real line is assumed to correspond to a real number and every real number to a point of it, so the real line is a vector space of dimension 1 over the field $\mathbb{R}$ of real numbers, that is to say, over itself: there is an one-to-one correspondence between points on the real line and real numbers. The set $\mathbb{R}$ of real numbers is a Euclidean space because it has a standard scalar product which is simply ordinary multiplication of real numbers and the standard norm on it is simply the absolute value function. Every real number of the x-axis is a point of $\mathcal{S}$. Since every possible value of X is a random event, all the possible values of X are events, all together and implicitly considered into $\mathcal{Q}$. In conformity with the possible values of X which constitute the set $\mathcal{Q}$, X can belong to a half-line, $X \geq x$, or to an interval, $x_1 \leq X \leq x_2$, or to an arbitrary set, $X \in \mathcal{I}$.

Remark 4 If we consider two random numbers, X and Y, $\mathcal{S}$ coincides with the Cartesian plane whose element, in general, is (x, y). For (X, Y), $\mathcal{Q}$ consists of pairs of possible values for X and Y. If we consider three random numbers, X, Y and Z, $\mathcal{S}$ coincides with the ordinary space whose element, in general, is (x, y, z) and if we consider more than three random numbers, the only restriction for $\mathcal{S}$ is that it is not visually intuitive to go beyond the third dimension.

4. Random Entities

Random points, random vectors, random matrices, random sets and random functions are random entities. An objective scheme of representation for random entities is given by the set $\mathcal{S}$ of "points" whose elements can be a finite or infinite number. Such points are not geometric points, but they are simply elements of $\mathcal{S}$, that is to say, they may be points in two-dimensional Euclidean space or in three-dimensional Euclidean space, vectors, matrices, sets of points and functions if $\mathcal{S}$ is, respectively, a set of points or vectors or matrices or sets of points or functions. Clearly, we need to consider each "point" of $\mathcal{Q}$ or $\mathcal{S}$ like a random event which is, a posteriori, true or false: among such "points", there is a very important "point" representing the alternative which, a posteriori, will really occur. It is, a priori, uncertain and for this reason it constitutes the essence of every problem concerning the alternatives $\mathcal{Q}$ which are contained in $\mathcal{S}$ in which $\mathcal{Q}$ is embedded.

Remark 5 On a plane the point which would be hit in firing at a target is a random point, with the geometric representation of this problem which is independent of any coordinate system. Similarly, in ordinary space the point where, at a precise moment, a stolen car is, such a car being equipped with a satellite alarm, is random. When the theft occurs, this alarm sends to a control center a radio signal through which it is possible to determine the exact position of the vehicle. The space of alternatives $\mathcal{S}$, corresponding to the usual physical space extended in length, width and height and in which bodies move or place themselves, provides an immediate geometric image which does not depend on coordinates.

Remark 6 A vector is an ordered list of n real numbers, $(x_1, \ldots, x_n) \in \mathbb{R}^n$, where n is a non-negative integer: real numbers $x_1, \ldots, x_n$ are called scalar components in the n-dimensional Euclidean space, with the number x_i which is the i-th scalar component of $(x_1, \ldots, x_n)$. Thus, the list of known unit prices of ten articles which are for sale in a given department store is the decuple $(p_1, \ldots, p_{10})$. Given n, for a certain individual, a vector is random when he does not know all scalar components of the finite ordered list of n real numbers, such a list being the true vector. For the same individual, different n-tuples of $\mathbb{R}^n$, which constitute $\mathcal{Q}$, are possible. The space of alternatives $\mathcal{S}$ is a vector space over the field $\mathbb{R}$ of real numbers because it coincides with all the n-tuples of $\mathbb{R}^n$. Evidently, each n-tuple of $\mathbb{R}^n$, belonging to $\mathcal{S}$, is a point of $\mathcal{S}$.

Remark 7 A matrix (a_{ij}) m × n, with m, $n \geq 1$, is a rectangular array of mn numbers, $(a_{ij}) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$, whose elements are arranged in m rows and n columns. The numbers of every row could represent known unit prices of n given articles which are for sale in m different department stores. The whole of all rectangular arrays of mn real numbers is a vector space over the field $\mathbb{R}$ and an isomorphism exists between it and $\mathbb{R}^{mn}$ because every array of mn numbers can be arranged into a row vector or column vector of $\mathbb{R}^{mn}$. For a certain individual, a matrix which has predetermined rows and columns is random when he does not know the real numbers of every row or column of the true matrix. For the same individual, possible matrices which constitute $\mathcal{Q}$ and all those of the vector space $\mathcal{S}$ over $\mathbb{R}$ have the same predetermined number of rows and columns. Clearly, each matrix of $\mathcal{S}$ is a "point" of $\mathcal{S}$.

Remark 8 Random curves and random sets on surfaces are random sets which give a non-linear structure to $\mathcal{S}$. The unknown path of an airplane, from takeoff to landing, is a random curve: every trajectory can be thought of as a set

which consists of infinite points and it is a "point" of $\mathcal{S}$. On the other hand, if a given individual does not know the part of the Italian territory, viewable via satellite map, on which rain fell in the last twelve hours starting from a certain instant, we have a random set on surface: each part of the Italian territory is a set of infinite points and it is a "point" of $\mathcal{S}$. Moreover, among the different parts of the Italian territory which constitute the abstract space of alternatives $\mathcal{S}$, there are both the empty part, that corresponds to the alternative according to which rain did not fall on the Italian territory in the last twelve hours, and the total part, that corresponds to the alternative according to which rain fell on the whole Italian territory in the last twelve hours.

Remark 9 For a certain individual, a function Y(t), with the variable t which is time, is random when its behavior is unknown to him, for which it can be said that chance exists instant by instant. If one knows the values of Y(t) because they have been calculated at any number of instants $t = t_1, \ldots, t_n$, however large the finite n, the value of Y(t) at a different instant t will still be uncertain. Every measurable function, where its values are $Y(t_1), \ldots, Y(t_n)$, is a "point" of $\mathcal{S}$. When we ask whether or not the numerical values of a function Y(t) of the set $\mathcal{S}$ at given instants fall inside fixed sets $a_h \leq Y(t_h) \leq b_h$ (h = 1, …, n) defined by two freely determined coordinates, random events given by $a_1 \leq Y(t_1) \leq b_1, \ldots, a_n \leq Y(t_n) \leq b_n$ can be true or false according to whether they occur or not inside intervals $[a_1, b_1], \ldots, [a_n, b_n]$.

Evidently, each problem concerning the possible alternatives of $\mathcal{Q}$ is usefully visualized by means of $\mathcal{S}$ whose nature is always and unequivocally objective (de Finetti, 1970).

5. Arithmetic and Boolean Operations

Putting the logical values true and false equal to the idempotent numbers 1 and 0 for which we have $1^n = 1$, $0^n = 0$, an event E is always a random number which can admit these two numbers, called indicators of E. Arithmetic and Boolean operations must be unified by applying arithmetic operations even to events and Boolean operations even to random numbers. For events, the arithmetic product is the same as the logical product $\wedge$, the arithmetic sum is the number of successes $Y = E_1 + \ldots + E_n$ and complementation is negation, that is to say, $\overline{E} = 1 - E$. Obviously, Y can yield a result outside the $\{0, 1\}$ set. The logical sum $\vee$ can be expressed by $A \vee B = 1 - (1 - A)(1 - B)$, where we must consider $A \vee B = (\overline{\overline{A} \wedge \overline{B}})$, with A and B which are random events. In $\mathbb{R}$ we can make the following definitions: $a \vee b = \max(a, b)$, $a \wedge b = \min(a, b)$ and $\bar{a} = 1 - a$, where a and b are real numbers; then, in case a and b have as values 1 or 0, the logical product, the logical sum and the negation are recovered. Moreover, it needs to unify the notation for the probability of an event E and for the mathematical expectation or prevision of a random number X; in fact, it is adopted **P**(E) for probability of E and **P**(X) for prevision of X, where **P** is linear, that is, additive and homogeneous.

6. Logic of Uncertainty

The subjectivistic conception of probability, through psychological analysis, vivifies notions that are mathematically correct but that is not sufficient to consider from the formal point of view. In fact, the instrument really propulsive of scientific thinking is not classical logic or, in the specific instance, logic of certainty that, as such, involves no affective demonstration, no judgement by anyone, but is probability and probability calculus. Therefore, when we consider any problem concerning the assignment of probability among possible cases and how to define it and to express it quantitatively, we find ourselves into the field, personal and subjective, of logic of uncertainty, clearly distinct and separate from that one of logic of certainty (de Finetti, 1931a). Indeed, when we say that we are not satisfied of logic of certainty, we mean that we are not satisfied of agnostic and undifferentiated attitude towards uncertainty. For all those things which, not being known to us with certainty, are uncertain or possible, any individual feels a more or less strong propensity to expect that some cases possible are true rather than others, to believe that the answer to a given question is no rather than yes, to estimate that the unknown value of a certain quantity is small rather than large. Evidently, these attitudes express, in the domain of uncertainty, different degrees of subjective probability, each of which is assigned to one of the possible alternatives identified by a given individual on the basis of his knowledge. So, to find oneself into the field of logic of prevision means to examine carefully desires or hopes that certain alternatives occur, anxieties and fears regarding the occurrence of unfavourable alternatives and to weigh up the pros and cons of each choice trying to reason about it in order to distribute, among all the possible alternatives and in the way which will appear most appropriate, one's own sensations of probability (de Finetti & Minisola 1961; de Finetti, 1955, 1963, 1969).

Remark 10 When a particular individual chooses to be guided only by the logic of certainty, after having distinguished a set more or less large of possible alternatives in the way which seems to him most effective, he has to stop because the question is closed. Remaining within the logic of certainty, the only thing that he could make is a prophecy, that is to say, among the cases that he believes possible, he might venture to guess the alternative that, according to him, will occur, transforming in this way, but unreasonably, the uncertainty in illusory certainty (de Finetti, 1967b, 1970).

Remark 11 The space of n random numbers coincides with the n-dimensional vector space $\mathcal{A}$ after the introduction of a coordinate system $x_1, \ldots, x_n$ in $\mathcal{A}$; by virtue of the fact that each event is a random number, a set of n possible events

$E_1, \ldots, E_n$ is embedded in $\mathcal{A}$. From such a set other events, called constituents, are originated: they are identified by particular ordered lists of n numbers $(x_1, \ldots, x_n)$, with $x_i = 0$ or $x_i = 1$, $i = 1, \ldots, n$, each of which is a possible point of $\mathcal{Q}$ contained in the vector space $\mathcal{A}$. Such considerations make clear, from the point of view of the logic of certainty, why the probability of an event is automatically incorporated in the prevision of a random number. In fact, going beyond the domain of the logic of certainty, we enter into the field of the logic of uncertainty and in the event that X is a random number, $\mathbf{P}(X)$ is the prevision of X: if $I(X) = \{x_1, \ldots, x_n\}$, when we assign to each value x_i of X the probability p_i ($i = 1, \ldots, n$), with $0 \leq p_i \leq 1$ and $\sum p_i = 1$, it turns out to be $\mathbf{P}(X) = x_1p_1 + \ldots + x_np_n$. The prevision of X coincides with the probability of an event E when and only when X, admitting only two possible values, 1 and 0, is an event, thus prevision and probability are two different words that express the same concept extra-logical, subjective and personal (de Finetti & Emanuelli, 1967a).

Remark 12 The die symmetry and the knowledge of an observed frequency are elements which any individual carefully examines to express his opinion from which the subjective probability is originated. According to the subjectivistic conception, the only probability that exists in any case is the subjective probability. It must be understood as the degree of belief of a certain individual in the occurrence of a specific event; anyway, probability of an event E is not an intrinsic characteristic of E because it depends on the information that the individual making the probabilistic evaluation has, so it is always subordinate to his present state of knowledge which can change for the possible attainment of new essential information and for the passage of time (de Finetti, 1963).

Remark 13 A probabilistic evaluation, known over a set of whatever events, always expresses the opinion of a given individual, real or hypothetical; the only admissible restriction is that this opinion is coherent, consequently, if it is not coherent it should be corrected by the individual in order to make it coherent (de Finetti, 1930a, 1930b).

7. Objective Statements and Subjective Evaluations

We reason in accordance with probability theory, although without awareness and in a rather approximate way, when we incessantly make our forecasts and assumptions which constitute the usual object of our thinking in all the practical circumstances of life, more than the much rarer judgements which are logically certain. In fact, we reason in all the circumstances of life, where we base ourselves on probabilities, by applying without awareness the two fundamental theorems of probability calculus, the theorem of total probability and the theorem of composite probability: our way of thinking is not forced by logical requirements but is only suggested by psychological motives when we judge on the probability that it will rain or not in order to decide to take or not the umbrella, on the probability for an individual to arrive in time at the postal office on foot in order to decide to go by bus or taxi or not, on the probability that different performances which have been announced for tonight are more or less interesting to decide whether to go and where and so on. Regarding these examples, nobody can certainly think that they are cases of objective probability because it is not be able to solve such problems. Instead, according to the subjectivistic viewpoint, every question has an exact and satisfactory answer, because it is always based on the psychological degree of confidence of a certain individual in relation to a certain assumption. In all cases, including the gambling games or statistics or molecular physics cases or any other case whose objective probability coincides with the subjective one, it is evidently only a matter of a pure psychological feeling. Anyway, the theorems of probability theory are always valid, thus justifying one of our most important empirical ways of reasoning.

Hence, any statement of probability calculus has an objective or logical meaning unlike probability evaluations whose meaning can only be empirical. For example, we consider a deck of Italian playing cards which consists of 40 cards divided into 4 suits; in particular, Neapolitan playing cards are divided into swords, cups, coins and clubs, whose 3 face cards per suit are knave or fante in Italian, knight or cavallo in Italian, king or re in Italian. Thus, if we suppose that the probability of drawing a fante or cavallo or re is $\mathbf{P}(E_f) = \mathbf{P}(E_c) = \mathbf{P}(E_r) = 1/10$, then we conclude that the probability of E, where E consists in drawing a face card, is given by $\mathbf{P}(E) = \mathbf{P}(E_f) + \mathbf{P}(E_c) + \mathbf{P}(E_r) = 3/10$: we make a purely logical reasoning because it is logically true that the three considered events are mutually exclusive and under such a condition it is logically certain that the theorem of total probability is valid. However, probability evaluations have an empirical or subjective meaning: if the probability of drawing a fante is $1/10$ for us, we always express a subjective opinion. In accordance with the subjectivistic viewpoint, we do not believe that the probability of any event E, $\mathbf{P}(E)$, with $0 \leq \mathbf{P}(E) \leq 1$, is objectively determined because we consider, on the contrary, all the functions $\mathbf{P}$ as formally admissible laws when they are not in conflict with theorems of probability calculus. Evidently, the choice of one of these functions is left to each individual who chooses according to his subjective opinion. Regarding the previous example, we consider admissible all the ∞^3 functions $\mathbf{P}$ for which it turns out to be $\mathbf{P}(E_f) = x$, $\mathbf{P}(E_c) = y$, $\mathbf{P}(E_r) = z$, $\mathbf{P}(E) = x + y + z$, with $x, y, z \geq 0$ and $x + y + z \leq 1$. The choice of functions for which we have $\mathbf{P}(E_f) = \mathbf{P}(E_c) = \mathbf{P}(E_r) = 1/10$, although suggested by spontaneous and universally approved remarks, is a very particular case and it is not forced by logical requirements of which mathematics can or must be interested. Obviously, recognizing if certain premises are sufficient or not in order to involve a certain conclusion becomes very difficult when the problem under consideration is not as

simple as in the previous example. However, such a problem is never solved when there is not a clear separation between all that is logical or objective and all that is empirical or subjective (de Finetti, 1930a, 1930b).

8. Criteria for the Probabilistic Evaluations

It is representative of one of the primary necessities of science the fact that it must not run the risk of taking as notions illusory combinations of terms of a metaphysical nature, but it must work with concepts of verified validity in a practical meaning. Therefore, its definitions must be operational, that is to say, must reduce a scientific concept not simply to sentences having only an apparent meaning, but to real experiences which are at least theoretically possible. Thus, the criteria which may be used to reveal concretely $\mathbf{P}(X)$ or, in particular, $\mathbf{P}(E)$ according to the opinion of a certain individual are two and entirely equivalent: they are based upon the identification of the practical consequences that a given individual knows to accept and accepts when he expresses his evaluation of $\mathbf{P}(X)$ or $\mathbf{P}(E)$ and, if coherently applied, lead to the same $\mathbf{P}(X) = \bar{x}$ in the event that X is estimated or to the same $\mathbf{P}(E) = p$ in the case that E is evaluated. If X is evaluated, both criteria contain the random magnitude $X - \bar{x}$, expressed by the difference between the real value X and the one chosen by a certain individual at his own will, $\mathbf{P}(X) = \bar{x}$. The first criterion provides that, after the subjective choice of $\bar{x}$, the individual is obliged to accept any bet unilaterally determined by an opponent, whose gain is $c(X - \bar{x})$, with c any betting amount, positive or negative, equally determined by the opponent; in particular, if $c = 1$, the gain of the bet is $(X - \bar{x})$, while if $c = -1$, it is $(\bar{x} - X)$. On the contrary, the second criterion provides that, after choosing $\bar{x}$, the individual must suffer the penalty $(X - \bar{x})^2$, positively proportional to the square of the difference between X and $\bar{x}$ (de Finetti, 1970). In particular, if an event E is evaluated, both criteria contain the magnitude $E - p$ given by the difference between the real value E, 1 or 0 according to whether E occurs or does not occur, and the one chosen by a certain individual according to his subjective opinion, $\mathbf{P}(E) = p$. The first criterion provides that, after the choice of p by a determined individual, he is obliged to accept any bet unilaterally determined by an opponent, whose gain is $c(E - p)$, where c is any betting amount, positive or negative, established by the opponent; in particular, if $c = 1$, the gain is $(E - p)$, while if $c = -1$, it is given by $(p - E)$. On the contrary, the second criterion provides that, after the subjective choice of p, the individual must suffer the penalty $(E - p)^2$. Evidently, in order to measure subjective probabilities, that is to say, to translate our degree of uncertainty, regarding judgements, into numerical determinations, the degree of confidence that we have in the occurrence of events is expressed by the conditions at which one would bet. There is a difference between judging if a bet is fair and judging how convenient it is for a certain individual, at a certain time, under certain circumstances, to accept it; moreover, the convenience will be judged differently, depending on the character of the individual and his love of risk. In other words, there is an essential difference between the case of one occasional and well defined betting and the abnormal case of an individual who would consistently and interminably be driven to betting.

9. Necessary and Sufficient Conditions of Coherence

The choice of $\mathbf{P}(X)$ or $\mathbf{P}(E)$, even if it is subjective, should not be contradictory and takes place within the set of coherent previsions of X or in that one of coherent probabilities of E; both the sets contain values objectively admissible which are independent of the personal views of any individual and also of the judgements about others' opinions. The necessary and sufficient conditions for coherence are two and completely equivalent, one for each evaluation criterion (de Finetti, 1970).

Regarding the first definition of coherence, it is assumed that the individual who subjectively evaluates $\mathbf{P}(X_i)$ or $\mathbf{P}(E_i)$, with $i = 1, \ldots, n$, does not want to make bets on X_i or E_i that give him an inevitable loss, therefore a set of his previsions or probabilities is not intrinsically contradictory when and only when, among the linear combinations of bets that he is obliged to accept, there are not combinations with gains all uniformly negative. Analytically, this means that for the numerical values of the random magnitude $Y = c_1(X_1 - \bar{x}_1) + \ldots + c_n(X_n - \bar{x}_n)$ or the random magnitude $Y = c_1(E_1 - p_1) + \ldots + c_n(E_n - p_n)$ must not be, objectively, that sup I(Y) is negative; conversely, we have that inf I(Y) cannot be positive. Even if the bets are an infinite number, Y is always linear combination of a finite number of them.

Regarding the second definition of coherence, it is assumed that the individual who subjectively evaluates $\mathbf{P}(X_i)$ or $\mathbf{P}(E_i)$, with $i = 1, \ldots, n$, does not prefer a given penalty if he can choose another penalty certainly smaller, therefore a set of his previsions or probabilities is coherent when and only when he could not choose them in order to make his penalty certainly and uniformly smaller. Analytically, this means that there are not any evaluations $\mathbf{P}^*(X_i)$ or $\mathbf{P}^*(E_i)$ that replaced with the evaluations $\mathbf{P}(X_i)$ or $\mathbf{P}(E_i)$ are such that for all the possible points, $(X_1, \ldots, X_n)$ or $(E_1, \ldots, E_n)$, the penalty $L^* = \sum_i (X_i - \mathbf{P}^*(X_i))^2 \cdot (1/k_i)^2$ is uniformly smaller than the penalty $L = \sum_i (X_i - \mathbf{P}(X_i))^2 \cdot (1/k_i)^2$ or the penalty $L^* = \sum_i (E_i - \mathbf{P}^*(E_i))^2 \cdot (1/k_i)^2$ is uniformly smaller than $L = \sum_i (E_i - \mathbf{P}(E_i))^2 \cdot (1/k_i)^2$, with $k_1, \ldots, k_n$ which are arbitrarily predetermined and homogeneous towards X_i or E_i.

A prevision $\mathbf{P}$ is coherent if its use cannot lead to an inadmissible decision such that a different possible decision would have certainly led to better results, whatever happened. If the sets of possible values for X and Y are, respectively, I(X)

= $\{x_1, \ldots, x_n\}$ and $I(Y) = \{y_1, \ldots, y_n\}$, when we assign the same weights p_i $(i = 1, \ldots, n)$, where we have $0 \leq p_i \leq 1$ and $\sum p_i = 1$, to each x_i and y_i we will have $\mathbf{P}(X + Y) = \mathbf{P}(X) + \mathbf{P}(Y)$, that is to say, **P** is additive; a prevision **P** of the random number X must satisfy the inequality $\inf I(X) \leq \mathbf{P}(X) \leq \sup I(X)$, that is, $\mathbf{P}(X)$ must not be less than the lower bound of the set of possible values for X, which is $\inf I(X)$, nor greater than the upper bound, which is $\sup I(X)$. A prevision **P** of X must also be linear, that is, we have $\mathbf{P}(aX) = a\mathbf{P}(X)$, for every real number a. In general, we have $\mathbf{P}(aX + bY + cZ + \ldots) = a\mathbf{P}(X) + b\mathbf{P}(Y) + c\mathbf{P}(Z) + \ldots$, with a, b, c, … whatever real numbers, for any finite number of summands. So, coherence reduces to linearity, which contains additivity property, and convexity. Similarly, if E is an event, when we have $0 \leq \mathbf{P}(E) \leq 1$, its evaluation is coherent; if $E_1, \ldots, E_n$ are mutually exclusive events, their evaluations are coherent when we have $\mathbf{P}(E_1 + \ldots + E_n) = \mathbf{P}(E_1) + \ldots + \mathbf{P}(E_n)$ (de Finetti, 1970).

10. Geometric Interpretation of Conditions of Coherence

Given in □ n random numbers $X_1, \ldots, X_n$, with $\mathcal{A}$ n-dimensional vector space having coordinate system $x_1, \ldots, x_n$, every prevision, coherent or not, of each random number X_i is always a point $(\mathbf{P}(X_1), \ldots, \mathbf{P}(X_n))$ of $\mathcal{A}$. In this space, moreover, the coordinates of the points $\mathcal{Q}$ of the set $\mathcal{Q}$ of possible points are identified by ordered lists $(x_1, \ldots, x_n)$ of n real numbers, with x_1 that is a possible value of $X_1, \ldots, x_n$ that is a possible value of X_n. Thus, on the basis of the geometric interpretation of the necessary and sufficient conditions for coherence, the set $\mathcal{P}$ of coherent previsions **P** is the closed convex hull of the set $\mathcal{Q}$ of the possible points Q of $\mathcal{A}$ (de Finetti, 1970).

Remark 14 The first condition of coherence involves that a point P of $\mathcal{A}$, with coordinates $(\mathbf{P}(X_1), \ldots, \mathbf{P}(X_n))$, is an admissible prevision if and only if no hyperplane separates it from the set $\mathcal{Q}$ of the possible points $\mathcal{Q}$ of $\mathcal{A}$: this characterizes the points of the convex hull, for which it is said that every linear equation between the numbers $X_i, c_1X_1 + \ldots + c_nX_n = c$, must also apply to the previsions $\mathbf{P}(X_i)$, $c_1\mathbf{P}(X_1) + \ldots + c_n\mathbf{P}(X_n) = c$, as well as any inequation between them, $c_1X_1 + \ldots + c_nX_n \geq c$, must also be satisfied by the previsions, $c_1\mathbf{P}(X_1) + \ldots + c_n\mathbf{P}(X_n) \geq c$.

Remark 15 The vector space $\mathcal{A}$ is Euclidean when it is provided with a scalar product positive-definite: by virtue of the metric $\rho^2 = \sum_i (x_i/k_i)^2$, it results $L = (P - Q)^2$, that is to say, the penalty L coincides with the square of the distance between the prevision-point P and the outcome-point Q. Thus, regarding the second condition of coherence, the points of the convex hull also enjoy the property according to which P cannot be moved in such a way as to reduce its distance from all points Q.

The points which are admissible in terms of coherence can be obtained as barycentres of, at most, n + 1 points Q_j of $\mathcal{Q}$ in the n-dimensional space or they are adherent points of $\mathcal{Q}$, but not belonging to $\mathcal{Q}$. More explicitly, every prevision-point P of $\mathcal{P}$ is admissible in terms of coherence when it is a barycentre of possible points Q_j of $\mathcal{Q}$, with non-negative weights, summing to 1: however, if all the weights are concentrated at a unique point Q_j, also the possible points turn out to be coherent previsions (de Finetti, 1970).

11. Conclusions

Probability exists only in our own judgement because it is always the degree of belief of a given individual for the occurrence of a given event. Nevertheless, when it needs, any individual can assess the probability of an event on the basis of an observed frequency or dividing the number of favourable outcomes to it by the total number of possible outcomes which are equally possible. In fact, the subjectivistic theory is not in contrast with any other provided that such different interpretations accept the role of particular criteria for the evaluation of the probability and give up the pretence of leading to a definition of probability. Each criterion for subjective evaluations furnishes an operational definition of probability or prevision **P** and together with the corresponding conditions of coherence can be taken as a foundation for the entire theory of probability. When we study this we show the dichotomy between the subjective or psychological aspect of probability and the objective or logical or geometrical one. Analytically and objectively, the first definition of coherence is similar to the property of stable equilibrium of the barycentre, while the second definition is similar to the property of minimum of the moment of inertia which characterizes the barycentre once again. When the properties of the barycentre are not satisfied, the set of previsions of a given individual cannot be coherent. Given the probabilities of the possible values, finite in number, of X, its barycentre, which is $\mathbf{P}(X)$, can be expressed as a function of them; the prevision of X does not presuppose the introduction of the concept of continuous probability distribution that, extending to the general case the concept of mathematical expectation or mean value of X, requires the use of mathematical tools more advanced than necessary.

References

de Finetti, B. (1930). Fondamenti logici del ragionamento probabilistico. *Bollettino dell'Unione Matematica Italiana,* 9, 258-261.

de Finetti, B. (1930). Problemi determinati e indeterminati nel calcolo delle probabilit à *Rendiconti della R. Accademia Nazionale dei Lincei,* 12, 367-373.

de Finetti, B. (1931). *Probabilismo. Saggio critico sulla teoria delle probabilità e sul valore della scienza (Biblioteca di Filosofia, diretta da A. Aliotta).* Napoli-Città di Castello, F. Perrella.

de Finetti, B. (1931). Sul significato soggettivo della probabilità. *Fundamenta Mathematicae,* 17, 298-329.

de Finetti, B. (1954). La nozione di evento. *Pubblicazioni delle Facoltà di Scienze e di Ingegneria dell'Università di Trieste,* Serie B 130, 170-174.

de Finetti, B. (1955). La probabilità e le scienze sociali. *L'industria,* 4, 467-495.

de Finetti, B., & Minisola, F. (1961). *La matematica per le applicazioni economiche.* Roma, Cremonese.

de Finetti, B. (1963). La decisione nell'incertezza. *Scientia,* 88, 61-68.

de Finetti, B., & Emanuelli, F. (1967). *Economia delle assicurazioni.* Trattato Italiano di Economia, vol. XVI, Torino, UTET.

de Finetti, B. (1967). L'adozione della concezione soggettivistica come condizione necessaria e sufficiente per dissipare secolari pseudoproblemi. In D. Fürst, & G. Parenti, *I fondamenti del calcolo delle probabilità. Atti della tavola rotonda tenuta a Poppi nei giorni 11-12 giugno 1966* (pp. 57-94). Scuola di Statistica dell'Università, Firenze.

de Finetti, B. (1969). *Un matematico e l'economia.* Milano, F. Angeli.

de Finetti, B. (1970). *Teoria delle probabilità, voll. I e II.* Torino, Einaudi.

International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Multivariate Kurtosis as a Tool for Comparing Copula Models

Ampalavanar Nanthakumar

Department of Mathematics, State University of New York at Oswego, Oswego NY 13126, U.S.A. E-mail: ampala.nanthakumar@oswego.edu

Received: March 26, 2016 Accepted: April 14, 2016 Online Published: June 22, 2016

doi:10.5539/ijsp.v5n4p67 URL: http://dx.doi.org/10.5539/ijsp.v5n4p67

Abstract

This paper studies the effectiveness of the Multivariate Kurtosis in comparing the Clayton Copula and the Farleigh-Gumbel-Morgenstern Copula in modeling when the actual populations follow either the bivariate exponential distribution or the bivariate normal distribution. The study shows that the Multivariate Kurtosis (as defined by Mardia) is a very effective tool in comparing Copulas and that Farleigh-Gumbel-Morgenstern Copula is slightly more accurate than the Clayton Copula for modeling.

Keywords: Multivariate, Kurtosis, Copula

1. Introduction

The copulas are used as a general way of formulating a multivariate distribution in such a way that the dependence can be infused in a reasonable manner. This is based on the simple idea that the joint distribution can be represented as a transformation of the underlying marginal distributions (see Sklar 1959). There are several types of copulas and each differ according to the strength of the dependence and the direction of the association. Ali et al (1978) studied the use of Copulas to construct the bivariate logistic distribution. Aas et al (2009) and Low et al (2013) investigated the applicability of paired (or "Vine") Copulas in the context of finance and portfolio management. Schölzel,and Friederichs (2008) studied the use of Copulas in climatology. Heinen and Rengifo (2007), Nikoloulopoulos and Karlis (2009, 2010), Karlis and Pedeli (2013) constructed bivariate integer-valued autoregressive models based on Copulas. For further literature review, the interested readers are referred to Nelson (2006). Due to the availability of many Copula models and its applications in many subject fields, there is a lot of interest in identifying the "best" Copula model. The question is "How do you compare these different Copula models ?". Here in this paper, we want to compare two particular copulas namely; Clayton Copula, and Farleigh-Gumbel-Morgenstern Copula. The Clayton Copula belongs to the family of Archimedean Copulas while the Farleigh-Gumbel-Morgenstern Copula does belong to the non-Archimedean family. We consider these copulas in the context of modeling the bivariate exponential distribution and the bivariate normal distribution. In order to compare these two Copulas, we use kurtosis as a tool. The kurtosis has been in use for a long time to study the 'peakedness' of the probability distributions. In fact, it is used as a measure to identify the distributions. For example, for the normal distribution the kurtosis is 3, for the bivariate normal distribution the kurtosis is 8, for the exponential distribution the kurtosis is 9, and so on.

In multivariate statistical analysis, normality of the sample is assumed in many cases. Hence, assessing for multivariate normality is an important problem. Similarly, in actuarial models, the bivariate exponential distribution is important and so verification of bivariate exponential is equally important. For this purpose, we will use the bivariate kurtosis as defined in Mardia (1970). In fact, there are several definitions for the Multivariate kurtosis (See, Mardia (1970), Malkovich and Afifi (1973), Srivastava (1984), and Mardia (1970, 1974)). Mardia defined multivariate kurtosis as a natural extension of the univariate case. To assess the multivariate normality, multivariate kurtosis has been defined and its asymptotic distributions under the multivariate normality have been given in Mardia (1970). Furthermore, Srivastava (1984) has considered another definition by using principal component scores and has derived their asymptotic distributions under the null hypothesis. Recently, Miyagawa et al (2011) proposed a sample measure of multivariate kurtosis of the form containing Mardia (1970) and Srivastava (1984). According to Mardia (1970), the multivariate kurtosis can be defined as $\beta_2 = E\left(\left(X-\mu\right)' \Sigma^{-1}\left(X-\mu\right)\right)^2$ where X is the observation vector; μ is the mean vector;

and Σ is the variance-covariance matrix.

In this paper, we propose a method to compare the copulas by using the bivariate kurtosis. We compare the copula through numerical calculations and graphs. Moreover, we check the effectiveness of this approach by estimating the Euclidean distance between the Copula and the actual distribution. The multivariate kurtosis seems to be a good measure for comparing the Copulas.

2. Methodology

Copulas:

According to the theory of probability, every joint probability distribution is a function of the marginal distributions (see Sklar (1959) for the details). Copula is a probability model that gives us a way to construct the joint distribution from the marginal distributions. In other words, the Copulas are mathematical models that approximate the multivariate distribution function. There are two families of Copula; Archimedean family of Copulas and the non-Archimedean family of Copulas. Each family consists of several kinds of Copulas. Some of these Copulas are of the discrete type while others are of continuous type.

Next, we present the definitions and the methodology necessary for the construction of the Copulas.

Definition:

A copula is a multivariate joint distribution defined on the k dimensional unit cube $[0,1]^k$ such that every marginal distribution is uniform on the interval $[0,1]$.

In other words, $C:[0,1]^k \to [0,1]$ is a $k-$dimensional copula if

(a). $C(u)=0$ whenever $u \in [0,1]^k$ has at least one component equal to 0.

(b). $C(u)=u_i$ whenever $u \in [0,1]^k$ has all the components equal to 1 except the i^{th} one which is equal to u_i.

(c). $C(u)$ is $k-$increasing.

3. Copula Construction

Archimedean Copula:

This is a family of copulas and the $k-$dimensional Archimedean Copula is defined as follows.

$$C(u_1,u_2,.....,u_k)=\Psi^{-1}\left(\sum_{i=1}^{k}\Psi(u_i)\right) \qquad (2.1)$$

where Ψ is known as the generator function and u_i is the marginal distribution of the i^{th} component.

Any generator function which satisfies the following properties is the basis for a copula.

$$\Psi(1)=0, \quad \text{limit } \Psi(x)=\infty \ , \quad \Psi'(x)<0 \ , \ \Psi''(x)>0 \qquad x \to 0$$

Special Case:

Clayton Copula

Let the generator function $\Psi(u)= u^{-\alpha}-1$ where $\alpha>1.$ One can show that the functional inverse,

$$\Psi^{-1}(u) = (1+u)^{-1/\alpha}.$$

In the bivariate case, the Archimedean formulation yields the Clayton Copula as

$$C(u_1,u_2)=\left(u_1^{-\alpha}+u_2^{-\alpha}-1\right)^{-1/\alpha} \qquad (2.2)$$

Non-Archimedian Copula:

The non-Archimedian copulas do not possess a generator function. These copulas are distinctly different from the Archimedian copulas.

Special Case:

Farleigh-Gumbel-Morgenstern Copula:

$$C(u_1, u_2) = u_1.u_2.(1 + \alpha.(1-u_1).(1-u_2)) \tag{2.3}$$

where α is the dependence parameter and $|\alpha| \le 1$.

Modeling the Joint Distribution by Copulas:

Let us suppose that the joint distribution of (X, Y) is $F(x, y)$ is unknown, but the marginal distributions are known.

Note that $F(x, y)$ is the joint distribution.

Let $u_1 = F_1(X)$ and $u_2 = F_2(Y)$ where F_1 and F_2 are the marginal distributions of X and Y respectively.

Theorem #1 (Sklar):

If X and Y are continuous random variables then there exists a continuous Copula $C(u_1, u_2)$ such that $C(u_1, u_2) = F(x, y)$.

Theorem #2:

The conditional distribution of Y given X = $\dfrac{\partial C(u_1, u_2)}{\partial u_1}$ (2.4)

Note: We will use the above results in Copula modeling.

Our objective here is to see whether the Copula models nearly resemble the actual populations and to identify the populations as either as “normal” or “non-normal” based on the kurtosis. Towards this, we will consider the bivariate exponential and the bivariate normal populations in this paper.

First, we will consider the bivariate exponential population. In other words, the bivariate observations (x_i, y_i) follows the bivariate exponential distribution with the cumulative distribution function (cdf) given by

$$F(x, y) = 1 - e^{-\lambda_1.x} - e^{-\lambda_2.y} - e^{-\lambda_1.x - \lambda_2.y + \lambda.\min(x,y)} \tag{2.5}$$

In the absence of any knowledge about the actual population distribution, we propose to use the Copulas to model this population distribution based on the marginal distributions. Here, we assume that the marginal distributions are known.

Say, that for the X variable, its marginal distribution $F_1(x)$ is given by $F_1(x) = 1 - e^{-\lambda_1.x}$. Similarly, the marginal distribution of the Y variable is given by $F_2(y) = 1 - e^{-\lambda_2 y}$.

Modeling Clayton Copula:

$$C(u_1, u_2) = \left(u_1^{-\alpha} + u_2^{-\alpha} - 1\right)^{\frac{-1}{\alpha}} \tag{2.6}$$

where u_1 and u_2 are the marginal distributions of X and Y respectively.

In order to generate the Clayton Copula, we will use Theorem #2.
For the Clayton Copula,

$$\frac{\partial C}{\partial u_1} = \left(u_1^{-\alpha} + u_2^{-\alpha} - 1\right)^{\frac{-1}{\alpha}} . u_1^{-\alpha} \;=\; v \quad \text{(say)} \tag{2.7}$$

where v is the conditional distribution of Y given X. So, v is uniformly distributed between 0 and 1.
This means that,

$$u_2 = F_2(y) \;=\; \left\{1 - u_1^{-\alpha} . \left(1 + v^{\frac{-\alpha}{(\alpha+1)}}\right)\right\}^{\frac{-1}{\alpha}} \tag{2.8}$$

$$\text{So, } 1 - e^{-\lambda_2 . y} = \left\{1 - u_1^{-\alpha}\left(1 + v^{\frac{-\alpha}{(\alpha+1)}}\right)\right\}^{\frac{-1}{\alpha}} \tag{2.9}$$

$$\text{and } y = \frac{-1}{\lambda_2} . \ln\left\{1 - \left(1 - u_1^{-\alpha}\left(1 + v^{\frac{-\alpha}{(\alpha+1)}}\right)\right)^{\frac{-1}{\alpha}}\right\} \tag{2.10}$$

Also, note that $u_1 = 1 - e^{-\lambda_1 x}$

This means that,

$$x = \frac{-1}{\lambda_1} . \ln\{1 - u_1\} \tag{2.11}$$

Note that α is the dependence parameter and it can be estimated by using the relationship that Kendall's Tau, $\tau = \frac{\alpha}{(\alpha + 2)}$. By using equations (2.10) and (2.11) the Clayton copula based samples can be generated.

Modeling Farleigh-Gumbel-Morgenstern Copula:

$$C(u_1, u_2) = u_1 . u_2 . (1 + \alpha . (1 - u_1) . (1 - u_2)) \tag{2.12}$$

where α is the dependence parameter and $|\alpha| \le 1$.

$$\text{Note that } \frac{\partial C}{\partial u_2} = u_1 + \alpha . \left(u_1 - u_1^{\,2}\right) . (1 - 2 . u_2) = z \tag{2.13}$$

where z represents the conditional distribution of X given $Y = y$.

This yields the quadratic equation,

$$\alpha . (1 - 2 . u_2) . u_1^{\,2} - \{1 + \alpha . (1 - 2 . u_2)\} . u_1 + z = 0 \tag{2.14}$$

The solution is given by

$$u_1 = \frac{\{1+\alpha.(1-2.u_2)\} - \sqrt{\{1+\alpha.(1-2.u_2)\}^2 - 4.\alpha.(1-2.u_2).z}}{2.\alpha.(1-2.u_2)} \quad (2.15)$$

Note that, $u_1 = 1-e^{-\lambda_1.x}$ and $u_2 = 1-e^{-\lambda_2.y}$ (2.16)

Now, let $A = 1+\alpha.(1-2.u_2)$ and $B = [1+\alpha.(1-2.u_2)]^2 - 4.\alpha.(1-2.u_2).z$

This means,

$$x = \frac{1}{\lambda_1}.\ln\left\{\frac{2.(A-1)}{A+\sqrt{B}-2}\right\} \text{ and } y = \frac{-\ln(1-u_2)}{\lambda_2} \quad (2.17)$$

Note that again α is the dependence parameter and it can be estimated by using the relationship that the Pearson Coefficient of Correlation $\rho = \frac{\alpha}{4}$. By using equation (2.17), Farleigh-Gumbel-Morgenstern copula based samples can be generated.

Next, in order to simulate the samples directly from the Bivariate Exponential distribution, we will do the following.

Let us choose a value λ such that $\lambda < \min\{\lambda_1, \lambda_2\}$. Next, simulate random observations $\tilde{U}$ according to the exponential distribution with mean = $\frac{1}{(\lambda_1 - \lambda)}$. Similarly, simulate random observations $\tilde{V}$ according to the exponential distribution with mean = $\frac{1}{(\lambda_2 - \lambda)}$. Also, we can simulate $\tilde{W}$ according to an exponential distribution with mean = $\frac{1}{\lambda}$.

Now, let us define, $\tilde{X} = \min\{\tilde{U}, \tilde{W}\}$ and $\tilde{Y} = \min\{\tilde{V}, \tilde{W}\}$. We can easily show that $(\tilde{X}, \tilde{Y})$ jointly follow the bivariate exponential distribution as given by (2.6). We are interested in studying two things in this paper.

(i).The suitability of the Copula models to study the covariance structure.

(ii).The use of kurtosis as a tool to check the validity of the Copula models.

Kurtosis:

Let us first introduce the notations that we will use in the context of kurtosis calculation.

μ_X = Mean of X

μ_Y = Mean of Y

$\sum$ = Covariance Matrix of (X, Y)

Next, we present Mardia's definition for Kurtosis for the multivariate situation.

Kurtosis (Mardia)

The bivariate kurtosis, κ is given by

$$\kappa = E\left\{(X-\mu_X, Y-\mu_Y)\Sigma^{-1}(X-\mu_X, Y-\mu_Y)^T\right\}^2 \tag{2.18}$$

Next, to derive the kurtosis, let us note that the inverse of the variance-covariance matrix is

$$\sum^{-1} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{2.19}$$

Then, one can easily show that

$$\begin{aligned}\kappa = {} & a_{11}{}^2.E(X-\mu_X)^4 + a_{22}{}^2.E(Y-\mu_Y)^4 + 4.a_{12}{}^2.E\left((X-\mu_X)^2.(Y-\mu_Y)^2\right) \\ & + 2.a_{11}.a_{22}.E\left((X-\mu_X)^2.(Y-\mu_Y)^2\right) + 4.a_{11}.a_{12}.E\left((X-\mu_X)^3.(Y-\mu_Y)\right) \\ & + 4.a_{22}.a_{12}.E\left((X-\mu_X).(Y-\mu_Y)^3\right)\end{aligned} \tag{2.20}$$

Note that $a_{11} = \sigma_X{}^{-2}.(1-\rho^2)^{-1}$ (2.21)

$$a_{22} = \sigma_Y{}^{-2}.(1-\rho^2)^{-1} \tag{2.22}$$

$$a_{12} = \frac{-\rho.(1-\rho^2)^{-1}}{\sigma_X.\sigma_Y} \tag{2.23}$$

where ρ is the correlation coefficient between X and Y.

Next, we present the following results in order to evaluate the kurtosis. The proofs can be found in the Appendix.

If the joint probability distribution is a bivariate normal then

Result 1: $E\left((X-\mu_X)^2(Y-\mu_Y)^2\right) = \sigma_X{}^2.\sigma_Y{}^2.(1-\rho^2) + 3.\rho^2.\sigma_X{}^2.\sigma_Y{}^2$

Result 2: $E\left((X-\mu_X)^3(Y-\mu_Y)\right) = 3.\rho.\sigma_X{}^3.\sigma_Y$

Result 3: $E\left((X-\mu_X)(Y-\mu_Y)^3\right) = 3.\rho.\sigma_X.\sigma_Y{}^3$

Result 4: Kurtosis, $\kappa = 8$.

If the joint distribution is bivariate exponential with the density function

$$f(x,y) = \begin{cases} \lambda_1.(\lambda_2-\lambda).e^{-\lambda_1.x-(\lambda_2-\lambda).y}, & x > y \\ \lambda_2.(\lambda_1-\lambda).e^{-\lambda_2.y-(\lambda_1-\lambda).x}, & x < y \\ \lambda.e^{-(\lambda_1+\lambda_2-\lambda).y}, & x = y \end{cases}$$

then the following results are true.

Result 5: $E\left((X-\mu_X)(Y-\mu_Y)\right) = \dfrac{\lambda}{\lambda_1.\lambda_2.(\lambda_1+\lambda_2-\lambda)}$

Result 6: $$E\left((X-\mu_X)^2.(Y-\mu_Y)^2\right)=\frac{1}{{\lambda_1}^2.{\lambda_2}^2}+\frac{8.\lambda}{\lambda_1.\lambda_2.(\lambda_1+\lambda_2-\lambda)^3}$$

Result 7: $$E\left((X-\mu_X)^3.(Y-\mu_Y)\right)=\frac{3.\lambda}{{\lambda_1}^3.\lambda_2.(\lambda_1+\lambda_2-\lambda)}+\frac{6.\lambda}{\lambda_1.\lambda_2.(\lambda_1+\lambda_2-\lambda)^3}$$

Result 8: $$E\left((X-\mu_X).(Y-\mu_Y)^3\right)=\frac{3.\lambda}{\lambda_1.{\lambda_2}^3.(\lambda_1+\lambda_2-\lambda)}+\frac{6.\lambda}{\lambda_1.\lambda_2.(\lambda_1+\lambda_2-\lambda)^3}$$

Result 9: Kurtosis, $$\kappa=20+\frac{20.\rho^2}{(1-\rho^2)}+\frac{8.\lambda.\{2.\lambda_1.\lambda_2.(1+2.\rho^2)-3.\rho.({\lambda_1}^2+{\lambda_2}^2)\}}{(1-\rho^2)^2.(\lambda_1+\lambda_2-\lambda)^3}$$

Note that the kurtosis will be 20 when the components are independent.

4. Results

- Numerical Result:

Kurtosis Calculation (Bivariate Exponential):

The Kurtosis given by (2.20) is estimated for the data generated from the Morgenstern Copula and the Clayton Copula and in addition to the simulated data from the bivariate exponential distribution by using the following parameters.

$$\lambda_1=0.4,\qquad \lambda_2=0.3\quad,\qquad \lambda=0.1\quad,\ \alpha=0.7$$

Kurtosis Estimate (simulated bivariate exponential) = 18.036

Kurtosis (Actual) = 21.086

Kurtosis Estimate (Morgenstern Copula) = 22.585

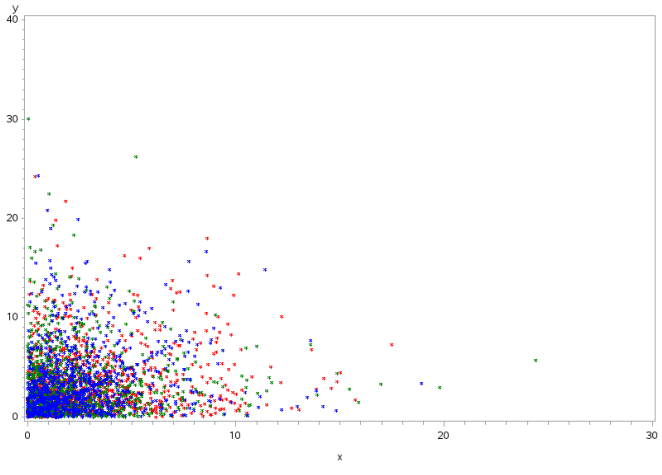
Average distance estimate (of Morgenstern Copula from actual Population) =5.631

Kurtosis Estimate (Clayton Copula) = 30.036

Average distance estimate (of Clayton Copula from actual Population) =5.927

- Graphical Result (Bivariate Exponential):

Here, we present the scatterplots based on the copulas. The blue dots in the scatterplot represent the data that was simulated from the bivariate exponential distribution with the sample size = 1000. The red dots in the scatterplot represent the data that was generated by using the Morgenstern Copula while the green dots represent the data generated from the Clayton Copula.



- Kurtosis Calculation (Bivariate Normal):

The Kurtosis given by (2.20) is estimated for the data generated from the Morgenstern Copula and the Clayton Copula and in addition to the simulated data from the bivariate normal distribution by using the following parameters.

$$\mu_1 = 2.25,\ \mu_2 = 3.33,\quad \rho = 0.175,\ \sigma_1 = 2.25,\ \sigma_2 = 3.33$$

Kurtosis Estimate (simulated bivariate normal) = 7.761

Kurtosis (Actual) = 8

Kurtosis Estimate (Morgenstern Copula) = 7.735

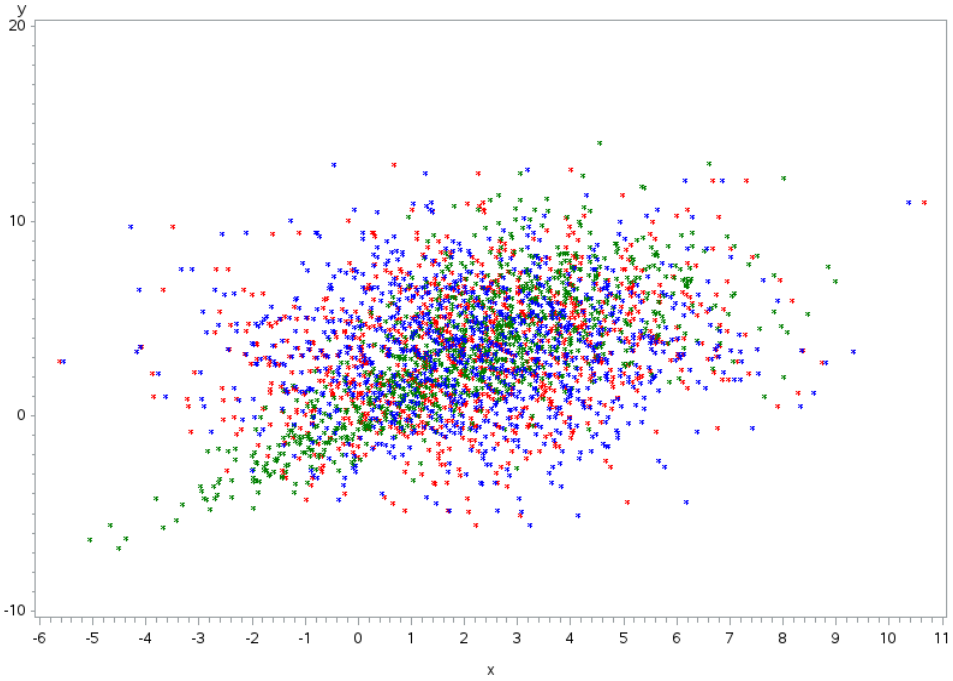
Average distance estimate (of Morgenstern Copula from actual Population) =0.515

Kurtosis Estimate (Clayton Copula) = 10.165

Average distance estimate (of Clayton Copula from actual Population) =5.637

- Graphical Result (Bivariate Normal):

Here, we again present the scatterplots based on the copulas. The blue dots in the scatterplot represent the data that was simulated from the bivariate normal distribution with the sample size = 1000. The red dots in the scatterplot represent the data that was generated by using the Morgenstern Copula while the green dots represent the data generated from the Clayton Copula.



5. Discussion and Conclusion

The Copulas allow us to model the multivariate distributions from the marginal distributions. There are many types of Copulas. The Farleigh-Gumbel-Morgenstern Copula is used in Actuarial models and also in Engineering related reliability studies. The Clayton Copula is used mostly in Finance and Marketing. This paper is focused on comparing the Farleigh-Gumbel-Morgenstern Copula with the Clayton Copula. Note that these two copulas belong to two different families with the Clayton Copula coming from the Archiemedian Family and the Farleigh-Gumbel-Morgenstern Copula from the Non-Archemedian Family. We chose the parameters so that the mean vector and the variance-covariance matrix would be the same for the bivariate exponential and the bivariate normal distributions. Note that from the numerical results, it is obvious that the Farleigh-Gumbel-Morgenstern Copula (or simply Morgenstern Copula) is slightly more accurate than the Clayton Copula for modeling both the bivariate exponential distribution and the

bivariate normal distribution. Also, the scatterplots seemed to support the numerical results based on the Kurtosis. Furthermore, from this paper it is evident that the multivariate kurtosis is a reasonable measure to compare these two Copulas.

References

Aas, K., Czado, C., & Bakken, H. (2009). Pair-copula constructions of multiple dependence, Insurance. *Mathematics and Economics, 44*(2), 182–198.

Ali, M. M., Mikhail, N. N., & Haq, M. S. (1978). A class of bivariate distributions including the bivariate logistic. *J. Multivariate Anal., 8,* 405-412.

Heinen, A., & Rengifo, E. (2007). Multivariate autoregressive modeling of time series count data using copulas. *Journal of Empirical Finance, 14,* 564 – 583.

Karlis, D., & Pedeli, X. (2013). Flexible bivariate INAR(1) processes using copulas. *Communications in Statistics-Theory and Methods, 42,* 723-740.

Low, R., Alcock, J., Brailsford, T., & Faff, R. (2013). Canonical vine copulas in the context of modern portfolio management: Are they worth it?. *Journal of Banking and Finance, 37*(8), 3085–3099.

Malkovich, J. F., & Afifi, A. A. (1973). On tests for multivariate normality. *Journal of the American Statistical Association, 68,* 176–179.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57,* 519–530.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya B, 36,* 115–128.

Miyagawa, C., Koizumi, K., & Seo, T. (2011). A new multivariate kurtosis and its asymptotic distribution. *SUT Journal of Mathematics, 47,* 55–71.

Nelson, R. B. (2006). An Introduction to Copulas. *Springer.*

Nikoloulopoulos, A., & Karlis, D. (2009). Modeling multivariate count data using copulas. *Communications in Statistics- Simulation and Computation, 39*(1), 172-187.

Nikoloulopoulos, A., & Karlis, D. (2010). Regression in a copula model for bivariate count data. *Journal of Applied Statistics, 37,* 1555-1568

Schölzel, C., & Friederichs, P. (2008). Multivariate non-normally distributed random variables in climate research – introduction to the copula approach. *Nonlinear Processes in Geophysics, 15*(5), 761–772.

Sklar, A. (1959). Fonctions de repartition a n dimensions et leurs marges. *Publications de L' Institut de Statistique de L' Universite de Paris, 8,* 229-231.

Srivastava, M. S. (1984). A measure of skewness and kurtosis and a graphical method for assessing multivariate normality. *Statistics & Probability Letters, 2,* 263-267.

Appendix:

Result 1: $E\left(\left(X-\mu_X\right)^2\left(Y-\mu_Y\right)^2\right) = {\sigma_X}^2.{\sigma_Y}^2.\left(1-\rho^2\right)+3.\rho^2.{\sigma_X}^2.{\sigma_Y}^2$

Proof: Note that, $E\left(\left(X-\mu_X\right)^2 \setminus Y\right)={\sigma_X}^2.\left(1-\rho^2\right)+\rho.\mu_X.\frac{\sigma_X}{\sigma_Y}.\left(Y-\mu_Y\right)+\rho^2.\frac{{\sigma_X}^2}{{\sigma_Y}^2}.\left(Y-\mu_Y\right)^2$

This implies that

$$E\left(\left(X-\mu_X\right)^2.\left(Y-\mu_Y\right)^2\right)=E\left\{\left(Y-\mu_Y\right)^2.E\left(\left(X-\mu_X\right)^2 \setminus Y\right)\right\}$$

$$= E\left({\sigma_X}^2.\left(1-\rho^2\right)\left(Y-\mu_Y\right)^2\right)+E\left(\rho.\mu_X.\frac{\sigma_X}{\sigma_Y}.\left(Y-\mu_Y\right)^3\right)+E\left(\rho^2.\frac{{\sigma_X}^2}{{\sigma_Y}^2}.\left(Y-\mu_Y\right)^4\right)$$

$$= {\sigma_X}^2.{\sigma_Y}^2.\left(1-\rho^2\right)+3.\rho^2.{\sigma_X}^2.{\sigma_Y}^2$$

Result 2: $E\left((X-\mu_X).(Y-\mu_Y)^3\right)=3.\rho.\sigma_X.{\sigma_Y}^3$

Proof: Note that, $E(X \setminus Y)=\mu_X+\rho.\frac{\sigma_X}{\sigma_Y}.(Y-\mu_Y)$

This implies that

$$E\left((X-\mu_X).(Y-\mu_Y)^3\right)=E\left\{(Y-\mu_Y)^3.E\left((X-\mu_X)\setminus Y\right)\right\}$$

$$= E\left\{(Y-\mu_Y)^3.\rho.\frac{\sigma_X}{\sigma_Y}.(Y-\mu_Y)\right\}$$

$$= \rho.\frac{\sigma_X}{\sigma_Y}.E(Y-\mu_Y)^4$$

$$= 3\rho.\sigma_X.{\sigma_Y}^3$$

Result 3: $E\left((X-\mu_X)^3.(Y-\mu_Y)\right)=3.\rho.{\sigma_X}^3.\sigma_Y$

Proof: Follows from symmetry.

Result 4: Kurtosis, $\kappa=8$.

Proof: Note that R.H.S of (2.15) can be written as

$$\left(1-\rho^2\right)^{-2}.3\frac{{\sigma_X}^4}{{\sigma_X}^4}+\left(1-\rho^2\right)^{-2}.3\frac{{\sigma_Y}^4}{{\sigma_Y}^4}+4\frac{\rho^2}{{\sigma_X}^2.{\sigma_Y}^2}.{\sigma_X}^2.{\sigma_Y}^2.\left(1-\rho^2\right)^{-1}$$

$$+4.\left(\frac{1}{{\sigma_X}^2}\right).\left(\frac{-\rho}{\sigma_X.\sigma_Y}\right).\left(1-\rho^2\right)^{-2}.3\rho.{\sigma_X}^3.\sigma_Y \quad +4.\left(\frac{1}{{\sigma_Y}^2}\right).\left(\frac{-\rho}{\sigma_X.\sigma_Y}\right).\left(1-\rho^2\right)^{-2}.3\rho.{\sigma_Y}^3.\sigma_X$$

$$+\ 2.\left(\frac{1}{{\sigma_X}^2}\right).\left(\frac{1}{{\sigma_Y}^2}\right)\left(1-\rho^2\right)^{-2}.{\sigma_X}^2.{\sigma_Y}^2\left(1-\rho^2\right)\ +\ 2.\left(\frac{1}{{\sigma_X}^2}\right).\left(\frac{1}{{\sigma_Y}^2}\right)\left(1-\rho^2\right)^{-2}.{\sigma_X}^2.{\sigma_Y}^2.3\rho^2$$

$$+\ 12.\left(\frac{\rho^2}{{\sigma_X}^2.{\sigma_Y}^2}\right)\left(1-\rho^2\right)^{-2}.\rho^2.{\sigma_X}^2.{\sigma_Y}^2$$

$$= \left(1-\rho^2\right)^{-2}.\left(8-16\rho^2+8\rho^4\right)$$

$$= 8.\left(1-\rho^2\right)^{-2}\left(1-\rho^2\right)^2$$

$$= 8$$

Result 5: $E\left((X-\mu_X)(Y-\mu_Y)\right)=\frac{\lambda}{\lambda_1+\lambda_2-\lambda}$

Result 6: $E\left((X-\mu_X)^2(Y-\mu_Y)^2\right)=\frac{1}{\lambda_1^{\,2}.\lambda_2^{\,2}}+\frac{8.\lambda}{\lambda_1.\lambda_2.(\lambda_1+\lambda_2-\lambda)^3}$

Proof:

Note that $E\left((X-\mu_X)^2(Y-\mu_Y)^2\right)= E\left(X^2.Y^2\right)-2.\mu_X.E\left(X.Y^2\right)+\mu_X^{\,2}.E\left(Y^2\right)$

$$-2.\mu_Y.E\left(X^2.Y\right)+4.\mu_X.\mu_Y.E\left(X.Y\right)$$

$$-2.\mu_X^{\,2}.\mu_Y.E\left(Y\right)+\mu_Y^{\,2}.E\left(X^2\right)$$

$$-2.\mu_X.\mu_Y^{\,2}.E\left(X\right)+\mu_X^{\,2}.\mu_Y^{\,2} \qquad (A.1)$$

One can easily show that for the bivariate exponential density,

(1). $E\left(X.Y\right)=\frac{(\lambda_1+\lambda_2)}{\lambda_1.\lambda_2.(\lambda_1+\lambda_2-\lambda)}$

(2). $E\left(X^2.Y\right)=\frac{6.(\lambda_1-\lambda)}{(\lambda_1+\lambda_2-\lambda)^4}+\frac{2.(\lambda_1-\lambda)}{\lambda_2.(\lambda_1+\lambda_2-\lambda)^3}+\frac{6.(\lambda_2-\lambda)}{(\lambda_1+\lambda_2-\lambda)^4}$

$$+\frac{4.(\lambda_2-\lambda)}{\lambda_1.(\lambda_1+\lambda_2-\lambda)^3}+\frac{2.(\lambda_2-\lambda)}{\lambda_1^{\,2}.(\lambda_1+\lambda_2-\lambda)^2}+\frac{6.\lambda}{(\lambda_1+\lambda_2-\lambda)^4}$$

(3). $E\left(X.Y^2\right)=\frac{6.(\lambda_2-\lambda)}{(\lambda_1+\lambda_2-\lambda)^4}+\frac{2.(\lambda_2-\lambda)}{\lambda_1.(\lambda_1+\lambda_2-\lambda)^3}+\frac{6.(\lambda_1-\lambda)}{(\lambda_1+\lambda_2-\lambda)^4}$

$$+\frac{4.(\lambda_1-\lambda)}{\lambda_2.(\lambda_1+\lambda_2-\lambda)^3}+\frac{2.(\lambda_1-\lambda)}{\lambda_2^{\,2}.(\lambda_1+\lambda_2-\lambda)^2}+\frac{6.\lambda}{(\lambda_1+\lambda_2-\lambda)^4}$$

(4). $E\left(X^2.Y^2\right)=\frac{24}{(\lambda_1+\lambda_2-\lambda)^4}+\frac{12.(\lambda_1-\lambda)}{\lambda_2.(\lambda_1+\lambda_2-\lambda)^4}+\frac{12.(\lambda_2-\lambda)}{\lambda_1.(\lambda_1+\lambda_2-\lambda)}$

$$+\frac{4.(\lambda_1-\lambda)}{\lambda_2^{\,2}.(\lambda_1+\lambda_2-\lambda)^3}+\frac{4.(\lambda_2-\lambda)}{\lambda_1^{\,2}.(\lambda_1+\lambda_2-\lambda)^3}$$

By substituting these equations in $(A.1)$, we get Result 6.

Result 7: $E\left((X-\mu_X)^3.(Y-\mu_Y)\right)=\frac{3.\lambda}{\lambda_1^{\,3}.\lambda_2.(\lambda_1+\lambda_2-\lambda)}+\frac{6.\lambda}{\lambda_1.\lambda_2.(\lambda_1+\lambda_2-\lambda)^3}$

Proof:

Note that

$$E\left(X^3.Y\right)=\frac{24}{\left(\lambda_1+\lambda_2-\lambda\right)^4}+\frac{18.\left(\lambda_2-\lambda\right)}{\lambda_1.\left(\lambda_1+\lambda_2-\lambda\right)^4}+\frac{12.\left(\lambda_2-\lambda\right)}{{\lambda_1}^2.\left(\lambda_1+\lambda_2-\lambda\right)^3}+\frac{6.\left(\lambda_1-\lambda\right)}{\lambda_2.\left(\lambda_1+\lambda_2-\lambda\right)^4}$$

$$+\frac{6.\left(\lambda_2-\lambda\right)}{{\lambda_1}^3.\left(\lambda_1+\lambda_2-\lambda\right)^2}$$

By substituting the above equations, we get the result.

Result 8: $E\left(\left(X-\mu_X\right).\left(Y-\mu_Y\right)^3\right)=\frac{3.\lambda}{\lambda_1.{\lambda_2}^3.\left(\lambda_1+\lambda_2-\lambda\right)}+\frac{6.\lambda}{\lambda_1.\lambda_2.\left(\lambda_1+\lambda_2-\lambda\right)^3}$

Proof: Follows by interchanging the variables X and Y.

Result 9: Kurtosis, $\kappa=20+\frac{20.\rho^2}{\left(1-\rho^2\right)}+\frac{8.\lambda.\left\{2.\lambda_1.\lambda_2.\left(1+2.\rho^2\right)-3.\rho.\left({\lambda_1}^2+{\lambda_2}^2\right)\right\}}{\left(1-\rho^2\right)^2.\left(\lambda_1+\lambda_2-\lambda\right)^3}$

Proof: Follows from combining all the previous results.

International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

The Principle of Indifference Does Not Lead to Contradictions

Wolfgang Tschirk[1]

[1] mathecampus, Vienna, Austria

Correspondence: Wolfgang Tschirk, mathecampus, Mariahilfer Straße 136, 1150 Wien, Austria. Tel: 43-680-126-8416. E-mail: wolfgang.tschirk@mathecampus.at

Received: May 6, 2016 Accepted: May 20, 2016 Online Published: June 22, 2016

doi:10.5539/ijsp.v5n4p79 URL: http://dx.doi.org/10.5539/ijsp.v5n4p79

Abstract

The Principle of Indifference says that if there are a finite number of propositions and a state of knowledge according to which none of the propositions is more plausible than any other, then, conditional on that knowledge, all of them have the same probability. Most researchers reject the principle because there exist counterexamples believed to prove that it leads to contradictions. We analyse three examples representative of the objections to the principle and show that, rather than disproving it, they suffer from a common error in applying it. From this and the fact that the Principle of Indifference complies with plausible reasoning we conclude that it does not lead to contradictions.

Keywords: Principle of Indifference, Principle of Insufficient Reason

1. Introduction

The Principle of Indifference, which dates back to Jakob Bernoulli, can be expressed as follows:

> If, given some proposition C, one of the propositions $A_1, \cdots, A_n$ must be true and the others must be false and none of the A_i is more plausible than any other, then, conditional on C, all A_i have the same probability $p(A_i|C) = 1/n$.

C is often referred to as someone's (prior) knowledge, which may or may not contain reasons to favour some proposition over some other. The rule became known as the *Principle of Insufficient Reason*. In 1921, Keynes introduced the term *Principle of Indifference* (Keynes, 1921). Keynes raised doubt whether the principle was valid and presented counterexamples intended to prove that it leads to contradictions and is therefore invalid. Other authors have commented on these examples and added their own.

Due to the presumed contradictions, most researchers have rejected and still reject the principle (Bartelborth, 2012; Carnap, 1966; Howie, 2002; Robert, 2007; van Fraassen, 1989). However, in the 1950s, Jaynes derived it from a set of desiderata for plausible reasoning (Jaynes 1958, 2003). In his probability theory, the principle is a proven theorem; it cannot be contradictory there, unless the whole theory were contradictory. If it is not, it must be possible to eliminate the contradictions on the ground of Jaynes' desiderata. This is what we are going to undertake.

2. Counterexamples to the Principle

We start with three examples Keynes gave in order to demonstrate where the Principle of Indifference fails. For ease of presentation, we adapt the wording without substantially altering the content.

1. *The colour of the book*: A person who knows nothing about the colour of a certain book must, according to the Principle of Indifference, assign the proposition $R :=$ *The book is red* the same probability as its contrary $\overline{R} =$ *The book is not red*, namely $p(R) = 1/2$. For the same reason the person must assign the proposition $B :=$ *The book is blue* the probability $p(B) = 1/2$ and the proposition $G :=$ *The book is green* the probability $p(G) = 1/2$. But now the proposition $R \vee B \vee G =$ *The book is red or blue or green* would have the probability $p(R \vee B \vee G) = p(R) + p(B) + p(G) = 3/2$, which is impossible as probabilities do not exceed unity.

2. *Sizes of countries*: A person who knows nothing about the sizes of countries must, according to the Principle of Indifference, assign the proposition $E :=$ *England is bigger than France* the same probability as its contrary $\overline{E} =$ *England is not bigger than France*, namely $p(E) = 1/2$. For the same reason the person must assign the proposition $B :=$ *The British Isles are bigger than France* the probability $p(B) = 1/2$. But this is impossible, since England is only a part of the British Isles; the Isles must surpass France in size with higher probability than England alone.

3. *Specific volume vs. specific density*: A person who only knows that the specific volume of some substance lies between 1 and 3 (measured in some unit) must, according to the Principle of Indifference, assign the propositions $V :=$ *The specific volume lies between 1 and 2* and its contrary $\overline{V} =$ *The specific volume lies between 2 and 3* the probabilities $p(V) = p(\overline{V}) = 1/2$. The reciprocal of the specific volume is the specific density. The person only knows that it lies between 1/3 and 1, and must therefore assign the proposition $D :=$ *The specific density lies between 1/3 and 2/3* the same probability as its contrary $\overline{D} =$ *The specific density lies between 2/3 and 1*, namely $p(D) = 1/2$. But this is impossible, since D says (though in other words) that the specific volume lies between 3/2 and 3 and must be more probable than $\overline{V}$.

As far as we know, every serious objection to the Principle of Indifference is a modification of one of these examples, corresponding to one of the following patterns:

1. The pattern of *The colour of the book*: If one out of two equally probable propositions (*The book is red* and *The book is not red*) is dissected (as *The book is not red* is dissected into *The book is blue*, *The book is green*, and maybe others, too), the resulting ones (*The book is red*, *The book is blue*, *The book is green*, etc.) cannot be equally probable; yet the Principle of Indifference says they are equally probable.
2. The pattern of *Sizes of countries*: A proposition (*England is bigger than France*) cannot have the same probability as an obviously more probable one (*The British Isles are bigger than France*); yet the Principle of Indifference says it has the same probability.
3. The pattern of *Specific volume vs. specific density*: Uniform distribution on one scale (equal probabilities for equally large intervals on the specific volume scale) leads to non-uniform distribution on a different scale (non-equal probabilities for equally large intervals on the specific density scale); yet the Principle of Indifference requires uniform distributions on both scales.

Among the objections that correspond to these patterns are: the *partitioning incoherence of Laplace's equiprobability* claimed by Robert (2007), corresponding to pattern 1; Carnap's *life on Mars* (Carnap, 1966), corresponding to pattern 2; Keynes' *urn with black and white balls* (Keynes, 1921), van Fraassen's *cube factory* (van Fraassen, 1989) and von Mises' *wine/water-paradox* (Mikkelson, 2004), all corresponding to pattern 3.

At first sight, the examples seem to disprove the Principle of Indifference; however, we will find that they suffer from a common error in applying it.

3. Desiderata for Plausible Reasoning

We base our analysis on Jaynes' desiderata for plausible reasoning, i.e. for assigning degrees of plausibility to propositions (in Jaynes' terminology, such plausibility assignment is called a *conclusion*):

(I) Degrees of plausibility are represented by real numbers.

(II) Plausible reasoning qualitatively corresponds with common sense.

(IIIa) If a conclusion can be reasoned in more than one way, every way leads to the same result.

(IIIb) Every conclusion is based on all available knowledge.

(IIIc) Equivalent states of knowledge lead to equivalent conclusions.

Desideratum (I), together with the convention that a greater plausibility shall correspond to a greater number, guarantees that 1) any two propositions A and B can be compared with respect to plausibility such that either A is more plausible than B, or B is more plausible than A, or A and B are equally plausible, and 2) if A is more plausible than B, and B is more plausible than C, then A is more plausible than C.

From desideratum (II), the following rule, which we call *implication rule*, can be derived:

If, given C, A implies B and B does not imply A, then, given C, B is more plausible than A.

It corresponds with common sense because, given C, B is true whenever A is true, but B can even be true when A is false.

Desiderata (IIIa)–(IIIc) ensure that reasoning is consistent.

4. Analysis of the Counterexamples

4.1 Example 1: The colour of the book

We start the examination by asking for the knowledge based on which the probabilities are assigned. Let C_R be some knowledge according to which R is equally plausible as $\overline{R}$; then, following the Principle of Indifference, the probability of R, given C_R, is $1/2$: $p(R|C_R) = 1/2$. Let C_B be some knowledge which makes B equally plausible as $\overline{B}$, and C_G some knowledge which makes G equally plausible as $\overline{G}$, then $p(B|C_B) = 1/2$ and $p(G|C_G) = 1/2$. Now each of the colours has a probability of $1/2$ to be the colour of the book, but these probabilities are conditional on *different* knowledge; and

$$p(R|C_R) + p(B|C_B) + p(G|C_G) = 3/2$$

does not violate any rule of probability (as C_R, C_B and C_G are different, the sum on the left side of the equation is not a probability). This has already been observed by Jeffreys in his review of Keynes' work (Jeffreys, 1922).

As long as C_R, C_B and C_G are not identical, Example 1 does not disprove the Principle of Indifference. The principle would fail only if there existed a state of knowledge which made R equally plausible as $\overline{R}$ and, at the same time, B equally plausible as $\overline{B}$ and G equally plausible as $\overline{G}$. Now we prove by contradiction that such knowledge cannot exist.

Assume that there exists some knowledge C according to which each of the propositions R, B and G is equally plausible as its respective contrary. First we note that C allows red, blue and green to be possible colours of the book (if, for instance, red were impossible, then R would be less plausible than $\overline{R}$). Then, given C, R implies $\overline{B}$ but $\overline{B}$ does not imply R (a red book is clearly non-blue, whereas a non-blue book does not have to be red); following the implication rule, $\overline{B}$ is more plausible than R, conditional on C. Using Jaynes' notation, where $A|B$ stands for the plausibility of A, given B, we have thus found

$$\overline{B}|C > R|C\,.$$

An analogous reasoning shows that $\overline{R}$ is more plausible than B, conditional on C:

$$\overline{R}|C > B|C\,.$$

Now remember that C is assumed to make R and $\overline{R}$ equally plausible:

$$R|C = \overline{R}|C\,.$$

Putting these relations together, we arrive at

$$\overline{B}|C > R|C = \overline{R}|C > B|C\,,$$

which contradicts the assumption that C makes B and $\overline{B}$ equally plausible. Therefore, a state of knowledge according to which each of the propositions R, B and G were equally plausible as its respective contrary does not exist.

From the above follows that Example 1 does not disprove the Principle of Indifference; its paradox results from the assumption of prior knowledge which cannot exist.

4.2 Example 2: Sizes of countries

As in Example 1, we start by asking for the knowledge based on which the probabilities are assigned. Let C_E be some knowledge according to which E is equally plausible as $\overline{E}$; then $p(E|C_E) = 1/2$. Let C_B be some knowledge which makes B equally plausible as $\overline{B}$; then $p(B|C_B) = 1/2$. We thus obtain

$$p(E|C_E) = p(B|C_B)\,;$$

but although England is a part of the British Isles, this equality is not impossible since the probabilities are conditional on different knowledge.

The Principle of Indifference would fail on Example 2 only if there existed a state of knowledge which contained the information that England is a part of the British Isles and, at the same time, made E equally plausible as $\overline{E}$ and B equally plausible as $\overline{B}$. We prove by contradiction that such knowledge cannot exist.

Assume that there exists some knowledge C according to which England is a part of the British Isles and each of the propositions E and B is equally plausible as its respective contrary. Then, given C, E implies B but B does not imply E (if England is bigger than France, then the British Isles also are; however, from the premise that the British Isles are bigger than France, one cannot conclude that England also is); following the implication rule, B is more plausible than E, conditional on C:

$$B|C > E|C\,.$$

On the other hand, given C, $\overline{B}$ implies $\overline{E}$ but $\overline{E}$ does not imply $\overline{B}$ (if the British Isles are not bigger than France, then England is also not; however, from the premise that England is not bigger than France, one cannot conclude that the British Isles are not); following the implication rule, $\overline{E}$ is more plausible than $\overline{B}$, conditional on C:

$$\overline{E}|C > \overline{B}|C\,.$$

Now remember that C is assumed to make E and $\overline{E}$ equally plausible:

$$E|C = \overline{E}|C\,.$$

Putting these relations together, we arrive at

$$B|C > E|C = \overline{E}|C > \overline{B}|C\,,$$

which contradicts the assumption that C makes B and $\overline{B}$ equally plausible. Therefore, a state of knowledge according to which England is a part of the British Isles and each of the propositions E and B were equally plausible as its respective contrary does not exist.

It follows that Example 2 does not disprove the Principle of Indifference; its paradox results from the assumption of prior knowledge which cannot exist.

4.3 Example 3: Specific volume vs. specific density

Again we ask for the knowledge based on which the probabilities are assigned. Let C_V be some knowledge according to which V is equally plausible as $\overline{V}$; then $p(\overline{V}|C_V) = 1/2$. Let C_D be some knowledge which makes D equally plausible as $\overline{D}$; then $p(D|C_D) = 1/2$. We thus obtain

$$p(\overline{V}|C_V) = p(D|C_D)\,;$$

but although specific volume and specific density are reciprocals of each other, this equality is not impossible since the probabilities are conditional on different knowledge.

The Principle of Indifference would fail on Example 3 only if there existed a state of knowledge which contained the information that specific volume and specific density are reciprocals of each other and, at the same time, made V equally plausible as $\overline{V}$ and D equally plausible as $\overline{D}$. We prove by contradiction that such knowledge cannot exist.

Assume that there exists some knowledge C according to which specific volume and specific density are reciprocals of each other and each of the propositions V and D is equally plausible as its respective contrary. Then, given C, $\overline{V}$ (*The specific volume lies between 2 and 3*) implies D (equivalent to *The specific volume lies between 3/2 and 3*) but D does not imply $\overline{V}$; following the implication rule, D is more plausible than $\overline{V}$, conditional on C:

$$D|C > \overline{V}|C\,.$$

On the other hand, given C, $\overline{D}$ (*The specific density lies between 2/3 and 1*) implies V (equivalent to *The specific density lies between 1/2 and 1*) but V does not imply $\overline{D}$; following the implication rule, V is more plausible than $\overline{D}$, conditional on C:

$$V|C > \overline{D}|C\,.$$

Now remember that C is assumed to make V and $\overline{V}$ equally plausible:

$$V|C = \overline{V}|C\,.$$

Putting these relations together, we arrive at

$$D|C > \overline{V}|C = V|C > \overline{D}|C\,,$$

which contradicts the assumption that C makes D and $\overline{D}$ equally plausible. Therefore, a state of knowledge according to which specific volume and specific density are reciprocals of each other and each of the propositions V and D were equally plausible as its respective contrary does not exist.

It follows that Example 3 does not disprove the Principle of Indifference; its paradox results from the assumption of prior knowledge which cannot exist.

5. Results

We have analysed three examples representative of the objections to the Principle of Indifference and found that, rather than disproving the principle, they suffer from a common error in applying it. The error can be described as follows: First,

a state of knowledge is assumed which contains no reason to favour any proposition (out of a certain set of propositions) over its contrary; second, based on this knowledge, each proposition is assigned a probability in accordance with the Principle of Indifference; third, the probability assignment is shown to be contradictory and the contradiction is attributed to the Principle of Indifference. However, as we have proven for each example, a state of knowledge which contains no reason to favour any of said propositions over its contrary and, at the same time, renders the resulting probability assignment inconsistent is impossible.

Therefore, none of the examples disproves the Principle of Indifference; the contradiction is always caused by the assumption of prior knowledge which cannot exist.

6. Discussion and Conclusions

The Principle of Indifference is a proven theorem in a probability theory that emerged from a set of desiderata for plausible reasoning. On the other hand, there exist counterexamples believed to show that it leads to contradictions. Being proven, the principle cannot be contradictory, unless the whole theory were contradictory. If it is not, it must be possible to eliminate the contradictions on the ground of the desiderata the theory is built upon.

Based on these desiderata, we have analysed three counterexamples and found that in none of them the principle fails. Whenever a contradiction arises, it is caused by the erroneous assumption that there be no reason to favour any relevant proposition over its contrary. In each of the examples we have found such reason; it has never been an *empirical* reason, it was always a *logical* one: Whatever prior knowledge one has with respect to the colour of a certain book and whether or not it contains reasons to favour or disfavour red, to favour or disfavour blue or to favour or disfavour green – it is logically impossible that each of the propositions *The book is red*, *The book is blue* and *The book is green* be exactly as plausible as its negation; and analogous arguments hold with respect to the other examples.

Some opponents to the Principle of Indifference consider only empirical knowledge as relevant knowledge. In his *Sizes of countries* example, Keynes explicitely rejects a way out of the contradiction based on logic. However, the use of logical evidence in plausibility assignments is mandatory because of two reasons: First, it is required by desideratum (IIIb); second, consistent reasoning would be impossible if the rules of logic could be arbitrarily ignored, in particular, fulfillment of the desiderata (IIIa) and (IIIc) would not be guaranteed.

As far as we know, every serious objection to the Principle of Indifference corresponds to the pattern of one of the examples we have analysed. If this is true, then the principle does not lead to contradictions at all.

References

Bartelborth, T. (2012). *Die erkenntnistheoretischen Grundlagen induktiven Schließens*. Ebook, URN=http://nbn-resolving.de/urn:nbn:de:bsz:15-qucosa-84565

Carnap, R. (1966). *Philosophical Foundations of Physics. An Introduction to the Philosophy of Science*. New York, NY: Basic Books.

Howie, D. (2002). *Interpreting Probability. Controversies and Developments in the Early Twentieth Century*. New York, NY: Cambridge University Press.

Jaynes, E. T. (1958). *Probability Theory in Science and Engineering*. Dallas, TX: Socony Mobile Oil Company.

Jaynes, E. T. (2003). *Probability Theory. The Logic of Science*. New York, NY: Cambridge University Press.

Jeffreys, H. (1922). The Theory of Probability. *Nature* 109, 132-133. http://dx.doi.org/10.1038/109132a0

Keynes, J. M. (1921). *A Treatise on Probability*. London, England: MacMillan & Co.

Mikkelson, J. M. (2004). Dissolving the Wine/Water Paradox. *British Journal for the Philosophy of Science* 55, 137–145. http://dx.doi.org/10.1093/bjps/55.1.137

Robert, C. P. (2007). *The Bayesian Choice. From Decision-Theoretic Foundations to Computational Implementation*. New York, NY: Springer.

Van Fraassen, B. C. (1989). *Laws and Symmetry*. New York, NY: Oxford University Press.

Copyrights

International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

On the Existence Conditions for Balanced Fractional 2^m Factorial Designs of Resolution $\mathrm{R}^*(\{1\}|\Omega_\ell)$ with $N < \nu_\ell(m)$

Yoshifumi Hyodo[1,2], Masahide Kuwada[2], Hiromu Yumiba[2]

[1] Graduate School of Informatics, Okayama University of Science, Okayama, Japan

[2] International Institute for Natural Sciences, Kurashiki, Japan

Correspondence: Hiromu Yumiba, International Institute for Natural Sciences, Kurashiki, 710-0821, Japan.
E-mail: yumiba@edu.kake.ac.jp

Received: January 29, 2016 Accepted: February 22, 2016 Online Published: June 22, 2016

doi:10.5539/ijsp.v5n4p84 URL: http://dx.doi.org/10.5539/ijsp.v5n4p84

Abstract

We consider a fractional 2^m factorial design derived from a simple array (SA) such that the $(\ell+1)$-factor and higher-order interactions are assumed to be negligible, where $2\ell \le m$. Under these situations, if at least the main effect is estimable, then a design is said to be of resolution $\mathrm{R}^*(\{1\}|\Omega_\ell)$. In this paper, we give a necessary and sufficient condition for an SA to be a balanced fractional 2^m factorial design of resolution $\mathrm{R}^*(\{1\}|\Omega_\ell)$ for $\ell = 2, 3$, where the number of assemblies is less than the number of non-negligible factorial effects. Such a design is concretely characterized by the suffixes of the indices of an SA.

Keywords: association algebra, balanced fractional factorial design, estimable parametric function, factorial effect, resolution, simple array

1. Introduction

As a generalization of an orthogonal array, the concept of a balanced array (BA) was first introduced by Chakravarti (1956) as a partially BA. However it is a generalization of the BIB design rather than of the PBIB design. Thus Srivastava and Chopra (1971) called it by BA. A BA of strength t, size N, m constraints, two symbols and index set $\{\mu_i^{(t)} \mid 0 \le i \le t\}$ is briefly written by $\mathrm{BA}(N, m, 2, t; \{\mu_i^{(t)}\})$. In particular, a BA of strength $t = m$ is called a simple array (SA) (see Shirakura, 1977), and it is written by $\mathrm{SA}(m; \{\lambda_x\})$ for brevity, where $\lambda_x = \mu_x^{(m)}$. When $t < m$, a BA of strength t does not always exist for given indices $\mu_i^{(t)}$. On the other hand, an SA always exists for any λ_x and any m. The existence conditions for a BA of strength t were given by Srivastava (1972) for $m = t+1, t+2$, and Shirakura (1977) for $m = t+3$. If the variance-covariance matrix of the estimators of the factorial effects to be of interest is invariant under any permutation on the factors, then a design is said to be balanced. Under certain conditions, a BA of strength 2ℓ turns out to be a balanced fractional 2^m factorial (2^m-BFF) design of resolution $2\ell+1$ (see for $\ell = 2$, Srivastava, 1970, and for general ℓ, Yamamoto *et al*., 1975), where $2\ell \le m$. The characteristic roots of the information matrix of a 2^m-BFF design of resolution V, i.e., $\ell = 2$, were obtained by Srivastava and Chopra (1971). By using the triangular multidimensional partially balanced (TMDPB) association scheme and its algebra, their results were generalized by Yamamoto *et al*. (1976) and Hyodo (1992) for a resolution $2\ell+1$ design, where $2\ell \le m$ and $m < 2\ell \le 2m$, respectively. The concept of the MDPB association scheme, which is a generalization of an ordinary association scheme (e.g., Bailey, 2004), was introduced by Bose and Srivastava (1964). The existence conditions for a BA of strength 2ℓ to be a 2^m-BFF design of resolution 2ℓ for general ℓ were obtained by Shirakura (1975,1980). Some algebraic properties of the information matrix of a fractional 2^m factorial (2^m-FF) design derived from an SA were investigated by Hyodo and Yamamoto (1988) and Hyodo (1989). As the extension of the concept of resolution, Yamamoto and Hyodo (1984) discussed the extended concept of resolution for 2^m fractions.

Definition 1.1. Under the assumption that the $(\ell+1)$-factor and higher-order interactions are negligible, if the p_1-factor, the p_2-factor,$\cdots$, and the p_r-factor interactions are estimable, where $0 \le p_1 < p_2 < \cdots < p_r \le \ell$, and furthermore if the

remaining interactions are not estimable (including the general mean and the main effect), then a design is said to be of resolution R($\{p_1, p_2, \cdots, p_r\}|\Omega_\ell$), where $\Omega_\ell = \{0, 1, \cdots, \ell\}$. In particular, when $p_i = i - 1\,(1 \le i \le r = \ell + 1)$, it is of resolution $2\ell + 1$, and when $p_i = i\,(1 \le i \le r = \ell - 1)$ (or $p_i = i - 1\,(1 \le i \le r = \ell)$), it is of resolution 2ℓ.

By relaxing the conditions of Definition 1.1, we give the following definition of resolution:

Definition 1.2. Under the same assumptions as Definition 1.1, if at least the p_1-factor, the p_2-factor,$\cdots$, and the p_r-factor interactions are estimable, where $0 \le p_1 < p_2 < \cdots < p_r \le \ell$, then a design is said to be of resolution R*($\{p_1, p_2, \cdots, p_r\}|\Omega_\ell$).

Note that the set of resolution R($\{p_1, p_2, \cdots, p_r\}|\Omega_\ell$) designs is a subset of resolution R*($\{p_1, p_2, \cdots, p_r\}|\Omega_\ell$) designs. For example, a resolution R*($\{1\}|\Omega_3$) design is of resolution R($\omega|\Omega_3$), where $\omega = \{1\}$, $\{0, 1\}$, $\{1, 2\}$, $\{1, 3\}$, $\{0, 1, 2\}$, $\{0, 1, 3\}$, $\{1, 2, 3\}$ or $\{0, 1, 2, 3\}$. Here when a design is derived from an SA, where the number of assemblies (or treatment combinations) is less than the number of non-negligible factorial effects, there does not exist a resolution R($\{1, 2\}|\Omega_3$), R($\{1, 3\}|\Omega_3$), R($\{0, 1, 3\}|\Omega_3$), R($\{1, 2, 3\}|\Omega_3$) or R($\{0, 1, 2, 3\}|\Omega_3$) design (see Table 4.1 in latter).

In a practical experiment, the most interesting factorial effect may be the main effect, next may be the two-factor interaction, and so on. Using the algebraic structure of the TMDPB association scheme and the matrix equations, Kuwada *et al.* (2003) obtained a 2^m-BFF design of resolution R*($\{1\}|\Omega_3$). However their results are very complex. A necessary and sufficient condition for an SA to be a 2^m-BFF design of resolution $2\ell + 1$ for general ℓ has been obtained by Hyodo *et al.* (2015), where $2\ell \le m$.

In this paper, we consider a 2^m-BFF design derived from an SA such that the number of assemblies is less than the number of factorial effects up to the ℓ-factor interaction, where $\ell = 2, 3$. Under these situations, using the suffixes of the indices λ_x of an SA, we give a necessary and sufficient condition for an SA to be a 2^m-BFF design of resolution R*($\{1\}|\Omega_2$), i.e., of resolution IV, and also we rewrite a necessary and sufficient condition for an SA to be a 2^m-BFF design of resolution R*($\{1\}|\Omega_3$).

2. Preliminaries

We consider a 2^m-FF design T with m factors and N assemblies, where $2\ell \le m$, and the $(\ell + 1)$-factor and higher-order interactions are assumed to be negligible. Then the linear model is given by $\boldsymbol{y}(T) = E_T\boldsymbol{\theta} + \boldsymbol{e}_T$, where $\boldsymbol{y}(T)$ is an $N \times 1$ observation vector, E_T is the $N \times \nu_\ell(m)$ design matrix, $\boldsymbol{\theta}' = (\boldsymbol{\theta}_0'; \boldsymbol{\theta}_1'; \cdots; \boldsymbol{\theta}_\ell')$, and $\boldsymbol{e}_T$ is an $N \times 1$ error vector with mean $\boldsymbol{0}_N$ and variance-covariance matrix $\sigma^2 I_N$. Here $\boldsymbol{\theta}_0$, $\boldsymbol{\theta}_1$, $\cdots$, and $\boldsymbol{\theta}_\ell$ are the general mean, the vector of the main effect,$\cdots$, and the vector of the ℓ-factor interaction, respectively, $\nu_\ell(m) = \binom{m}{0} + \binom{m}{1} + \cdots + \binom{m}{\ell}$, and I_p is the identity matrix of order p. The normal equations for estimating $\boldsymbol{\theta}$ are given by $M_T\hat{\boldsymbol{\theta}} = E_T'\boldsymbol{y}(T)$, where $M_T = E_T'E_T$ is the information matrix of order $\nu_\ell(m)$. If M_T is non-singular, then T is of resolution $2\ell + 1$.

Let $A_\alpha^{(u,v)}(= A_\alpha^{(v,u)\prime})$ $(0 \le \alpha \le u \le v \le \ell)$ be the local association matrices of size $\binom{m}{u} \times \binom{m}{v}$ of the TMDPB association scheme, and further let $A_\beta^{\#(u,v)}(= A_\beta^{\#(v,u)\prime})$ $(0 \le \beta \le u \le v \le \ell)$ be the matrices of size $\binom{m}{u} \times \binom{m}{v}$ (see Yamamoto *et al.*, 1976), where the relation between $A_\alpha^{(u,v)}$ and $A_\beta^{\#(u,v)}$ is given by

$$A_\alpha^{(u,v)} = \sum_{\beta=0}^{u} z_{\beta\alpha}^{(u,v)} A_\beta^{\#(u,v)} \quad \text{for } 0 \le \alpha \le u \le v$$

and

$$A_\beta^{\#(u,v)} = \sum_{\alpha=0}^{u} z_{(u,v)}^{\beta\alpha} A_\alpha^{(u,v)} \quad \text{for } 0 \le \beta \le u \le v.$$

Here

$$z_{\beta\alpha}^{(u,v)} = \sum_{b=0}^{\alpha} (-1)^{\alpha-b} \binom{u-\beta}{b}\binom{u-b}{u-\alpha}\binom{m-u-\beta+b}{b} \sqrt{\binom{m-u-\beta}{v-u}\binom{v-\beta}{v-u}} \Big/ \binom{v-u+b}{b} \quad \text{for } u \le v \tag{1}$$

and

$$z_{(u,v)}^{\beta\alpha} = \phi_\beta z_{\beta\alpha}^{(u,v)} \Big/ \left\{\binom{m}{u}\binom{u}{\alpha}\binom{m-u}{v-u+\alpha}\right\} \quad \text{for } u \le v$$

(see Shirakura and Kuwada, 1976, and Yamamoto *et al*., 1976), where $\phi_\beta = \binom{m}{\beta} - \binom{m}{\beta-1}$. Then some properties of $A_\beta^{\#(u,v)}$ are cited in the following:

$$A_\beta^{\#(u,w)} A_\gamma^{\#(w,v)} = \delta_{\beta\gamma} A_\beta^{\#(u,v)},$$
$$\sum_{\beta=0}^{u} A_\beta^{\#(u,u)} = I_{\binom{m}{u}} \tag{2}$$

and

$$\mathrm{rank}\{A_\beta^{\#(u,v)}\} = \phi_\beta,$$

where $\delta_{\beta\gamma}$ is the Kronecker delta.

Let $D_\alpha^{(u,v)} (= D_\alpha^{(v,u)\prime})$ $(0 \le \alpha \le u \le v \le \ell)$ and $D_\beta^{\#(u,v)} (= D_\beta^{\#(v,u)\prime})$ $(0 \le \beta \le u \le v \le \ell)$ be the matrices of order $\nu_\ell(m)$ such that the $(u+1)$-th row block and the $(v+1)$-th column block of $D_\alpha^{(u,v)}$ and $D_\beta^{\#(u,v)}$ are given by $A_\alpha^{(u,v)}$ and $A_\beta^{\#(u,v)}$, respectively, and zero at elsewhere. Then the information matrix M_T is given by

$$M_T = \sum_{\beta=0}^{\ell} \sum_{u=0}^{\ell-\beta} \sum_{v=0}^{\ell-\beta} \kappa_\beta^{u,v} D_\beta^{\#(u+\beta,v+\beta)}$$

(see Yamamoto *et al*., 1976), where T is a BA$(N, m, 2, 2\ell; \{\mu_i^{(2\ell)}\})$. Here the relation between $\kappa_\beta^{u,v}$ and $\mu_i^{(2\ell)}$ is given by

$$\kappa_\beta^{u,v} \left(= \kappa_\beta^{v,u}\right) = \sum_{\alpha=0}^{u+\beta} z_{\beta\alpha}^{(u+\beta,v+\beta)} \gamma_{v-u+2\alpha} \quad \text{for } 0 \le u \le v \le \ell-\beta \text{ and } 0 \le \beta \le \ell, \tag{3}$$

where

$$\gamma_j = \sum_{i=0}^{2\ell} \sum_{p=0}^{j} (-1)^p \binom{j}{p} \binom{2\ell-j}{i-j+p} \mu_i^{(2\ell)} \quad \text{for } 0 \le j \le 2\ell. \tag{4}$$

The relation between the indices $\mu_i^{(2\ell)}$ of a BA of strength 2ℓ and λ_x of an SA is given by

$$\mu_i^{(2\ell)} = \sum_{x=0}^{m} \binom{m-2\ell}{x-i} \lambda_x \quad \text{for } 0 \le i \le 2\ell. \tag{5}$$

Note that size N(=number of assemblies) of an SA$(m; \{\lambda_x\})$ is given by $N = \sum_{x=0}^{m} \binom{m}{x} \lambda_x$. Furthermore M_T is isomorphic to the symmetric matrices $\|\kappa_\beta^{u,v}\| (= K_\beta,$ say) of order $(\ell - \beta + 1)$, i.e., there exists an orthogonal matrix P of order $\nu_\ell(m)$ such that

$$P' M_T P = \mathrm{diag}\,[K_0; K_1, \cdots, K_1; K_2, \cdots, K_2; \cdots; K_\ell, \cdots, K_\ell], \tag{6}$$

where K_β $(0 \le \beta \le \ell)$ are with multiplicities ϕ_β. From (6), the following is immediately:

Lemma 2.1. *Let T be an* SA$(m; \{\lambda_x\})$. *Then the information matrix M_T is non-singular, i.e., T is of resolution $2\ell+1$, if and only if every K_β $(0 \le \beta \le \ell)$ is non-singular, i.e.,* $\mathrm{rank}\{K_\beta\} = \ell - \beta + 1$ *for all β.*

From (1), and (3) through (5), we have the following (see Hyodo and Yamamoto, 1988):

Lemma 2.2. *Let T be an* SA$(m; \{\lambda_x\})$. *Then we have*

$$\kappa_\beta^{u,v} = \sum_{x=\beta}^{m-\beta} \left[\left\{ 2^\beta \Big/ \sqrt{\binom{m-2\beta}{u}} \right\} \left\{ \sqrt{\lambda_x} \sum_{p=0}^{u} (-1)^p \binom{x-\beta}{u-p} \binom{m-\beta-x}{p} \right\} \sqrt{\binom{m-2\beta}{x-\beta}} \right]$$
$$\times \left[\left\{ 2^\beta \Big/ \sqrt{\binom{m-2\beta}{v}} \right\} \left\{ \sqrt{\lambda_x} \sum_{q=0}^{v} (-1)^q \binom{x-\beta}{v-q} \binom{m-\beta-x}{q} \right\} \sqrt{\binom{m-2\beta}{x-\beta}} \right]$$
$$\textit{for } 0 \le u \le v \le \ell-\beta,\ 0 \le \beta \le \ell, \textit{ and } 2\ell \le m.$$

Let $F_\beta\,(0 \le \beta \le \ell)$ be the $(\ell-\beta+1)\times(m-2\beta+1)$ matrices such that the column vector corresponding to the index λ_x $(x \in V_\beta)$ is given by $\boldsymbol{F}_\beta(x)$, where the $(u+1)$-th row of $\boldsymbol{F}_\beta(x)$ is given by

$$\sqrt{\lambda_x}\sum_{p=0}^{u}(-1)^p\binom{x-\beta}{u-p}\binom{m-\beta-x}{p} \text{ for } 0 \le u \le \ell-\beta \tag{7}$$

and $V_\beta = \{x \in N_0\,|\,\beta \le x \le m-\beta\}$. Here N_0 is a set of non-negative integers. The $(u+1)$-th row and the $(v+1)$-th column of K_β $(0 \le u, v \le \ell-\beta; 0 \le \beta \le \ell)$ correspond to the $(u+\beta)$-factor interaction and the $(v+\beta)$-factor one, respectively. Thus the $(u+1)$-th row of F_β corresponds to the $(u+\beta)$-factor interaction. Then from (7), we can easily obtain the following theorem (e.g., Hyodo *et al*., 2015):

Theorem 2.1. *Let T be an* SA$(m;\{\lambda_x\})$, *where $2\ell \le m$. Then the matrices $K_\beta\,(0 \le \beta \le \ell)$ can be expressed as $K_\beta = (D_\beta F_\beta \Lambda_\beta)(D_\beta F_\beta \Lambda_\beta)'$, where D_β and Λ_β are the diagonal matrices such that the $(u+1)$-th element $(0 \le u \le \ell-\beta)$ of D_β and the element of Λ_β corresponding to λ_x are given by $2^\beta/\sqrt{\binom{m-2\beta}{u}}$ and $\sqrt{\binom{m-2\beta}{x-\beta}}$, respectively.*

It follows from Theorem 2.1 that $\text{rank}\{K_\beta\} = \text{r-rank}\{F_\beta\}$, where r-rank$\{A\}$ denotes the row rank of a matrix A. In order to obtain the rank of a matrix A, we sometimes apply the "elementary row operations" on it. In this case, we positively use the notation "r-rank" instead of the rank. Let $SV_\beta = \{x \in V_\beta\,|\,\lambda_x \ne 0\}\,(0 \le \beta \le \ell)$, and further let NSV_β be the cardinal number of SV_β. Then the following is obtained (see Hyodo *et al*., 2015):

Theorem 2.2. *Let T be an* SA$(m;\{\lambda_x\})$, *where $2\ell \le m$. Then it holds that* r-rank$\{F_\beta(x_1, x_2, \cdots, x_{n_\beta})\} = \min(n_\beta, \ell-\beta+1)$ *for $\{x_1, x_2, \cdots, x_{n_\beta}\} \subset SV_\beta\,(0 \le \beta \le \ell)$, where $F_\beta(x_1, x_2, \cdots, x_{n_\beta}) = (\boldsymbol{F}_\beta(x_1), \boldsymbol{F}_\beta(x_2), \cdots, \boldsymbol{F}_\beta(x_{n_\beta}))$. Furthermore the first* $\min(n_\beta, \ell-\beta+1)$ *rows of $F_\beta(x_1, x_2, \cdots, x_{n_\beta})$ are linearly independent.*

The following is due to Ghosh and Kuwada (2001):

Lemma 2.3. *Let $K = \|K_{ij}\|$ and $L = \|L_{ij}\|$ $(i, j = 1, 2, 3)$ be a positive semi-definite matrix of order n with* $\text{rank}\{K\} = \text{rank}\left\{\begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}\right\} = n_1 + n_2 (\ge 1)$ *and some matrix of order n such that $L_{11} = I_{n_1}$ and $L_{1j} = L'_{j1} = 0_{n_1\times n_j}$ $(j = 2, 3)$, respectively, where K_{ij} and L_{ij} are both of size $n_i \times n_j$, and $n_1 + n_2 + n_3 = n$. Then a matrix equation $XK = L$ with parameter matrix X of order n has a solution if and only if*

(i) $n_3 = 0$, *where if $n_2 \ge 1$, then L_{22} is arbitrary, or*

(ii) $n_3 \ge 1$, *and moreover*

(1) *when $n_2 = 0$, $K_{33} = 0_{n_3\times n_3}$, and furthermore $L_{33} = 0_{n_3\times n_3}$, or*

(2) *when $n_2 \ge 1$, there exists a matrix W of size $n_3 \times n_2$ such that $K_{3j} = WK_{2j}$ $(j = 1, 2, 3)$, and furthermore $L_{i3} = L_{i2}W'$ $(i = 2, 3)$, where L_{i2} are arbitrary.*

Remark 2.1. In Lemma 2.3.(i) and (ii)(2), when $n_2 \ge 1$, without loss of generality, we can put $L_{22} = I_{n_2}$ and $L_{23} = L'_{32}$ $(= W')$ (if $n_3 \ge 1$), and hence $L_{33} = WW'$. Furthermore we have $W = K_{32}K_{22}^{-1}$.

Let T be an SA$(m;\{\lambda_x\})$, where $2\ell \le m$. Then a set of parametric functions $H\boldsymbol{\Theta}$ is estimable if and only if there exists a matrix X of order $\nu_\ell(m)$ such that $XM_T = H$, where H and X are given by

$$H = \sum_{\beta=0}^{\ell}\sum_{u=0}^{\ell-\beta}\sum_{v=0}^{\ell-\beta} h_\beta^{u,v} D_\beta^{\#(u+\beta,v+\beta)} \text{ and } X = \sum_{\beta=0}^{\ell}\sum_{u=0}^{\ell-\beta}\sum_{v=0}^{\ell-\beta} \chi_\beta^{u,v} D_\beta^{\#(u+\beta,v+\beta)},$$

where $2\ell \le m$. Thus there exist matrices X_β such that $X_\beta K_\beta = H_\beta$ for all $\beta\,(0 \le \beta \le \ell)$ if and only if T is of resolution $\text{R}^*(\{1\}|\Omega_\ell)$, where

$$H_0 = \begin{pmatrix} h_0^{0,0} & 0 & h_0^{0,2} & \cdots & h_0^{0,\ell} \\ 0 & 1 & 0 & \cdots & 0 \\ h_0^{2,0} & 0 & h_0^{2,2} & \cdots & h_0^{2,\ell} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_0^{\ell,0} & 0 & h_0^{\ell,2} & \cdots & h_0^{\ell,\ell} \end{pmatrix},\ H_1 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & h_1^{1,1} & \cdots & h_1^{1,\ell-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & h_1^{\ell-1,1} & \cdots & h_1^{\ell-1,\ell-1} \end{pmatrix}$$

$$\text{and } H_\gamma = \begin{pmatrix} h_\gamma^{0,0} & \cdots & h_\gamma^{0,\ell-\gamma} \\ \vdots & \ddots & \vdots \\ h_\gamma^{\ell-\gamma,0} & \cdots & h_\gamma^{\ell-\gamma,\ell-\gamma} \end{pmatrix} \text{ for } \gamma \ge 2.$$

Let $B = \mathrm{diag}[B_1, B_2, B_3]$, $C' = (C_1'\ C_2'\ C_3')$ and Δ be a diagonal and non-singular matrix of order $n(= n_1+n_2+n_3)$, a matrix of size $n \times p$ with r-rank$\{C\}$ = r-rank $\left\{\begin{pmatrix} C_1 \\ C_2 \end{pmatrix}\right\} = n_1 + n_2$, and a diagonal and non-singular matrix of order p, respectively, where B_i and C_i $(i = 1, 2, 3)$ are of order n_i and of size $n_i \times p$, respectively. Then we have the following:

Lemma 2.4. *Let $Z = \|Z_{ij}\|$ $(i, j = 1, 2, 3)$ be a matrix of order $n(= n_1 + n_2 + n_3)$, where Z_{ij} are of size $n_i \times n_j$, and let $K(= \|K_{ij}\|) = (BC\Delta)(BC\Delta)'$ and $K_{ij} = (B_iC_i\Delta)(B_jC_j\Delta)'$, and further let $L = \|L_{ij}\|$ be a matrix given by* Lemma 2.3. *Then a matrix equation $ZK = L$ has a solution if and only if*

(i) $n_3 = 0$, *where if $n_2 \geq 1$, then L_{22} is arbitrary, or*

(ii) $n_3 \geq 1$, *and moreover*

(1) *when $n_2 = 0$, it holds $C_3 = 0_{n_3\times p}$, and furthermore $L_{33} = 0_{n_3\times n_3}$, or*

(2) *when $n_2 \geq 1$, there exists a matrix $W_2^*(= C_3C_2'(C_2C_2')^{-1})$ of size $n_3 \times n_2$ such that $C_3 = W_2^*C_2$, and furthermore $L_{i3} = L_{i2}(B_3W_2^*B_2^{-1})'$ $(i = 2, 3)$, where L_{i2} are arbitrary.*

Proof. (i) When $n_3 = 0$, if $n_2 \geq 1$, then from Lemma 2.3.(i), we have the required result.

(ii)(1) When $n_3 \geq 1$ and $n_2 = 0$, it follows from Lemma 2.3.(ii)(1) that $K_{33} = (B_3C_3\Delta)(B_3C_3\Delta)' = B_3C_3\Delta\Delta'C_3'B_3' = 0_{n_3\times n_3}$, and hence $C_3\Delta\Delta'C_3' = 0_{n_3\times n_3}$, which implies $C_3\Delta = 0_{n_3\times p}$. Thus we get $C_3 = 0_{n_3\times p}$, and hence $L_{33} = 0_{n_3\times n_3}$.

(2) When $n_i \geq 1$ $(i = 2, 3)$, from Lemma 2.3.(ii)(2), there exists a matrix W_2 of size $n_3 \times n_2$ such that $K_{3j} = W_2K_{2j}$ $(j = 1, 2, 3)$. Since r-rank$\{C\}$ = r-rank $\left\{\begin{pmatrix} C_1 \\ C_2 \end{pmatrix}\right\} = n_1 + n_2$, there exists a matrix $(W_1^*\ \ W_2^*)$ of size $n_3 \times (n_1 + n_2)$ such that $C_3 = (W_1^*\ \ W_2^*)\begin{pmatrix} C_1 \\ C_2 \end{pmatrix}$, where W_k^* $(k = 1, 2)$ are of size $n_3 \times n_k$. Then $K_{3j} = (B_3C_3\Delta)(B_jC_j\Delta)' = \{B_3(W_1^*\ W_2^*)\begin{pmatrix} C_1 \\ C_2 \end{pmatrix}\Delta\}(B_jC_j\Delta)' = B_3\{W_1^*B_1^{-1}(B_1C_1\Delta)\}(B_jC_j\Delta)' + B_3\{W_2^*B_2^{-1}(B_2C_2\Delta)\}(B_jC_j\Delta)' = B_3W_1^*B_1^{-1}K_{1j} + B_3W_2^* \times B_2^{-1}K_{2j} = (B_3W_1^*B_1^{-1}\ \ B_3W_2^*B_2^{-1})\begin{pmatrix} K_{1j} \\ K_{2j} \end{pmatrix}$. Thus we have $(K_{31}\ \ K_{32}) = (B_3W_1^*B_1^{-1}\ \ B_3W_2^*B_2^{-1})\begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} = (O_{n_3\times n_1}\ \ W_2)\begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}$. Since $\begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}$ is non-singular, we get $B_3W_1^*B_1^{-1} = O_{n_3\times n_1}$, i.e., $W_1^* = O_{n_3\times n_1}$, and $W_2 = B_3W_2^*B_2^{-1}$.

From Lemma 2.3, the converse is obvious, and hence the required result is obtained.

Let $\overline{T}$ be an array obtained by interchanging all of symbols 0 and 1 of T, where T is an SA$(m; \{\lambda_x\})$. Then it can be easily shown that $\overline{T}$ is also an SA$(m; \{\overline{\lambda}_x\})$, where $\overline{\lambda}_x = \lambda_{m-x}$ for $0 \leq x \leq m$ (e.g., Shirakura and Kuwada, 1975). Note that $\overline{T}$ is called a complementary SA (CSA) of T. Furthermore if T is of resolution R$^*(\{1\}|\Omega_\ell)$, then $\overline{T}$ is also of resolution R$^*(\{1\}|\Omega_\ell)$.

If $N \geq \nu_\ell(m)$, then there always exists a 2^m-BFF design of resolution $2\ell + 1$ (see Hyodo *et al.*, 2015). Thus in the rest of this paper, we consider a 2^m-BFF design of resolution R$^*(\{1\}|\Omega_\ell)$ derived from an SA with $N < \nu_\ell(m)$ for $\ell = 2, 3$.

3. Resolution R$^*(\{1\}|\Omega_2)$ designs

We now consider case $\ell = 2$. Then it follows from (7) that $\boldsymbol{F}_\beta(x)$ $(0 \leq \beta \leq 2)$ are given by

$$\boldsymbol{F}_0(x) = \sqrt{\lambda_x}\begin{pmatrix} 1 \\ 2x - m \\ \{(2x-m)^2 - m\}/2 \end{pmatrix} \quad \text{for } x \in V_0, \tag{8a}$$

$$\boldsymbol{F}_1(x) = \sqrt{\lambda_x}\begin{pmatrix} 1 \\ 2x - m \end{pmatrix} \quad \text{for } x \in V_1 \tag{8b}$$

and

$$F_2(x) = \sqrt{\lambda_x}(1) \quad \text{for } x \in V_2. \tag{8c}$$

Let

$$f_\beta(x; m) = \{(2x - m)^2 + (2\beta - m)\}/2 \quad \text{for } x \in V_\beta\ (\beta = 0, 1). \tag{9}$$

Then we have the following:

Lemma 3.1. *If* $f_\beta(x_1;m) = f_\beta(x_2;m)$ *for* $\{x_1, x_2\} \subset V_\beta$ $(\beta = 0, 1)$, *then* $x_1 + x_2 - m = 0$.

Proof. If $f_\beta(x_1;m) = f_\beta(x_2;m)$ for $\{x_1, x_2\} \subset V_\beta$ $(\beta = 0, 1)$, i.e., $\{(2x_1 - m)^2 + (2\beta - m)\}/2 = \{(2x_2 - m)^2 + (2\beta - m)\}/2$, then we have $(2x_1 - m)^2 - (2x_2 - m)^2 = 4(x_1 - x_2)(x_1 + x_2 - m) = 0$. Thus we get the required result.

The following is the main results of this section:

Theorem 3.1. *Let* T *be an* SA$(m;\{\lambda_x\})$, *where* $SV_0 = \{x_1, x_2, \cdots, x_{\mathrm{NSV}_0}\}$, $N < \nu_2(m)$ *and* $m \geq 4$. *Then a necessary and sufficient condition for* T *to be a* 2^m-BFF *design of resolution* $\mathrm{R}^*(\{1\}|\Omega_2)$, *i.e., of resolution* IV, *is that non-zero indices of an* SA *satisfy the following*:

(i) *When* $\mathrm{NSV}_0 = 2$, $x_1 = 1$ *and* $x_2 = m - 1$,

(ii) *when* $\mathrm{NSV}_0 = 3$,

(1) $x_1 = 0$, $x_2 = 1$ *and* $x_3 = m - 1$, *or its* CSA, *or*

(2) $x_1 = 0$, $x_2 = 2$ *and* $x_3 = 4$, *where* $m = 4$

and

(iii) *when* $\mathrm{NSV}_0 = 4$, $x_1 = 0$, $x_2 = 1$, $x_3 = m - 1$ *and* $x_4 = m$.

Proof. See Appendix.

From Theorems 2.1 and 2.2, Lemmas 2.3 and 2.4, and Remark 2.1, we have the following:

Theorem 3.2. *If* T *is a* 2^m-BFF *design of resolution* $\mathrm{R}^*(\{1\}|\Omega_2)$, *i.e., of resolution* IV, *derived from an* SA$(m;\{\lambda_x\})$, *where* $m \geq 4$ *and* $N < \nu_2(m)$, *and moreover*

(i) *when* $\mathrm{NSV}_0 = 2$, $A_0^{\#(0,0)}\boldsymbol{\theta}_0 + \left[\left\{1/\sqrt{\binom{m}{2}}\right\} f_0(x_i;m)\right] A_0^{\#(0,2)}\boldsymbol{\theta}_2$ $(i = 1, 2)$ *and* $A_0^{\#(1,1)}\boldsymbol{\theta}_1$ *are estimable, where* $\{x_1, x_2\} = SV_0$ *and* $f_0(x;m)$ *is given by* (9), *and furthermore if* $f_0(x_i;m) = 0$ *for all* i, *then* $A_0^{\#(0,0)}\boldsymbol{\theta}_0$ *is estimable and* $A_0^{\#(2,2)}\boldsymbol{\theta}_2$ *is not estimable,*

(ii) *when* $\mathrm{NSV}_1 = 1$, $A_1^{\#(1,1)}\boldsymbol{\theta}_1$ *is estimable and* $A_1^{\#(2,2)}\boldsymbol{\theta}_2$ *is not estimable*,

(iii) *when* $\mathrm{NSV}_2 = 0$, $A_2^{\#(2,2)}\boldsymbol{\theta}_2$ *is not estimable*

and

(iv) *when* $\mathrm{NSV}_\beta \geq 3 - \beta$ $(0 \leq \beta \leq 2)$, $A_\beta^{\#(u,u)}\boldsymbol{\theta}_u$ $(\beta \leq u \leq 2)$ *are estimable*.

Note from (2) that if $A_\beta^{\#(u,u)}\boldsymbol{\theta}_u$ are estimable for all β $(0 \leq \beta \leq u \leq 2)$, then $\boldsymbol{\theta}_u$ is estimable. The results of Theorem 3.1, and estimable parametric functions and the resolution $\mathrm{R}(\omega|\Omega_2)$ for each design are summarized in Table 3.1.

Let $K_\beta^{(0)}$ $(0 \leq \beta < \ell)$ be the matrices of order $(\ell - \beta)$ obtained from K_β by cutting off its last row and column, and further let $\boldsymbol{k}_\beta^{1\prime} = (\kappa_\beta^{0,0}\ \kappa_\beta^{0,1}\ \cdots \kappa_\beta^{0,\ell-\beta})$ and $\boldsymbol{k}_\beta^{2\prime} = (\kappa_\beta^{\ell-\beta,0}\ \kappa_\beta^{\ell-\beta,1}\ \cdots \kappa_\beta^{\ell-\beta,\ell-\beta})$. Then we have the following due to Shirakura (1980):

Proposition 3.1. *A necessary and sufficient condition for a* $\mathrm{BA}(N, m, 2, 2\ell;\{\mu_i^{(2\ell)}\})$, T, *say, to be a* 2^m-BFF *design of resolution* 2ℓ *is that* T *satisfies the following condition*:

For r *integers* $0 \leq \beta_1 < \beta_2 < \cdots < \beta_r \leq \ell$ *with* $|K_{\beta_j}| = 0$ *and* $|K_\alpha| \neq 0$ $(\alpha \neq \beta_j\ (1 \leq j \leq r);\ 0 \leq \alpha \leq \ell)$,

(i) *when* $\beta_1 = 0$, *there exists a scalar* d *such that* $\boldsymbol{k}_0^2 = d\boldsymbol{k}_0^1$, $|K_0^{(0)}| \neq 0$, $\kappa_{\beta_j}^{\ell-\beta_j,\ell-\beta_j} = 0\ (1 \leq \beta_j \leq \ell)$ *and* $|K_{\beta_j}^{(0)}| \neq 0\ (1 \leq \beta_j \leq \ell - 1)$

and

(ii) *when* $\beta_1 \geq 1$, $\kappa_{\beta_j}^{\ell-\beta_j,\ell-\beta_j} = 0\ (1 \leq \beta_j \leq \ell)$ *and* $|K_{\beta_j}^{(0)}| \neq 0\ (1 \leq \beta_j \leq \ell - 1)$,

where $|A|$ *denotes the determinant of a matrix* A.

In a theoretical sense, Proposition 3.1 above is a very useful result. However it is not always practical. Because the elements $\kappa_\beta^{u,v}$ of K_β $(0 \leq u \leq v \leq \ell - \beta;\ 0 \leq \beta \leq \ell)$ are given by some linear combinations of the indices $\mu_i^{(2\ell)}$ of a BA (or λ_x of an SA) (see (3) and (4)). Hence it is not always easy to obtain $\mu_i^{(2\ell)}$ (or λ_x) such that these indices satisfy some conditions. As an example, we consider case $\ell = 2$:

Let T be an SA($m;\{\lambda_x\}$), where $m \geq 4$. Then from (3) through (5), we have

$$\kappa_0^{0,0} = \sum_{x=0}^{m} \binom{m}{x}\lambda_x (= N),$$

$$\kappa_0^{0,1}\left(=\kappa_0^{1,0}\right) = \left\{1\Big/\sqrt{\binom{m}{1}}\right\}\sum_{x=0}^{m}(2x-m)\binom{m}{x}\lambda_x,$$

$$\kappa_0^{0,2}\left(=\kappa_0^{2,0}\right) = \left[1\Big/\left\{2\sqrt{\binom{m}{2}}\right\}\right]\sum_{x=0}^{m}\{(2x-m)^2-m\}\binom{m}{x}\lambda_x,$$

$$\kappa_0^{1,1} = \left\{1\Big/\binom{m}{1}\right\}\sum_{x=0}^{m}(2x-m)^2\binom{m}{x}\lambda_x,$$

$$\kappa_0^{1,2}\left(=\kappa_0^{2,1}\right) = \left[1\Big/\left\{2\sqrt{\binom{m}{1}\binom{m}{2}}\right\}\right]\sum_{x=0}^{m}(2x-m)\{(2x-m)^2-m\}\binom{m}{x}\lambda_x,$$

$$\kappa_0^{2,2} = \left[1\Big/\left\{4\binom{m}{2}\right\}\right]\sum_{x=0}^{m}\{(2x-m)^2-m\}^2\binom{m}{x}\lambda_x,$$

$$\kappa_1^{0,0} = 2^2\sum_{x=1}^{m-1}\binom{m-2}{x-1}\lambda_x,$$

$$\kappa_1^{0,1}\left(=\kappa_1^{1,0}\right) = \left\{2^2\Big/\sqrt{\binom{m-2}{1}}\right\}\sum_{x=1}^{m-1}(2x-m)\binom{m-2}{x-1}\lambda_x,$$

$$\kappa_1^{1,1} = \left\{2^2\Big/\binom{m-2}{1}\right\}\sum_{x=1}^{m-1}(2x-m)^2\binom{m-2}{x-1}\lambda_x$$

and

$$\kappa_2^{0,0} = \sum_{x=2}^{m-2}\binom{m-4}{x-2}\lambda_x.$$

Thus it can be easily shown that $\kappa_1^{1,1} = 0$ if and only if there exists λ_{x^*} such that $(2x^*-m)^2 = 0$ for $x^* \in SV_1$, and $\kappa_2^{0,0} = 0$ if and only if $\lambda_{x^{**}} = 0$ for any $x^{**} \in V_2$. However it is not so easy to obtain the indices λ_x such that $\boldsymbol{k}_0^{2\prime}(= (\kappa_0^{2,0}\ \kappa_0^{2,1}\ \kappa_0^{2,2}))= d\boldsymbol{k}_0^{1\prime}(= d(\kappa_0^{0,0}\ \kappa_0^{0,1}\ \kappa_0^{0,2}))$ for $x \in SV_0$ and some d, that is to say, to obtain λ_x such that the system of the linear equations $\kappa_0^{2,u} = d\kappa_0^{0,u}\,(u = 0, 1, 2)$ satisfies for $x \in SV_0$ and some d. On the other hand, the elements of $F_\beta\,(0 \leq \beta \leq 2)$ are given by some polynomial of x of the indices λ_x of an SA (e.g., Hyodo *et al*., 2015) as seen from (8). In particular, the element of the first row of F_β is all one. Thus it follows from Theorem 2.2 and (8) that if r-rank$\{F_0\} = 2 < 3$, then there exist two indices λ_{x_i} $(i = 1, 2)$ such that $x_i \in SV_0$ and the elements of the last row of $F_0(x_1, x_2)$ are the same constants, i.e., $f_0(x_1; m) = f_0(x_2; m)$, where $f_0(x; m)$ is given by (9). Next if r-rank$\{F_1\} = 1 < 2$, then there exists an index λ_{x^*} such that $x^* \in SV_1$ and the last row of $\boldsymbol{F}_1(x^*)$ is 0, i.e., $2x^* - m = 0$, and if r-rank$\{F_2\} = 0 < 1$, then $F_2(x^{**}) = 0$ for any $x^{**} \in V_2$, i.e., $\lambda_{x^{**}} = 0$. Thus in order to obtain the indices λ_x of an SA such that they satisfy some conditions, the matrices F_β and Theorem 2.2 are very powerful.

Table 3.1. 2^m-BFF designs of resolution $R^*(\{1\}|\Omega_2)$ with $N < \nu_2(m)$ $(m \geq 4)$

Theorem 3.1		Existence conditions ($\lambda_{x_i} \neq 0 (i = 1, 2, \cdots, \mathrm{NSV}_0)$)	Estimable parametric functions	Resolution $R(\omega|\Omega_2)$
(i)		$x_1 = 1,\ x_2 = m - 1$	$A_0^{\#(0,0)}\boldsymbol{\theta}_0 + \left[(m-1)(m-4) \Big/ \left\{2\sqrt{\binom{m}{2}}\right\}\right] A_0^{\#(0,2)}\boldsymbol{\theta}_2$: estimable, $A_0^{\#(1,1)}\boldsymbol{\theta}_1$: estimable, $A_0^{\#(2,2)}\boldsymbol{\theta}_2$: not estimable (if $m = 4$), $A_1^{\#(u,u)}\boldsymbol{\theta}_u$: estimable $(u = 1, 2)$, $A_2^{\#(2,2)}\boldsymbol{\theta}_2$: not estimable	ω={0, 1} (if $m = 4$) {1} (if $m \geq 5$)
(ii)	(1)	$x_1 = 0,\ x_2 = 1,\ x_3 = m - 1$, or its CSA	$A_\gamma^{\#(u,u)}\boldsymbol{\theta}_u$: estimable $(\gamma \leq u \leq 2;\ \gamma = 0, 1)$, $A_2^{\#(2,2)}\boldsymbol{\theta}_2$: not estimable	ω={0, 1}
	(2)	$m = 4;\quad x_1 = 0,\ x_2 = 2,\ x_3 = 4$	$A_\gamma^{\#(u,u)}\boldsymbol{\theta}_u$: estimable $(\gamma \leq u \leq 2;\ \gamma = 0, 2)$, $A_1^{\#(1,1)}\boldsymbol{\theta}_1$: estimable, $A_1^{\#(2,2)}\boldsymbol{\theta}_2$: not estimable	
(iii)		$x_1 = 0,\ x_2 = 1,\ x_3 = m - 1,\ x_4 = m$	$A_\gamma^{\#(u,u)}\boldsymbol{\theta}_u$: estimable $(\gamma \leq u \leq 2;\ \gamma = 0, 1)$, $A_2^{\#(2,2)}\boldsymbol{\theta}_2$: not estimable	

4. Resolution $R^*(\{1\}|\Omega_3)$ designs

In this section, we consider case $\ell = 3$. By use of the properties of the TMDPB association algebra and the matrix equations, a necessary and sufficient condition for an SA to be a 2^m-BFF design of resolution $R^*(\{1\}|\Omega_3)$ was already given by Kuwada *et al*. (2003). However their results are very complex. On the other hand, the elements of F_β $(0 \le \beta \le 3)$ considered here are given by some polynomial of x of the indexes λ_x of an SA as in (10) below, and they are very simple. Thus using these matrices F_β and Theorem 2.2, we shall rewrite the existence conditions for a 2^m-BFF design of resolution $R^*(\{1\}|\Omega_3)$ with $N < \nu_3(m)$.

From (7), $\boldsymbol{F}_\beta(x)$ $(0 \le \beta \le 3)$ are given by

$$\boldsymbol{F}_0(x) = \sqrt{\lambda_x}\begin{pmatrix} 1 \\ 2x-m \\ \{(2x-m)^2-m\}/2 \\ (2x-m)\{(2x-m)^2-(3m-2)\}/6 \end{pmatrix} \quad \text{for } x \in V_0, \tag{10a}$$

$$\boldsymbol{F}_1(x) = \sqrt{\lambda_x}\begin{pmatrix} 1 \\ 2x-m \\ \{(2x-m)^2-(m-2)\}/2 \end{pmatrix} \quad \text{for } x \in V_1, \tag{10b}$$

$$\boldsymbol{F}_2(x) = \sqrt{\lambda_x}\begin{pmatrix} 1 \\ 2x-m \end{pmatrix} \quad \text{for } x \in V_2 \tag{10c}$$

and

$$F_3(x) = \sqrt{\lambda_x}\,(1) \quad \text{for } x \in V_3. \tag{10d}$$

Let

$$g(x;m) = (2x-m)\{(2x-m)^2-(3m-2)\}/6 \quad \text{for } x \in V_0 \text{ and } m \ge 6. \tag{11}$$

Then we have the following:

Lemma 4.1. (I)(i) *If* $d_{02} = f_0(x_i;m)$ *and* $d_{03} = g(x_i;m)$ $(i = 1,2)$ *for* $\{x_1,x_2\} \subset V_0$ *and* $m \ge 6$, *where* d_{0k} $(k = 2,3)$ *are constants, and* $f_0(x;m)$ *and* $g(x;m)$ *are given by* (9) *and* (11), *respectively, then*

(1) *when* $m = 3t^2+2t+1$ $(t \ge 1)$, *we get* $x_p = t(3t-1)/2(\ge 1)$ *and* $x_q = (t+1)(3t+2)/2(\ge 5)$ *for* $\{p,q\} = \{1,2\}$ *and*

(2) *when* $m = 3t^2+4t+2$ $(t \ge 1)$, *we get* $x_p = t(3t+1)/2(\ge 2)$ *and* $x_q = (t+1)(3t+4)/2(\ge 7)$ *for* $\{p,q\} = \{1,2\}$.

Here in (1) *and* (2) *just above, we have* $d_{02} = m-1$ *and* $d_{03} = 0$.

(ii) *If* $d_0 + f_0(x_i;m)d_2 = g(x_i;m)$ $(i = 1,2,3)$ *for* $\{x_1,x_2,x_3\} \subset V_0$ *and* $m \ge 6$, *where* d_k $(k = 0,2)$ *are constants, then* $x_p+x_q-m \ne 0$ *for some* $\{p,q\} \subset \{1,2,3\}$ *and* $(2x_1-m)(2x_2-m)+(2x_2-m)(2x_3-m)+(2x_3-m)(2x_1-m)+(3m-2) = 0$. *Here*

$$d_0 = -\Big[(2x_p-m)^2(2x_q-m)^2 - m\{4(x_p-x_q)^2-(3m-2)\} - 2(2x_p-m)(2x_q-m)\Big]/\{12(x_p+x_q-m)\}$$

and

$$d_2 = \{(2x_p-m)^2+(2x_p-m)(2x_q-m)+(2x_q-m)^2-(3m-2)\}/\{6(x_p+x_q-m)\}.$$

(II)(i) *There does not exist a integer* $x_1^* \in V_1$ *such that* $2x_1^*-m = 0$ *and* $f_1(x_1^*;m) = 0$ *for* $m \ge 6$, *where* $f_1(x^*;m)$ *is given by* (9).

(ii) *If* $(2x_j^*-m)d^* = f_1(x_j^*;m)$ $(j = 1,2)$ *for* $\{x_1^*,x_2^*\} \subset V_1$ *and* $m \ge 6$, *where* d^* *is a constant, then* $(2x_1^*-m)(2x_2^*-m)+(m-2) = 0$. *Here* $d^* = x_1^*+x_2^*-m$.

Proof. (I)(i) If $d_{02} = f_0(x_i;m)$ $(i = 1,2)$ for $\{x_1,x_2\} \subset V_0$ and $m \ge 6$, i.e., $f_0(x_1;m) = f_0(x_2;m)$, then from Lemma 3.1, we have $x_1+x_2-m = 0$. In addition, if $d_{03} = g(x_i;m)$ $(i = 1,2)$, i.e., $g(x_1;m) = g(x_2;m)$, then $(2x_p-m)\{(2x_p-m)^2-(3m-2)\} = 0$ for some $p \in \{1,2\}$. If $2x_p-m = 0$, then $x_p = m/2 = x_q$ for $q \in \{1,2\} \setminus \{p\}$, and hence $2x_p-m \ne 0$. Thus it must be $(2x_p-m)^2-(3m-2) = 0$. Then it has solutions $x_p = (m \pm \sqrt{3m-2})/2$, which must be integers.

Thus we put $3m-2 = s^2$ ($s \geq 4$), and hence $m = (s^2+2)/3$. If $s = 3t$ ($t \geq 2$), then $m = 3t^2+2/3$, if $s = 3t+1$ ($t \geq 1$), then $m = 3t^2+2t+1$, and if $s = 3t+2$ ($t \geq 1$), then $m = 3t^2+4t+2$, and hence $s \neq 3t$. Since $(2x-m)^2-(3m-2) = 0$ for $x \in \{x_1, x_2\}$, we get $d_{02} = m-1$ and $d_{03} = 0$. Therefore (i) is proved.

(ii) It follows from Theorem 2.2 that the first three rows of $F_0(x_1, x_2, x_3)$ are linearly independent, and hence the first and the third rows of $F_0(x_1, x_2, x_3)$ are also linearly independent. Thus there exists $\{x_p, x_q\} \subset \{x_1, x_2, x_3\}$ such that $f_0(x_p; m) \neq f_0(x_q; m)$, i.e., $x_p + x_q - m \neq 0$. If $d_0 + f_0(x_i; m)d_2 = g(x_i; m)$ ($i = 1, 2, 3$), i.e., $d_0 + [\{(2x_i-m)^2-m\}/2]d_2 = (2x_i - m)\{(2x_i - m)^2 - (3m-2)\}/6$, where d_k ($k = 0, 2$) are constants, then we get $d_0 = -[(2x_p - m)^2(2x_q - m)^2 - m\{4(x_p - x_q)^2 - (3m-2)\} - 2(2x_p - m)(2x_q - m)]/\{12(x_p + x_q - m)\}$ and $d_2 = \{(2x_p - m)^2 + (2x_p - m)(2x_q - m) + (2x_q - m)^2 - (3m-2)\}/\{6(x_p + x_q - m)\}$ for some $\{p, q\} \subset \{1, 2, 3\}$, where $x_p + x_q - m \neq 0$. Substituting d_0 and d_2 into $d_0 + f_0(x_r; m)d_2 = g(x_r; m)$ for $r \in \{1, 2, 3\} \setminus \{p, q\}$, we get $(2x_p - m)(2x_q - m) + (2x_q - m)(2x_r - m) + (2x_r - m)(2x_p - m) + (3m-2) = (2x_1 - m)(2x_2 - m) + (2x_2 - m)(2x_3 - m) + (2x_3 - m)(2x_1 - m) + (3m-2) = 0$, and hence (ii) is established.

(II)(i) If $2x_1^* - m = 0$ and $f_1(x_1^*; m) = 0$ for $x_1^* \in V_1$, i.e., $\{(2x_1^* - m)^2 - (m-2)\}/2 = -(m-2)/2 = 0$, then $m = 2 < 6$. Thus the required result is obtained.

(ii) If $(2x_j^* - m)d^* = f_1(x_j^*; m)$ ($j = 1, 2$) for $\{x_1^*, x_2^*\} \subset V_1$ and $m \geq 6$, i.e., $(2x_j^* - m)d^* = \{(2x_j^* - m)^2 - (m-2)\}/2$, then $2(x_1^* - x_2^*)d^* = \{(2x_1^* - m)^2 - (2x_2^* - m)^2\}/2 = 2(x_1^* - x_2^*)(x_1^* + x_2^* - m)$. Thus we get $d^* = x_1^* + x_2^* - m$, and hence $(2x_1^* - m)(2x_2^* - m) + (m-2) = 0$, which is the required result.

The following is the main theorem of this section:

Theorem 4.1. *Let T be an* SA$(m; \{\lambda_x\})$, *where* $SV_0 = \{x_1, x_2, \cdots, x_{\mathrm{NSV}_0}\}$, $N < \nu_3(m)$ *and* $m \geq 6$. *Then a necessary and sufficient condition for T to be a* 2^m-BFF *design of resolution* $\mathrm{R}^*(\{1\}|\Omega_3)$ *is that non-zero indices of an* SA *satisfy the following*:

(i) *When* $\mathrm{NSV}_0 = 3$,

(1) $x_1 = 1$, $x_2 = 2$ *and* $x_3 = 5$, *where* $m = 6$, *or its* CSA,
(2) $x_1 = 1$, $x_2 = 2$ *and* $x_3 = 7$, *where* $m = 9$, *or its* CSA, *or*
(3) $x_1 = 1$, $x_2 = 3$ *and* $x_3 = 5$, *where* $m = 6$,

(ii) *when* $\mathrm{NSV}_0 = 4$,

(1) $x_1 = 0$, $x_2 = 1$, $x_3 = 2$ *and* $x_4 = m-1$, *or its* CSA,
(2) $x_1 = 0$, $x_2 = 1$, $x_3 = m-2$ *and* $x_4 = m-1$, *or its* CSA,
(3) $x_1 = 0$, $x_2 = 2$, $x_3 = 4$ *and* $x_4 = 6$, *where* $m = 6$,
(4) $x_1 = 0$, $x_2 = 1$, $x_3 = 2$ *and* $x_4 = m-2$, *or its* CSA,
(5) $x_1 = 0$, $x_2 = 2$, $x_3 = m-2$ *and* $x_4 = m-1$, *or its* CSA,
(6) $x_1 = 1$, $x_2 = 2$, $x_3 = m-2$ *and* $x_4 = m-1$, *where* $m \geq 7$,
(7) $x_1 = 0$, $x_2 = 1$, $x_3 = 4$ *and* $x_4 = 7$, *where* $m = 7$, *or its* CSA,
(8) $x_1 = 0$, $x_2 = 1$, $x_3 = 3$ *and* $x_4 = m-1$, *or its* CSA,
(9) $x_1 = 0$, $x_2 = 1$, $x_3 = m-3$ *and* $x_4 = m-1$, *where* $m \geq 7$, *or its* CSA, *or*
(10) $x_1 = 0$, $x_2 = 1$, $x_3 = 4$ *and* $x_4 = 7$, *where* $m = 8$, *or its* CSA,

(iii) *when* $\mathrm{NSV}_0 = 5$,

(1) $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = m-1$ *and* $x_5 = m$, *or its* CSA,
(2) $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = m-2$ *and* $x_5 = m$, *or its* CSA,
(3) $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = m-2$ *and* $x_5 = m-1$, *where* $m \geq 7$, *or its* CSA,
(4) $x_1 = 0$, $x_2 = 1$, $x_3 = 3$, $x_4 = m-1$ *and* $x_5 = m$, *or its* CSA(*if* $m \geq 7$), *or*
(5) $x_1 = 0$, $x_2 = 1$, $x_3 = 4$, $x_4 = 7$ *and* $x_5 = 8$, *where* $m = 8$

and

(iv) *when* $\mathrm{NSV}_0 = 6$, $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = m-2$, $x_5 = m-1$ *and* $x_6 = m$, *where* $m \geq 7$.

Proof. Proof is available in the Appendix.

It follows from Theorems 2.1 and 2.2, Lemmas 2.3, 2.4 and 4.1, and Remark 2.1 that we obtain the following:

Theorem 4.2. *If T is a 2^m-*BFF *design of resolution* $\mathrm{R}^*(\{1\}|\Omega_3)$ *derived from an* $\mathrm{SA}(m;\{\lambda_x\})$, *where* $m \geq 6$ *and* $N < \nu_3(m)$, *and furthermore*

(i) *when* $\mathrm{NSV}_0 = 3$, $A_0^{\#(0,0)}\boldsymbol{\theta}_0 + \left[\left\{1\Big/\sqrt{\binom{m}{3}}\right\}d_0\right]A_0^{\#(0,3)}\boldsymbol{\theta}_3$, $A_0^{\#(1,1)}\boldsymbol{\theta}_1$ *and* $A_0^{\#(2,2)}\boldsymbol{\theta}_2 + \left[\left\{\sqrt{\binom{m}{2}}\Big/\sqrt{\binom{m}{3}}\right\}d_2\right]A_0^{\#(2,3)}\boldsymbol{\theta}_3$ *are estimable*, *where* d_k $(k = 0, 2)$ *are given in* Lemma 4.1.(I)(ii). *In particular*, *if* $d_0 = 0$, *then* $A_0^{\#(0,0)}\boldsymbol{\theta}_0 = \boldsymbol{\theta}_0$ *is estimable and* $A_0^{\#(3,3)}\boldsymbol{\theta}_3$ *is not estimable*, *and also if* $d_2 = 0$, *then* $A_0^{\#(2,2)}\boldsymbol{\theta}_2$ *is estimable and* $A_0^{\#(3,3)}\boldsymbol{\theta}_3$ *is not estimable.*

(ii) *When* $\mathrm{NSV}_1 = 2$, $A_1^{\#(1,1)}\boldsymbol{\theta}_1$ *and* $A_1^{\#(2,2)}\boldsymbol{\theta}_2 + \left[\left\{\sqrt{\binom{m-2}{1}}\Big/\sqrt{\binom{m-2}{2}}\right\}d^*\right]A_1^{\#(2,3)}\boldsymbol{\theta}_3$ *are estimable*, *where* d^* *is given in* Lemma 4.1.(II)(ii). *Particularly if* $d^* = 0$, *then* $A_1^{\#(2,2)}\boldsymbol{\theta}_2$ *is estimable and* $A_1^{\#(3,3)}\boldsymbol{\theta}_3$ *is not estimable.*

(iii) *When* $\mathrm{NSV}_2 = 1$, $A_2^{\#(2,2)}\boldsymbol{\theta}_2 + \left[\left\{1\Big/\sqrt{\binom{m-4}{1}}\right\}d^{**}\right]A_2^{\#(2,3)}\boldsymbol{\theta}_3$ *is estimable, where* $d^{**} = 2x^{**} - m$ *for* $x^{**} \in SV_2$. *In particular*, *if* $d^{**} = 0$, *then* $A_2^{\#(2,2)}\boldsymbol{\theta}_2$ *is estimable and* $A_2^{\#(3,3)}\boldsymbol{\theta}_3$ *is not estimable.*

(iv) *When* $\mathrm{NSV}_3 = 0$, $A_3^{\#(3,3)}\boldsymbol{\theta}_3$ *is not estimable.*

(v) *When* $\mathrm{NSV}_\beta \geq 4-\beta$ $(0 \leq \beta \leq 3)$, $A_\beta^{\#(u,u)}\boldsymbol{\theta}_u$ $(\beta \leq u \leq 3)$ *are estimable.*

Analogously to Section 3, if $A_\beta^{\#(u,u)}\boldsymbol{\theta}_u$ are estimable for all β $(0 \leq \beta \leq u \leq 3)$, then $\boldsymbol{\theta}_u$ is estimable. In Table 4.1, the results of Theorem 4.1, and estimable parametric functions and the resolution $\mathrm{R}(\omega|\Omega_3)$ for each design are summarized.

5. Discussion

The class of BFF designs is a subset of FF designs. Thus there may exist a better FF design than a BFF design with respect to some criterion (e.g., Kuwada, 1982). However BFF designs possess the same advantage over unbalanced designs as a BIB design does over unbalanced or partially balanced designs. In this paper, we restrict our attention to the class of 2^m-BFF designs derived from SAs. Under these restrictions, we have given a necessary and sufficient condition for an SA to be a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_\ell)$ for $\ell = 2, 3$, where $N < \nu_\ell(m)$. As mentioned earlier, if T is an $\mathrm{SA}(m;\{\lambda_x\})$, then it is a $\mathrm{BA}(N, m, 2, t;\{\mu_i^{(t)}\})$ for any t $(1 \leq t \leq m)$, where the relation between $\mu_i^{(t)}$ and λ_x is given by (5). When $m = t+1$, if there exists a BA of strength t, then it is an SA. However when $m = t+2$, there exists a BA of strength t such that it is not always an SA (e.g., Kuriki and Yamamoto, 1984, and Shirakura, 1977). For example,

$$T' = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \end{pmatrix}$$

is a $\mathrm{BA}(N = 7, m = 4, 2, t = 2;\{\mu_0^{(2)} = 1,\ \mu_1^{(2)} = \mu_2^{(2)} = 2\})$, but it is not an SA.

Let T be an $\mathrm{SA}(m;\{\lambda_1 = \lambda_{m-1} = 1,\ \lambda_x = 0\ (x \neq 1, m-1)\})$ with $N = 2m$, where $m \geq 6$. Then under the assumption that the three-factor and higher-order interactions are negligible, $A_0^{\#(0,0)}\boldsymbol{\theta}_0 + \left[(m-1)(m-4)/\left\{2\sqrt{\binom{m}{2}}\right\}\right]A_0^{\#(0,2)}\boldsymbol{\theta}_2$, $A_0^{\#(1,1)}\boldsymbol{\theta}_1$ and $A_1^{\#(u,u)}\boldsymbol{\theta}_u$ $(u = 1, 2)$ are estimable, and $A_2^{\#(2,2)}\boldsymbol{\theta}_2$ is not estimable (see Table 3.1.(i)). Thus from (2), $A_0^{\#(1,1)}\boldsymbol{\theta}_1 + A_1^{\#(1,1)}\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1$ is estimable, and the general mean is confounded with some of the two-factor interaction. On the other hand, under the assumption that the four-factor and higher-order interactions are negligible,

Table 4.1. 2^m-BFF designs of resolution $R^*(\{1\}|\Omega_3)$ with $N < \nu_3(m)$ $(m \geq 6)$

Theorem 4.1		Existence conditions ($\lambda_{x_i} \neq 0\,(i=1,2,\cdots,\mathrm{NSV}_0)$)	Estimable parametric functions	Resolution $R(\omega|\Omega_3)$
(i)	(1)	$m=6$; $x_1=1$, $x_2=2$, $x_3=5$, or its CSA	$A_0^{\#(0,0)}\boldsymbol{\theta}_0+\left(\sqrt{5}/3\right)A_0^{\#(0,3)}\boldsymbol{\theta}_3$: estimable, $A_0^{\#(1,1)}\boldsymbol{\theta}_1$: estimable, $A_0^{\#(2,2)}\boldsymbol{\theta}_2-\left(1/\sqrt{3}\right)A_0^{\#(2,3)}\boldsymbol{\theta}_3$: estimable, $A_1^{\#(u,u)}\boldsymbol{\theta}_u$: estimable $(u=1,2,3)$, $A_2^{\#(2,2)}\boldsymbol{\theta}_2-\sqrt{2}A_2^{\#(2,3)}\boldsymbol{\theta}_3$: estimable, $A_3^{\#(3,3)}\boldsymbol{\theta}_3$: not estimable	$\omega=\{1\}$
	(2)	$m=9$; $x_1=1$, $x_2=2$, $x_3=7$, or its CSA	$A_0^{\#(0,0)}\boldsymbol{\theta}_0+\{28/(3\sqrt{21}\}A_0^{\#(0,3)}\boldsymbol{\theta}_3$: estimable, $A_0^{\#(1,1)}\boldsymbol{\theta}_1$: estimable, $A_0^{\#(2,2)}\boldsymbol{\theta}_2-(\sqrt{7/3})A_0^{\#(2,3)}\boldsymbol{\theta}_3$: estimable, $A_\gamma^{\#(u,u)}\boldsymbol{\theta}_u$: estimable $(\gamma\leq u\leq 3;\ \gamma=1,2)$, $A_3^{\#(3,3)}\boldsymbol{\theta}_3$: not estimable	
	(3)	$m=6$; $x_1=1$, $x_2=3$, $x_3=5$	$A_0^{\#(u,u)}\boldsymbol{\theta}_u$: estimable $(u=0,1,2)$, $A_0^{\#(3,3)}\boldsymbol{\theta}_3$: not estimable, $A_\gamma^{\#(u,u)}\boldsymbol{\theta}_u$: estimable $(\gamma\leq u\leq 3;\ \gamma=1,3)$, $A_2^{\#(2,2)}\boldsymbol{\theta}_2$: estimable, $A_2^{\#(3,3)}\boldsymbol{\theta}_3$: not estimable	$\omega=\{0,1,2\}$
(ii)	(1)	$x_1=0$, $x_2=1$, $x_3=2$, $x_4=m-1$, or its CSA	$A_\gamma^{\#(u,u)}\boldsymbol{\theta}_u$: estimable $(\gamma\leq u\leq 3;\ \gamma=0,1)$, $A_2^{\#(2,2)}\boldsymbol{\theta}_2+\{(2x_3-m)/\sqrt{m-4}\}A_2^{\#(2,3)}\boldsymbol{\theta}_3$: estimable, $A_\gamma^{\#(3,3)}\boldsymbol{\theta}_3$: not estimable	$\omega=\{0,1\}$
	(2)	$x_1=0$, $x_2=1$, $x_3=m-2$, $x_4=m-1$, or its CSA		
	(3)	$m=6$; $x_1=0$, $x_2=2$, $x_3=4$, $x_4=6$	$A_\gamma^{\#(u,u)}\boldsymbol{\theta}_u$: estimable $(\gamma\leq u\leq 3;\ \gamma=0,2)$, $A_1^{\#(u,u)}\boldsymbol{\theta}_u$: estimable $(u=1,2)$, $A_\gamma^{\#(3,3)}\boldsymbol{\theta}_3$: not estimable $(\gamma=1,3)$	$\omega=\{0,1,2\}$

Table 4.1. (Continued)

Theorem 4.1		Existence conditions ($\lambda_{x_i} \neq 0\,(i=1,2,\cdots,\mathrm{NSV}_0)$)	Estimable parametric functions	Resolution $\mathrm{R}(\omega\|\Omega_3)$
	(4)	$x_1=0,\ x_2=1,\ x_3=2,\ x_4=m-2$, or its CSA	$A_\gamma^{\#(u,u)}\boldsymbol{\theta}_u$: estimable ($\gamma \le u \le 3;\ \gamma=0,1,2$), $A_3^{\#(3,3)}\boldsymbol{\theta}_3$: not estimable	$\omega=\{0,1,2\}$
	(5)	$x_1=0,\ x_2=2,\ x_3=m-2,\ x_4=m-1$, or its CSA		
	(6)	$m\ge 7;\quad x_1=1,\ x_2=2,\ x_3=m-2,\ x_4=m-1$		
	(7)	$m=7;\quad x_1=0,\ x_2=1,\ x_3=4,\ x_4=7$, or its CSA	$A_\gamma^{\#(u,u)}\boldsymbol{\theta}_u$: estimable ($\gamma \le u \le 3;\ \gamma=0,3$), $A_1^{\#(1,1)}\boldsymbol{\theta}_1$: estimable, $A_1^{\#(2,2)}\boldsymbol{\theta}_2-\sqrt{2}A_1^{\#(2,3)}\boldsymbol{\theta}_3$: estimable, $A_2^{\#(2,2)}\boldsymbol{\theta}_2+\sqrt{1/3}A_2^{\#(2,3)}\boldsymbol{\theta}_3$: estimable	$\omega=\{0,1\}$
	(8)	$x_1=0,\ x_2=1,\ x_3=3,\ x_4=m-1$, or its CSA	$A_\gamma^{\#(u,u)}\boldsymbol{\theta}_u$: estimable ($\gamma \le u \le 3;\ \gamma=0,1,3$), $A_2^{\#(2,2)}\boldsymbol{\theta}_2+\left\{(2x_3-m)/\sqrt{m-4}\right\}A_2^{\#(2,3)}\boldsymbol{\theta}_3$: estimable, $A_2^{\#(3,3)}\boldsymbol{\theta}_3$: not estimable (if $2x_3-m=0$)	$\omega=\{0,1,2\}$ (if $m=6$) $\{0,1\}$ (if $m\ge 7$)
	(9)	$m\ge 7;\quad x_1=0,\ x_2=1,\ x_3=m-3,\ x_4=m-1$, or its CSA		$\omega=\{0,1\}$
	(10)	$m=8;\quad x_1=0,\ x_2=1,\ x_3=4,\ x_4=7$, or its CSA		$\omega=\{0,1,2\}$
(iii)	(1)	$x_1=0,\ x_2=1,\ x_3=2,\ x_4=m-1,\ x_5=m$, or its CSA	$A_\gamma^{\#(u,u)}\boldsymbol{\theta}_u$: estimable ($\gamma \le u \le 3;\ \gamma=0,1$), $A_2^{\#(2,2)}\boldsymbol{\theta}_2+\left\{(2x_3-m)/\sqrt{m-4}\right\}A_2^{\#(2,3)}\boldsymbol{\theta}_3$: estimable, $A_3^{\#(3,3)}\boldsymbol{\theta}_3$: not estimable	$\omega=\{0,1\}$
	(2)	$x_1=0,\ x_2=1,\ x_3=2,\ x_4=m-2,\ x_5=m$, or its CSA	$A_\gamma^{\#(u,u)}\boldsymbol{\theta}_u$: estimable ($\gamma \le u \le 3;\ \gamma=0,1,2$), $A_3^{\#(3,3)}\boldsymbol{\theta}_3$: not estimable	$\omega=\{0,1,2\}$
	(3)	$m\ge 7;\quad x_1=0,\ x_2=1,\ x_3=2,\ x_4=m-2,\ x_5=m-1$, or its CSA		
	(4)	$x_1=0,\ x_2=1,\ x_3=3,\ x_4=m-1,\ x_5=m$, or its CSA (if $m\ge 7$)	$A_\gamma^{\#(u,u)}\boldsymbol{\theta}_u$: estimable ($\gamma \le u \le 3;\ \gamma=0,1,3$), $A_2^{\#(2,2)}\boldsymbol{\theta}_2+\left\{(2x_3-m)/\sqrt{m-4}\right\}A_2^{\#(2,3)}\boldsymbol{\theta}_3$: estimable, $A_2^{\#(3,3)}\boldsymbol{\theta}_3$: not estimable (if $2x_3-m=0$)	$\omega=\{0,1,2\}$ (if $m=6$) $\{0,1\}$ (if $m\ge 7$)
	(5)	$m=8;\quad x_1=0,\ x_2=1,\ x_3=4,\ x_4=7,\ x_5=8$		$\omega=\{0,1,2\}$
(iv)		$m\ge 7;\quad x_1=0,\ x_2=1,\ x_3=2,\ x_4=m-2,\ x_5=m-1,\ x_6=m$	$A_\gamma^{\#(u,u)}\boldsymbol{\theta}_u$: estimable ($\gamma \le u \le 3;\ \gamma=0,1,2$), $A_3^{\#(3,3)}\boldsymbol{\theta}_3$: not estimable	

$$
\begin{aligned}
&\text{r-rank}\{F_0(1, m-1)\} \\
&= \text{r-rank}\left\{\begin{pmatrix} 1 & 1 \\ -(m-2) & m-2 \\ (m-1)(m-4)/2 & (m-1)(m-4)/2 \\ -(m-1)(m-2)(m-6)/6 & (m-1)(m-2)(m-6)/6 \end{pmatrix}\right\} \\
&= 2 < 4, \\
&\text{r-rank}\{F_1(1, m-1)\} \\
&= \text{r-rank}\left\{\begin{pmatrix} 1 & 1 \\ -(m-2) & m-2 \\ (m-2)(m-3)/2 & (m-2)(m-3)/2 \end{pmatrix}\right\} \\
&= 2 < 3
\end{aligned}
$$

and

$$\text{r-rank}\{F_2\} = \text{r-rank}\{F_3\} = 0.$$

Thus the third row of $F_0(1, m-1)$ equals $(m-1)(m-4)/2$ times the first and its last row equals $(m-1)(m-6)/6$ times the second, and the last row of $F_1(1, m-1)$ equals $(m-2)(m-3)/2$ times the first. Hence from Lemma 2.4, $A_0^{\#(0,0)}\boldsymbol{\theta}_0 + \{(m-4)\sqrt{(m-1)/(2m)}\}A_0^{\#(0,2)}\boldsymbol{\theta}_2$, $A_0^{\#(1,1)}\boldsymbol{\theta}_1 + [(m-6)\sqrt{(m-1)/\{6(m-2)\}}]A_0^{\#(1,3)}\boldsymbol{\theta}_3$, $A_1^{\#(1,1)}\boldsymbol{\theta}_1 + \left\{\sqrt{\binom{m-2}{2}}\right\}A_1^{\#(1,3)}\boldsymbol{\theta}_3$ and $A_1^{\#(2,2)}\boldsymbol{\theta}_2$ are estimable, and $A_\gamma^{\#(u,u)}\boldsymbol{\theta}_u$ $(2 \le \gamma \le u \le 3)$ are not estimable. This implies that the main effect is confounded with some of the three-factor interaction. Therefore if the three-factor and higher-order interactions are negligible, then T is of resolution $\text{R}^*(\{1\}|\Omega_2)$, and hence it is also of resolution $\text{R}(\{1\}|\Omega_2)$. However if the three-factor interaction is not negligible, then the main effect is not estimable, and hence it is not of resolution $\text{R}^*(\{1\}|\Omega_3)$.

Appendix

In this Appendix, we provide the proofs of Theorems 3.1 and 4.1.

Proof of Theorem 3.1. We shall prove the claim by listing out all possible cases. In these cases, since $SV_0 \supset SV_1 \supset SV_2$, we have $\text{NSV}_0 \ge \text{NSV}_1 \ge \text{NSV}_2$, and $\text{NSV}_\gamma - \text{NSV}_{\gamma+1} \le 2$ for $\gamma = 0, 1$. Furthermore when $\text{NSV}_0 \ge 5$, we have $N \ge 1 + 1 + m + m + \binom{m}{2} > \nu_2(m)$ for $m \ge 4$, and hence $\text{NSV}_0 \le 4$. Then the proof starts with case $\text{NSV}_2 = 0$.

[A] When $\text{NSV}_2 = 0$, i.e., $\lambda_{x^{**}} = 0$ for any $x^{**} \in V_2$, from Theorem 2.2, it must be that $1 \le \text{NSV}_1 \le 2$ and $\text{NSV}_0 \ge 2$. In addition,

[a] when $\text{NSV}_1 = 1$, i.e., $x_1^* = 1$ or $m-1$, we have $\text{r-rank}\{F_1(x_1^*)\} = 1 < 2$. Thus from (8b), the last row of $F_1(x_1^*)$ must be 0, i.e., $2x_1^* - m = 0$, where $m = 2s \ge 4$. However $2x_1^* - m = -(m-2) < 0$ and $m - 2 > 0$ for $m \ge 4$ according as $x_1^* = 1$ and $m-1$, respectively. Therefore in this case, there does not exist a 2^m-BFF design of resolution $\text{R}^*(\{1\}|\Omega_2)$.

[b] When $\text{NSV}_1 = 2$, i.e., $x_1^* = 1$ and $x_2^* = m-1$, we have $\text{r-rank}\{F_1(x_1^*, x_2^*)\} = 2$.

[1] When NSV_0=2, i.e., $x_i = x_i^*$ $(i = 1, 2)$, we have $\text{r-rank}\{F_0(x_1, x_2)\} = 2 < 3$. Thus from (8a), the elements of the last row of $F_0(x_1, x_2)$ must be the same, i.e., $f_0(x_1; m) = f_0(x_2; m)$, where $f_0(x; m)$ is given by (9). Since $x_1 + x_2 - m = 0$ and $\nu_2(m) - N \le \nu_2(m) - 2m = (m-4)(m+1)/2 + 3 > 0$ for $m \ge 4$, it follows from Lemma 3.1 that case (i) is established.

[2] When $\text{NSV}_0 = 2 + p$ $(p = 1, 2)$, i.e., $x_1 = 0$ or m (if $p = 1$) ($x_1 = 0$ (if $p = 2$)) and $x_{i+1} = x_i^*$ $(i = 1, 2)$ (and $x_4 = m$ (if $p = 2$)), we have $\text{r-rank}\{F_0(x_1, x_2, \cdots, x_{2+p})\} = 3$. Then $\nu_2(m) - N \le (m-4)(m+1)/2 + (3-p) > 0$ for $m \ge 4$ and $p = 1, 2$. Thus cases (ii)(1) and (iii) are proved.

[B] When $\text{NSV}_2 = 1$, i.e., $2 \le x_1^{**} \le m-2$, we have $\text{r-rank}\{F_2(x_1^{**})\} = 1$. When $\text{NSV}_1 = 3$, we have $N \ge 2m + \binom{m}{2} > \nu_2(m)$ for $m \ge 4$, and hence $(\text{NSV}_2 = 1 \le)\text{NSV}_1 \le 2$. Furthermore from Theorem 2.2, it must be $\text{NSV}_0 \ge 2$.

[a] When $\text{NSV}_1 = 1$, i.e., $x_1^* = x_1^{**}$, we have $\text{r-rank}\{F_1(x_1^*)\} = 1 < 2$. Thus from (8b), it must be $2x_1^* - m = 0$, and hence $x_1^* = m/2$, where $m = 2s \ge 4$. Furthermore

[1] when $\text{NSV}_0 = 2$, i.e., $x_1 = 0$ or m and $x_2 = x_1^*$, we have $\text{r-rank}\{F_0(x_1, x_2)\} = 2 < 3$. Thus it must be $f_0(x_1; m) = f_0(x_2; m)$ for $m \ge 4$. However $x_1 + x_2 - m \ne 0$ for $m \ge 4$, and hence, from Lemma 3.1, there does not exist a 2^m-BFF design of resolution $\text{R}^*(\{1\}|\Omega_2)$.

[2] When $\mathrm{NSV}_0 = 3$, i.e., $x_1 = 0$, $x_2 = x_1^*$ and $x_3 = m$, we have r-rank$\{F_0(x_1, x_2, x_3)\} = 3$. When $m = 4$, we have $8 \le N < \nu_2(4) = 11$, and when $m = 2s \ge 6$, $\nu_2(m) - N \le \nu_2(m) - \{2 + \binom{m}{s}\} \le \binom{m}{2} + m - \binom{m}{3} - 1 = -(m-6)\{(m-6)(m+6) + 35\}/6 \le 0$ for $m = 2s \ge 6$. Thus we get $m = 4$, and hence $x_1 = 0$, $x_2 = m/2 = 2$ and $x_3 = 4$, which is case (ii)(2).

[b] When $\mathrm{NSV}_1 = 2$, i.e., $x_1^* = 1$ or $m-1$ and $x_2^* = x_1^{**}$, we have r-rank$\{F_1(x_1^*, x_2^*)\} = 2$. In this case, $N \ge m + \binom{m}{2} = \nu_2(m) - 1$, and hence it must be $\mathrm{NSV}_0 = 2$, i.e., $x_i = x_i^*$ $(i = 1, 2)$. Then we have r-rank$\{F_0(x_1, x_2)\} = 2 < 3$. Thus from (8a), it must be $f_0(x_1; m) = f_0(x_2; m)$ for $m \ge 4$. However $x_1 + x_2 - m \ne 0$ for $m \ge 4$, and hence, from Lemma 3.1, a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_2)$ does not exist.

[C] When $\mathrm{NSV}_2 = q \ge 2$, where $(m-2) - 1 \ge q$, we have $N \ge q\binom{m}{2} \ge 2\binom{m}{2} > \nu_2(m)$ for $m \ge 3 + q$. Thus in this case, there does not exist a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_2)$ with $N < \nu_2(m)$ for $m \ge 3 + q$.

Proof of Theorem 4.1. Similarly to the proof of Theorem 3.1, it will be done by listing out all possible cases. In these cases, we have $SV_0 \supset SV_1 \supset SV_2 \supset SV_3$, and hence $\mathrm{NSV}_0 \ge \mathrm{NSV}_1 \ge \mathrm{NSV}_2 \ge \mathrm{NSV}_3$, and $\mathrm{NSV}_\gamma - \mathrm{NSV}_{\gamma+1} \le 2$ for $\gamma = 0, 1, 2$. Moreover when $\mathrm{NSV}_0 \ge 7$, $N \ge 1 + 1 + m + m + \binom{m}{2} + \binom{m}{2} + \binom{m}{3} > \nu_3(m)$ for $m \ge 6$, and hence $\mathrm{NSV}_0 \le 6$. We also begin the proof with $\mathrm{NSV}_3 = 0$.

[A] When $\mathrm{NSV}_3 = 0$ (and hence $\mathrm{NSV}_2 \le 2$), i.e., $\lambda_{x^{***}} = 0$ for any $x^{***} \in V_3$, it follows from Theorem 2.2 that $\mathrm{NSV}_1 \ge 1$ and $\mathrm{NSV}_0 \ge 2$. Moreover

[a] when $\mathrm{NSV}_2 = 0$ (and hence $\mathrm{NSV}_1 \le 2$), and furthermore

[1] when $\mathrm{NSV}_1 = 1$, i.e., $x_1^* = 1$ or $m-1$, we have r-rank$\{F_1(x_1^*)\} = 1 < 3$. Thus from Lemma 4.1.(II)(i), there does not exist a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_3)$.

[2] When $\mathrm{NSV}_1 = 2$, i.e., $x_1^* = 1$ and $x_2^* = m-1$, r-rank$\{F_1(x_1^*, x_2^*)\} = 2 < 3$. Then we have $(2x_1^* - m)(2x_2^* - m) + (m-2) = -(m-2)(m-3) < 0$ for $m \ge 6$. Therefore from Lemma 4.1.(II)(ii), a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_3)$ does not exist.

[b] When $\mathrm{NSV}_2 = 1$ (and hence $\mathrm{NSV}_1 \le 3$), i.e., $x_1^{**} = 2$ or $m-2$, we have r-rank$\{F_2(x_1^{**})\} = 1 < 2$.

[1] When $\mathrm{NSV}_1 = 1$, i.e., $x_1^* = x_1^{**}$, we have r-rank$\{F_1(x_1^*)\} = 1 < 3$. Then from Lemma 4.1.(II)(i), there does not exist a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_3)$.

[2] When $\mathrm{NSV}_1 = 2$, i.e., $x_1^* = 1$ or $m-1$ and $x_2^* = x_1^{**}$, we have r-rank$\{F_1(x_1^*, x_2^*)\} = 2 < 3$. When $x_1^* = 1$, $(2x_1^* - m)(2x_2^* - m) + (m-2) = (m-2)(m-3) > 0$ and $-(m-2)(m-5) < 0$ for $m \ge 6$ according as $x_2^* = 2$ and $m-2$, respectively. Thus, it follows from Lemma 4.1.(II)(ii) and the relation of the CSA that there does not exist a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_3)$.

[3] When $\mathrm{NSV}_1 = 3$ (and hence $\mathrm{NSV}_0 \le 5$), i.e., $x_1^* = 1$, $x_2^* = x_1^{**}$ and $x_3^* = m-1$, r-rank$\{F_1(x_1^*, x_2^*, x_3^*)\} = 3$. In addition,

[3.1] when $\mathrm{NSV}_0 = 3$, i.e., $x_i = x_i^*$ $(i = 1, 2, 3)$, r-rank$\{F_0(x_1, x_2, x_3)\} = 3 < 4$. Then $x_1 + x_2 - m \ne 0$ for $x_2 = 2$ or $m-2$, and $m \ge 6$. Furthermore from Lemma 4.1.(I)(ii), we have $(2x_1 - m)(2x_2 - m) + (2x_2 - m)(2x_3 - m) + (2x_3 - m)(2x_1 - m) + (3m - 2) = -(m-1)(m-6) = 0$ for $x_2 = 2$ or $m-2$. Thus we get $m = 6$, and hence $x_3 = m - 1 = 5$. In this case, $27 \le N < \nu_3(6) = 42$. Therefore case (i)(1) is established. Here if $x_2 = 2$, then $d_0 = 10/3$ and $d_2 = -2/3$, and if $x_2 = m - 2 = 4$, then $d_0 = -10/3$ and $d_2 = 2/3$, where d_k $(k = 0, 2)$ are constants given in Lemma 4.1.(I)(ii).

[3.2] When $\mathrm{NSV}_0 = 3 + p$ $(p = 1, 2)$, i.e., $x_1 = 0$ or m (if $p = 1$) $(x_1 = 0$ (if $p = 2$)), $x_2 = 1$, $x_3 = 2$ or $m-2$, and $x_4 = m-1$ (and $x_5 = m$ (if $p = 2$)), we have r-rank$\{F_0(x_1, x_2, \cdots, x_{3+p})\} = 4$. Then $\binom{m}{2} + 2m + p \le N < \nu_3(m)$ for $m \ge 6$ and $p = 1, 2$. Thus we obtain cases (ii)(1) and (2), and (iii)(1).

[c] When $\mathrm{NSV}_2 = 2$ (and hence $\mathrm{NSV}_1 \le 4$), i.e., $x_1^{**} = 2$ and $x_2^{**} = m-2$, r-rank$\{F_2(x_1^{**}, x_2^{**})\} = 2$.

[1] When $\mathrm{NSV}_1 = 2$ (and hence $\mathrm{NSV}_0 \le 4$), i.e., $x_j^* = x_j^{**}$ $(j = 1, 2)$, r-rank$\{F_1(x_1^*, x_2^*)\} = 2 < 3$. Thus from Lemma 4.1.(II)(ii), $(2x_1^* - m)(2x_2^* - m) + (m-2) = -(m-3)(m-6) = 0$ for $m \ge 6$, and hence we get $m = 6$ and $x_2^* = m - 2 = 4$. In this case, $d^* = 0$, where d^* is a constant given in Lemma 4.1.(II)(ii). Furthermore

[1.1] when $\mathrm{NSV}_0 = 2$, i.e., $x_i = x_i^*$ $(i = 1, 2)$, r-rank$\{F_0(x_1, x_2)\} = 2 < 4$. Since $m = 6$, from Lemma 4.1.(I)(i)(1), we get $t = 1$, and hence $x_1 = 1$ and $x_2 = 5$, which contradict $x_1 = 2$ and $x_2 = 4$. Thus in this case, there does not exist a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_3)$.

[1.2] When $\mathrm{NSV}_0 = 3$, i.e., $x_1 = 0$ or $m(= 6)$ and $x_{i+1} = x_i^*$ $(i = 1, 2)$, r-rank$\{F_0(x_1, x_2, x_3)\} = 3 < 4$. Since $m = 6$, we have $x_1 + x_2 - m = -4 \neq 0$ and $(2x_1 - m)(2x_2 - m) + (2x_2 - m)(2x_3 - m) + (2x_3 - m)(2x_1 - m) + (3m - 2) = 12 \neq 0$ for $x_1 = 0$. Therefore from Lemma 4.1.(I)(ii) and the relation of the CSA, a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_3)$ does not exist.

[1.3] When $\mathrm{NSV}_0 = 4$, i.e., $x_1 = 0$, $x_{i+1} = x_i^*$ $(i = 1, 2)$ and $x_4 = m(= 6)$, r-rank$\{F_0(x_1, x_2, \cdots, x_4)\} = 4$. In this case, $32 \leq N < \nu_3(6) = 42$, and hence case (ii)(3) is proved.

[2] When $\mathrm{NSV}_1 = 3$ (and hence $\mathrm{NSV}_0 \leq 5$), i.e., $x_1^* = 1$ or $m - 1$ and $x_{j+1}^* = x_j^{**}$ $(j = 1, 2)$, we have r-rank$\{F_1(x_1^*, x_2^*, x_3^*)\}$ =3.

[2.1] When $\mathrm{NSV}_0 = 3$, i.e., $x_i = x_i^*$ $(i = 1, 2, 3)$, r-rank$\{F_0(x_1, x_2, x_3)\} = 3 < 4$. Then $x_1 + x_2 - m = -(m-3)(\neq 0)$ and $1(\neq 0)$ for $m \geq 6$ according as $x_1 = 1$ and $m - 1$, respectively. Furthermore $(2x_1 - m)(2x_2 - m) + (2x_2 - m)(2x_3 - m) + (2x_3 - m)(2x_1 - m) + (3m - 2) = -(m - 2)(m - 9)$ for $x_1 = 1$ or $m - 1$. Thus from Lemma 4.1.(I)(ii), we get $m = 9$, and hence $x_1 = 1$ or $m - 1 = 8$ and $x_3 = m - 2 = 7$. In this case, $81 \leq N < \nu_3(9) = 130$. Thus case (i)(2) is established. If $x_1 = 1$, $x_2 = 2$ and $x_3 = 7$, then we get $d_0 = 56/3$ and $d_2 = -7/3$, and if $x_1 = 8$, $x_2 = 2$ and $x_3 = 7$, then $d_0 = -56/3$ and $d_2 = 7/3$.

[2.2] When $\mathrm{NSV}_0 = 3 + p$ $(p = 1, 2)$, i.e., $x_1 = 0$ or m (if $p = 1$) $(x_1 = 0$ (if $p = 2$)) and $x_{i+1} = x_i^*$ $(i = 1, 2, 3)$(and $x_5 = m$ (if $p = 2$)), we have r-rank$\{F_0(x_1, x_2, \cdots, x_{3+p})\} = 4$. Furthermore $\nu_3(m) - N \leq (m - 6)(m^2 + 5)/6 + (6 - p) > 0$ for $m \geq 6$ and $p = 1, 2$. Thus cases (ii)(4) and (5), and (iii)(2) are proved.

[3] When $\mathrm{NSV}_1 = 4$, i.e., $x_1^* = 1, x_2^* = 2, x_3^* = m - 2$ and $x_4^* = m - 1$, r-rank$\{F_1(x_1^*, x_2^*, \cdots, x_4^*)\} = 3$. Furthermore when $\mathrm{NSV}_0 = 4 + p$ $(p = 0, 1, 2)$, i.e., when $p = 0$, we have $x_i = x_i^*$ $(i = 1, 2, \cdots, 4)$, when $p = 1$, $x_1 = 0$ or m and $x_{i+1} = x_i^*$ $(i = 1, 2, \cdots, 4)$, and when $p = 2$, $x_1 = 0$, $x_{i+1} = x_i^*$ $(i = 1, 2, \cdots, 4)$ and $x_6 = m$, we have r-rank$\{F_0(x_1, x_2, \cdots, x_{4+p})\} = 4$. Then $\nu_3(m) - N \leq (m - 6)(m^2 - 1)/6 - p = (m - 7)(m^2 + m + 6)/6 + (8 - p)$ for $m \geq 6$ and $p = 0, 1, 2$. Thus when $m = 6$, we have $\nu_3(6) - N \leq 0$ for $p = 0, 1, 2$, and when $m \geq 7$, $\nu_3(m) - N > 0$ for $p = 0, 1, 2$. Therefore we establish cases (ii)(6), (iii)(3) and (iv).

[B] When $\mathrm{NSV}_3 = 1$ (and hence $\mathrm{NSV}_2 \leq 3$), i.e., $3 \leq x_1^{***} \leq m - 3$, it follows from Theorem 2.2 that r-rank$\{F_3(x_1^{***})\} = 1$, and it must be $\mathrm{NSV}_0 \geq 2$. In addition,

[a] when $\mathrm{NSV}_2 = 1$ (and hence $\mathrm{NSV}_1 \leq 3$), i.e., $x_1^{**} = x_1^{***}$, we have r-rank$\{F_2(x_1^{**})\} = 1 < 2$.

[1] When $\mathrm{NSV}_1 = 1$, i.e., $x_1^* = x_1^{**}$, r-rank$\{F_1(x_1^*)\} = 1 < 3$. Thus from Lemma 4.1.(II)(i), there does not exist a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_3)$.

[2] When $\mathrm{NSV}_1 = 2$ (and hence $\mathrm{NSV}_0 \leq 4$), i.e., $x_1^* = 1$ or $m - 1$ and $x_2^* = x_1^{**}$, r-rank$\{F_1(x_1^*, x_2^*)\} = 2 < 3$. Then from Lemma 4.1.(II)(ii), it must be $(2x_1^* - m)(2x_2^* - m) + (m - 2) = 0$ for $m \geq 6$. We consider case $x_1^* = 1$. Then we have $(2x_1^* - m)(2x_2^* - m) + (m - 2) = -(m - 2)\{2x_2^* - (m + 1)\} = 0$ for $m \geq 6$, and hence $x_2^* = (m + 1)/2$, where $m = 2s + 1 \geq 7$. Moreover

[2.1] when $\mathrm{NSV}_0 = 2$, i.e., $x_i = x_i^*$ $(i = 1, 2)$, we have r-rank$\{F_0(x_1, x_2)\} = 2 < 4$. Since $x_1 = 1$, from Lemma 4.1.(I)(i)(1), we get $t = 1$, and hence $m = 6$ and $x_2 = 5$. However $m = 6 < 7$ and $5 \notin SV_3$ for $m = 6$. Thus from Lemma 4.1.(I)(i)(1) and the relation of the CSA, a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_3)$ does not exist.

[2.2] When $\mathrm{NSV}_0 = 3$, i.e., $x_1 = 0$ or m and $x_{i+1} = x_i^*$ $(i = 1, 2)$, we have r-rank$\{F_0(x_1, x_2, x_3)\} = 3 < 4$. Then it holds $x_p + x_q - m \neq 0$ for any $\{p, q\} \subset \{1, 2, 3\}$ and $m \geq 7$. However when $x_2 = 1$, $(2x_1 - m)(2x_2 - m) + (2x_2 - m)(2x_3 - m) + (2x_3 - m)(2x_1 - m) + (3m - 2) = m(m - 1) > 0$ and $-m(m - 5) < 0$ for $m \geq 7$ according as $x_1 = 0$ and m, respectively. Therefore it follows from Lemma 4.1.(I)(ii) and the relation of the CSA that there does not exist a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_3)$.

[2.3] When $\mathrm{NSV}_0 = 4$, i.e., $x_1 = 0$ and $x_{i+1} = x_i^*$ $(i = 1, 2)$ and $x_4 = m$, r-rank$\{F_0(x_1, x_2, \cdots, x_4)\} = 4$. In this case, when $m = 7$, we have $44 \leq N < \nu_3(7) = 64$, and when $m = 2s + 1 \geq 9$, $\nu_3(m) - N \leq \binom{m}{2} + \binom{m}{3} - \binom{m}{4} - 1 < \binom{m}{2} + \binom{m}{3} - \binom{m}{4} = -[(m - 9)\{(m - 9)(m^2 + 8m + 74) + 682\}/24 + 6] < 0$ for $m \geq 9$. Thus we get $m = 7$, and hence $x_1 = 0$, $x_2 = 1$ or $m - 1 = 6$, $x_3 = (m + 1)/2 = 4$ and $x_4 = m = 7$. Therefore case (ii)(7) is proved.

[3] When $\mathrm{NSV}_1 = 3$ (and hence $\mathrm{NSV}_0 \leq 5$), i.e., $x_1^* = 1$, $x_2^* = x_1^{**}$ and $x_3^* = m - 1$, we have r-rank$\{F_1(x_1^*, x_2^*, x_3^*)\}$ = 3.

[3.1] When NSV_0=3, i.e., $x_i = x_i^*$ $(i = 1, 2, 3)$, r-rank$\{F_0(x_1, x_2, x_3)\} = 3 < 4$. Then $x_1 + x_2 - m < 0$ for $m \geq 6$. Furthermore from Lemma 4.1.(I)(ii), it must be $(2x_1 - m)(2x_2 - m) + (2x_2 - m)(2x_3 - m) + (2x_3 - m)(2x_1 - m) + (3m - 2) = -(m - 1)(m - 6) = 0$ for $m \geq 6$. Thus we get $m = 6$, and hence, $x_1 = 1$, $x_2 = 3$ and $x_3 = m - 1 = 5$. In this case, $32 \leq N < \nu_3(6) = 42$, and hence we have case (i)(3), and $d_0 = d_2 = 0$.

[3.2] When $\mathrm{NSV}_0 = 3 + p$ ($p = 1, 2$), i.e., $x_1 = 0$ or m (if $p = 1$) ($x_1 = 0$ (if $p = 2$)), $x_{i+1} = x_i^*$ ($i = 1, 2$) and $x_4 = m - 1$ (and $x_5 = m$ (if $p = 2$)), r-rank$\{F_0(x_1, x_2, \cdots, x_{3+p})\} = 4$. When $x_3 = 3$ or $m - 3$ (if $m \geq 7$), $\nu_3(m) - N \leq (m-6)(m+3)/2 + (10 - p) > 0$ for $m \geq 6$ and $p = 1, 2$, and hence we obtain cases (ii)(8) and (9), and (iii)(4). When $x_3 = 4$ and $m = 8$, $\nu_3(8) - N \leq 7 - p > 0$ for $p = 1, 2$. Thus cases (ii)(10) and (iii)(5) are obtained. Furthermore when $4 \leq x_3 \leq m - 4$ and $m \geq 9$, then $N \geq p + 2m + \binom{m}{x_3} \geq p + 2m + \binom{m}{4}$, and hence $\nu_3(m) - N \leq \binom{m}{2} + \binom{m}{3} - m - \binom{m}{4} - (p-1) < \binom{m}{2} + \binom{m}{3} - \binom{m}{4} < 0$ for $m \geq 9$ and $p = 1, 2$ (see [B][a][2][2.3] above). Thus there does not exist a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_3)$ with $N < \nu_3(m)$ for $m \geq 9$.

[b] When NSV_2=2 (and hence $\mathrm{NSV}_1 \leq 4$), i.e., $x_1^{**} = 2$ or $m - 2$ and $x_2^{**} = x_1^{***}$, we have r-rank$\{F_2(x_1^{**}, x_2^{**})\} = 2$.

[1] When NSV_1=2, i.e., $x_j^* = x_j^{**}$ ($j = 1, 2$), we have r-rank$\{F_1(x_1^*, x_2^*)\} = 2 < 3$. Thus from Lemma 4.1.(II)(ii), it must be $(2x_1^* - m)(2x_2^* - m) + (m - 2) = 0$ for $m \geq 6$. When $x_1^* = 2$, we have $x_2^* = (m+1)/2 + 1/(m-4)$, which is an integer for $m = 6$ only, and hence $x_2^* = 4$. However 4 does not belong to SV_3 for $m = 6$. Therefore from Lemma 4.1.(II)(ii) and the relation of the CSA, there does not exist a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_3)$.

[2] When $\mathrm{NSV}_1 = 3$ (and hence $\mathrm{NSV}_0 \leq 5$), i.e., $x_1^* = 1$ or $m - 1$ and $x_{j+1}^* = x_j^{**}$ ($j = 1, 2$), we have r-rank$\{F_1(x_1^*, x_2^*, x_3^*)\} = 3$. In this case, $N \geq m + \binom{m}{2} + \binom{m}{x_3^*}$ for $m \geq 6$. Thus when $4 \leq x_3^* \leq m - 4$ and $m \geq 8$, $N \geq m + \binom{m}{2} + \binom{m}{4} > \nu_3(m)$, and hence we only consider case $x_3^* = 3$ or $m - 3$ (if $m \geq 7$) for $m \geq 6$.

[2.1] When $\mathrm{NSV}_0 = 3$, i.e., $x_i = x_i^*$ ($i = 1, 2, 3$), we have r-rank$\{F_0(x_1, x_2, x_3)\} = 3 < 4$. Then $x_s + x_t - m \neq 0$ for any $\{s, t\} \subset \{1, 2, 3\}$. Moreover when $x_1 = 1$, $(2x_1 - m)(2x_2 - m) + (2x_2 - m)(2x_3 - m) + (2x_3 - m)(2x_1 - m) + (3m - 2) = 3(m^2 - 7m + 14)$, $-(m^2 - 15m + 30)$, $-(m^2 - 11m + 22)$ and $-(m^2 - 7m - 2)$ according as $x_2 = 2$ and $x_3 = 3$, $x_2 = 2$ and $x_3 = m - 3$, $x_2 = m - 2$ and $x_3 = 3$, and $x_2 = m - 2$ and $x_3 = m - 3$, respectively. However $m^2 - 7m + 14 = (m-6)(m-1) + 8 > 0$ for $m \geq 6$, and furthermore three quadratic equations $m^2 - 15m + 30 = 0$, $m^2 - 11m + 22 = 0$ and $m^2 - 7m - 2 = 0$ do not have an integer solution for $m \geq 6$. Therefore from Lemma 4.1.(I)(ii) and the relation of the CSA, there does not exist a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_3)$.

[2.2] When $\mathrm{NSV}_0 = 3 + p$ ($p = 1, 2$), we have $N \geq p + m + \binom{m}{2} + \binom{m}{3} \geq \nu_3(m)$ for $m \geq 6$ and $p = 1, 2$. Thus there does not exist a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_3)$ with $N < \nu_3(m)$.

[3] When $\mathrm{NSV}_1 = 4$, we have $N \geq 2m + \binom{m}{2} + \binom{m}{3} > \nu_3(m)$ for $m \geq 6$. Hence there does not exist a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_3)$ with $N < \nu_3(m)$.

[c] When $\mathrm{NSV}_2 = 3$, we have $N \geq 2\binom{m}{2} + \binom{m}{3} > \nu_3(m)$, and hence a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_3)$ with $N < \nu_3(m)$ does not exist.

[C] When $\mathrm{NSV}_3 = q \geq 2$, where $(m - 3) - 2 \geq q$, it holds that $\binom{m}{3} > 1 + m + \binom{m}{2}$ for $m \geq 5 + q \geq 7$. Thus we have $N \geq q\binom{m}{3} \geq 2\binom{m}{3} > \nu_3(m)$. Therefore there does not exist a 2^m-BFF design of resolution $\mathrm{R}^*(\{1\}|\Omega_3)$ with $N < \nu_3(m)$ for $m \geq 5 + q$.

Therefore the proof is complete.

Acknowledgments

The authors would like to thank the editor and the reviewer for their helpful comments and suggestions that have lead to an improved version of this paper.

References

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

Bose, R. C., & Srivastava, J. N. (1964). Multidimensional partially balanced designs and their analysis, with applications to partially balanced factorial fractions. *Sankhyā: The Indian Journal of Statistics*, *26*, 145-168.

Chakravarti, I. M. (1956). Fractional replication in asymmetrical factorial designs and partially balanced arrays. *Sankhyā: The Indian Journal of Statistics*, *17*, 143-164.

Ghosh, S., & Kuwada, M. (2001). Some estimable parametric functions for balanced fractional 3^m factorial designs. *Technical Report #01-6*, *Statistical Research Group*, Hiroshima University, Japan.

Hyodo, Y. (1989). Structure of fractional factorial designs derived from two-symbol balanced arrays and their resolution. *Hiroshima Mathematical Journal*, *19*, 457-475.

Hyodo, Y. (1992). Characteristic polynomials of information matrices of some balanced fractional 2^m factorial designs of resolution $2\ell + 1$. *Journal of Statistical Planning and Inference*, *31*, 245-252. http://dx.doi.org/10.1016/0378-3758(92)90033-O

Hyodo, Y., & Yamamoto, S. (1988). Algebraic structure of information matrices of fractional factorial designs derived from simple two-symbol balanced arrays and its applications. In K. Matusita (Ed.), *Statistical Theory and Data Analysis II* (pp. 457-468). North-Holland, Amsterdam.

Hyodo, Y., Yumiba, H., & Kuwada, M. (2015). Existence conditions for balanced fractional 2^m factorial designs of resolution $2\ell + 1$ derived from simple arrays. *Communications in Statistics-Theory and Methods*, *44,* 2564-2570. http://dx.doi.org/10.1080/03610926.2013.788716

Kuriki, S., & Yamamoto, S. (1984). Nonsimple 2-symbol balanced arrays of strength t and $t + 2$ constraints. *TRU Mathematics*, *20*, 249-263.

Kuwada, M. (1982). On some optimal 2^m factorial designs of resolution V. *Journal of Statistical Planning and Inference*, *7*, 39-48. http://dx.doi.org/10.1016/0378-3758(82)90018-0

Kuwada, M., Hyodo, Y., & Han, D. (2003). Characterization of balanced fractional 2^m factorial designs of resolution $R^*(\{1\}|3)$ and GA-optimal designs. *Journal of the Japan Statistical Society*, *33*, 181-201.

Shirakura, T. (1975). On balanced arrays of 2 symbols, strength 2ℓ, m constraints and index set $\{\mu_0, \mu_1, \cdots, \mu_{2\ell}\}$ with $\mu_\ell = 0$. *Journal of the Japan Statistical Society*, *5*, 53-56.

Shirakura, T. (1977). Contributions to balanced fractional 2^m factorial designs derived from balanced arrays of strength 2ℓ. *Hiroshima Mathematical Journal*, *7*, 217-285.

Shirakura, T. (1980). Necessary and sufficient condition for a balanced array of strength 2ℓ to be a balanced fractional 2^m factorial design of resolution 2ℓ. *The Australian Journal of Statistics*, *22*, 69-74. http://dx.doi.org/10.1111/j.1467-842X.1980.tb01155.x

Shirakura, T., & Kuwada, M. (1975). Note on balanced fractional 2^m factorial designs of resolution $2\ell + 1$. *Annals of the Institute of Statistical Mathematics*, *27*, 377-386. http://dx.doi.org/10.1007/BF02504657

Shirakura, T., & Kuwada, M. (1976). Covariance matrices of the estimates for balanced fractional 2^m factorial designs of resolution $2\ell + 1$. *Journal of the Japan Statistical Society*, *6*, 27-31.

Srivastava, J. N. (1970). Optimal balanced 2^m fractional factorial designs. In R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao, & K. J. C. Smith (Eds.), *Essays in Probability and Statistics* (pp. 689-706). The University of North Carolina Press, Chapel Hill.

Srivastava, J. N. (1972). Some general existence conditions for balanced arrays of strength t and 2 symbols. *Journal of Combinatorial Theory*, *13*, 198-206.

Srivastava, J. N., & Chopra, D. V. (1971). On the characteristic roots of the information matrix of 2^m balanced factorial designs of resolution V, with applications. *Annals of Mathematical Statistics*, *42*, 722-734. http://dx.doi.org/10.1214/aoms/1177693421

Yamamoto, S., & Hyodo, Y. (1984). Extended concept of resolution and the design derived from balanced arrays. *TRU Mathematics*, *20*, 341-349.

Yamamoto, S., Shirakura, T., & Kuwada, M. (1975). Balanced arrays of strength 2ℓ and balanced fractional 2^m factorial designs. *Annals of the Institute of Statistical Mathematics*, *27*, 143-157. http://dx.doi.org/10.1007/BF02504632

Yamamoto, S., Shirakura, T., & Kuwada, M. (1976). Characteristic polynomials of the information matrices of balanced fractional 2^m factorial designs of higher $(2\ell + 1)$ resolution. In S. Ikeda, T. Hayakawa, H. Hudimoto, M. Okamoto, M. Siotani, & S. Yamamoto (Eds.), *Essays in Probability and Statistics* (pp. 73-94). Shinko Tsusho, Tokyo.

International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Mean-correction and Higher Order Moments for a Stochastic Volatility Model with Correlated Errors

Sujay Mukhoti[1] & Pritam Ranjan[1]

[1] Operations Management and Quantitative Techniques, Indian Institute of Management Indore, Madhya Pradesh, India

Correspondence: Sujay Mukhoti, C-111, OM&QT, Prabandh Shikhar, Rau-Pithampur Road, IIM Indore, M.P., India, 453556. Tel: 91-731-243-9487. E-mail: sujaym@iimidr.ac.in

Received: May 6, 2016 Accepted: June 6, 2016 Online Published: June 22, 2016

doi:10.5539/ijsp.v5n4p102 URL: http://dx.doi.org/10.5539/ijsp.v5n4p102

Abstract

In an efficient stock market, the log-returns and their time-dependent variances are often jointly modelled by stochastic volatility models (SVMs). Many SVMs assume that errors in log-return and latent volatility process are uncorrelated, which is unrealistic. It turns out that if a non-zero correlation is included in the SVM (e.g., Shephard (2005)), then the expected log-return at time t conditional on the past returns is non-zero, which is not a desirable feature of an efficient stock market. In this paper, we propose a mean-correction for such an SVM for discrete-time returns with non-zero correlation. We also find closed form analytical expressions for higher moments of log-return and its lead-lag correlations with the volatility process. We compare the performance of the proposed and classical SVMs on S&P 500 index returns obtained from NYSE.

Keywords: Leverage Effect, Martingale Difference, Return Skewness, Volatility Asymmetry.

1. Introduction

Over the last few decades different aspects of stock price movements in discrete time have been the focus of numerous research avenues. Suppose P_t denotes the price of a stock at time t, then the continuously compounded return or log-return (here onwards referred to as return) of the stock is defined as $r_t = \log(P_t/P_{t-1})$. A stock market is said to be efficient if the price of a stock contains every available information about it. In such a market the risk involved in investing on a stock is measured by the standard deviation of r_t, often termed as the *volatility* of the stock in finance literature. It has been noted that volatility varies over time (Engle (1982)). Stochastic Volatility Models (SVMs) is a popular class of models for describing the time-varying volatility of stock returns (Shephard (2005)).

Although there are a plethora of SVMs for describing the stock returns, one of the simplest yet most popular discrete-time SVM is given by Taylor (1982), where the return process r_t is a non-linear product of two independent stochastic processes, viz. an i.i.d. error process ϵ_t, and a latent volatility process h_t, which is further modelled as an $AR(1)$. That is,

$$\begin{aligned} r_t &= \exp\left\{\frac{h_t}{2}\right\}\epsilon_t \\ h_t &= \alpha + \phi(h_{t-1} - \alpha) + \sigma\eta_t, \ \forall t = 1, 2, \ldots, \end{aligned} \tag{1}$$

where $\alpha = E(h_t)$ is the long-range volatility, ϕ is the stationarity parameter, σ measures the variability of the volatility process h_t, and ϵ_t and η_t are uncorrelated i.i.d. $N(0, 1)$ errors. Hereafter this model will be referred as SVM_0.

As in (1), many of the new generation SVMs which are being used in the finance literature assume that ϵ_t and η_t are independent $N(0, 1)$ errors. In reality, however, ϵ_t and η_t are often correlated (Harvey & Siddique (1999)). Though discrete-time SVMs with non-zero $corr(\epsilon_t, \eta_t)$ have been developed earlier and are being used, they assume that h_{t+1} (instead of h_t as in (1)) depends on η_t via AR(1) (see *e.g.* Meyer & Yu (2000); Berg *et al.* (2004)). In this paper, we focus on the SVM presented in (1) with correlated errors (denoted as SVM_ρ). That is, the additional assumption in (1) is $corr(\epsilon_t, \eta_t) = \rho$ (Jacquier *et al.* (2004)).

It turns out that introducing a non-zero correlation between η_t and ϵ_t in (1) has an adverse effect on the admissibility of the SVM from an efficient market's viewpoint. In particular, the conditional expectation of r_t given the past data, $E[r_t \mid \mathcal{F}_{t-1}]$, is not zero, where $\mathcal{F}_{t-1}$ is the space (σ-field) generated with $r_1, ..., r_{t-1}$. This zero conditional expectation of the return is a necessary requirement for an *efficient market hypothesis* (EMH) (see Yu (2005) for a review).

In this paper, we propose a mean-correction for SVM_ρ - model (1) with correlated errors, such that $E[r_t \mid \mathcal{F}_{t-1}]$ becomes zero and the corrected SVM would satisfy EMH. The proposed mean-corrected model is denoted by $SVM_{\rho\mu}$. Further, Black (1976) mentioned that, usually, the amount of increment in volatility due to price fall is larger than the magnitude

of reduction in the volatility due to price increase. In turn, this indicates the volatility of positive returns, $var(r_t|r_t > 0)$, is less than the volatility of the negative returns, $var(r_t|r_t < 0)$ resulting in skewness in return distribution. Moreover, the kurtosis quantifies the proportion of extreme values, that occur during crashes, explained by the model. We find the closed form expressions for the higher-order moments and the lead-lag correlation of the underlying return process. These descriptive statistics indicate the influence of past/future volatility on today's return.

The remainder of the article is organized as follows. Section 2 presents the main results: $SVM_{\rho\mu}$ - the mean-corrected SVM with non-zero correlation that satisfies EMH, and the closed form analytical expressions for the higher order moments and lead-lag correlation for the proposed model. For the returns of S&P 500 NSYE, Section 3 presents a comparison between the standard zero correlation model (1) and the ones with non-zero correlation. Finally Section 4 outlines the concluding remarks and a few possible future directions.

2. Main Results

For this section, we assume that the error terms ϵ_t and η_t in (1) have not only a constant correlation ρ and i.i.d. $N(0, 1)$ marginals, but they also follow a bivariate normal distribution. The proposed mean-corrected model ($SVM_{\rho\mu}$) contains an additional term μ, i.e.,

$$\begin{aligned} r_t &= \mu + \exp\left\{\frac{h_t}{2}\right\}\epsilon_t \\ h_t &= \alpha + \phi(h_{t-1} - \alpha) + \sigma\eta_t, \ \forall t = 1, 2, \ldots T. \end{aligned} \tag{2}$$

Theorem 1 establishes the value of μ for which the proposed mean-corrected model (2) gives zero conditional expectation $E[r_t \mid \mathcal{F}_{t-1}]$ and hence satisfy EMH. Later in this section, we derive closed form expressions for the higher-order moments, i.e., variance, skewness, and kurtosis of r_t, and lead-lag correlations between r_t and $h_{t\pm k}$.

Theorem 1. *For $SVM_{\rho\mu}$ in (2) with $|\phi| \leq 1, \sigma > 0$ and $-\infty < \alpha < \infty$, if (ϵ_t, η_t) follows a standard bivariate normal distribution with correlation ρ, the mean term*

$$\mu = -\frac{\rho\sigma}{2}\exp\left\{\frac{\alpha}{2} + \frac{\sigma^2}{8(1-\phi^2)}\right\} \tag{3}$$

gives $E[r_t \mid \mathcal{F}_{t-1}] = 0$ and vice-versa.

Proof. The conditional expected return $E[r_t \mid \mathcal{F}_{t-1}] = 0$ gives

$$\begin{aligned} -\mu &= E\left[\exp\left\{\frac{h_t}{2}\right\}\epsilon_t\right] = E\left[\exp\left\{\frac{\alpha + \phi(h_{t-1} - \alpha) + \sigma\eta_t}{2}\right\}\epsilon_t\right] \\ &= \exp\left\{\frac{\alpha}{2}\right\} \times E\left[\exp\left\{\frac{\phi\sigma}{2}\sum_{j=1}^{\infty}\phi^{j-1}\eta_{t-j}\right\}\right] \times E\left[\exp\left\{\frac{\sigma\eta_t}{2}\right\}\epsilon_t\right]. \end{aligned} \tag{4}$$

Since (ϵ_t, η_t) follows a standard bivariate normal with correlation ρ, the condition distribution of $\epsilon_t \mid \eta_t$ is given by $N\left(\rho\eta_t, 1-\rho^2\right)$. This conditional normal distribution and the moment generating function (mgf) of a normal distribution simplifies the third term in (4) as

$$E\left[\exp\left\{\frac{\sigma\eta_t}{2}\right\}\epsilon_t\right] = E_{\eta_t}\left[\exp\left\{\frac{\sigma\eta_t}{2}\right\}\rho\eta_t\right] = \frac{\rho\sigma}{2}\exp\left\{\frac{\sigma^2}{8}\right\}, \tag{5}$$

and the second term to

$$\prod_{j=1}^{\infty} E\left[\exp\left\{\frac{\sigma\phi^j}{2}\eta_{t-j}\right\}\right] = \exp\left\{\frac{\sigma^2}{8}\sum_{j=1}^{\infty}\phi^{2j}\right\} = \exp\left\{\frac{\sigma^2\phi^2}{8(1-\phi^2)}\right\}. \tag{6}$$

Hence the final expression for μ follows from (4)-(6). □

Yu (2005) tried to compute $E[r_t \mid \mathcal{F}_{t-1}]$, but the final expression appears to be incorrect. Note that the proposed mean-correction (in Theorem 1) makes the model (2) usable in the stock market, as it now satisfies EMH (in particular, $E[r_t \mid \mathcal{F}_{t-1}] = 0$). Further, the proof of the above theorem prohibits the usage of heavy-tail distributions (like t distribution) as the volatility error distribution (Wang *et al.* (2011)) as its moment generating function would not exist resulting in

in-existence of expected returns. In Section 3, we discuss the usage of this model for the index returns of S&P500 index of New York Stock Exchange (NYSE) observed during 1^{st} April, 2002 - 30^{th} March, 2006.

2.1 Higher-order moments

For additional key features on the distribution of returns, we estimate higher order moments, in particular, variance, skewness and kurtosis conditional on $\mathcal{F}_{t-1}$.

Theorem 2. *For $SVM_{\rho\mu}$ in (2), if Theorem 1 holds, then the variance of returns conditional on $\mathcal{F}_{t-1}$ is given by*

$$V(r_t \mid \mathcal{F}_{t-1}) = \exp\left\{\alpha + \frac{\sigma^2}{2(1-\phi^2)}\right\}\left(1 + \rho^2\sigma^2 - \frac{\rho^2\sigma^2}{4}\exp\left\{-\frac{\sigma^2}{4(1-\phi^2)}\right\}\right). \tag{7}$$

Proof. Following the definition of variance,

$$\begin{aligned} V(r_t \mid \mathcal{F}_{t-1}) &= E[r_t^2 \mid \mathcal{F}_{t-1}] - 0^2 \\ &= E\left[\exp\{h_t\}\epsilon_t^2\right] - \mu^2 \\ &= \exp\{\alpha\} \times E\left[\exp\left\{\sigma\sum_{j=1}^{\infty}\phi^j\eta_{t-j}\right\}\right] \times E\left[\exp\{\sigma\eta_t\}\,\epsilon_t^2\right] - \mu^2 \\ &= \exp\left\{\alpha + \frac{\sigma^2\phi^2}{2(1-\phi^2)}\right\}(1+\rho^2\sigma^2)\exp\left\{\frac{\sigma^2}{2}\right\} - \mu^2 \quad \text{(as in (4)-(6)).} \end{aligned}$$

The final result follows by substituting the value of μ from Theorem 1. □

The expressions of the conditional mean and variance are the most crucial components in finding the skewness and kurtosis statistics. For $SVM_{\rho\mu}$ in (2), under the same conditions as in Theorem 2, the skewness conditional on $\mathcal{F}_{t-1}$ is measured by $\mu_3/(Var(r_t \mid \mathcal{F}_{t-1}))^{3/2}$, where

$$\begin{aligned} \mu_3 &= \frac{3\rho\sigma}{2}exp\left\{\frac{3\alpha}{2} + \frac{9\sigma^2}{8(1-\phi^2)}\right\}\left[3 + \frac{9\sigma^2\rho^2}{4} + \frac{\rho^2\sigma^2}{6}exp\left\{-\frac{3\sigma^2}{4(1-\phi^2)}\right\}\right. \\ &\quad \left. -\left(1+\rho^2\sigma^2\right)exp\left\{-\frac{\sigma^2}{2(1-\phi^2)}\right\}\right]. \end{aligned} \tag{8}$$

The proof of (8) starts with $\mu_3 = E[r_t^3 \mid \mathcal{F}_{t-1}]$, and proceeds in the exact same manner as in Theorems 1 and 2. Similarly the closed form expression of kurtosis can also be found as $\mu_4/(Var(r_t \mid \mathcal{F}_{t-1}))^2$, where

$$\begin{aligned} \mu_4 &= \exp\left\{2\alpha + \frac{2\sigma^2}{(1-\phi^2)}\right\} \times \left[\frac{3}{2}\rho^2\sigma^2(1+\sigma^2\rho^2)\exp\left\{\frac{-5\sigma^2}{4(1-\phi^2)}\right\}\right. \\ &\quad + \left(3 + 24\rho^2\sigma^2 + 16\rho^4\sigma^4\right) - \frac{3}{16}\rho^4\sigma^4\exp\left\{\frac{-3}{2}\frac{\sigma^2}{(1-\phi^2)}\right\} \\ &\quad \left. -9\rho^2\sigma^2\left(1 + \frac{3}{4}\rho^2\sigma^2\right)\exp\left\{\frac{-3\sigma^2}{4(1-\phi^2)}\right\}\right]. \end{aligned} \tag{9}$$

As expected all four descriptive statistics found here depends heavily on $corr(\epsilon_t, \eta_t) = \rho$. On a closer inspection of these statistics, we see that $\rho = 0$ (i.e., the classical SVM by Taylor (1982)) gives $\mu = 0, \mu_3 = 0$,

$$Var(r_t \mid \mathcal{F}_{t-1}) = \exp\left\{\alpha + \frac{\sigma^2}{2(1-\phi^2)}\right\} \quad \text{and} \quad \mu_4 = 3\exp\left\{2\alpha + \frac{2\sigma^2}{(1-\phi^2)}\right\}.$$

The simplified expressions found here are consistent with the ones reported by Ghysels *et al.* (1996), and hence the proposed model $SVM_{\rho\mu}$ (2) is a generalization of the classical model SVM_0 in (1). Next we investigate the conditional (on $\mathcal{F}_{t-1}$) dependence between the current returns and past, current and future volatility.

2.2 Lead-lag correlations

In this section, we wish to estimate three quantities: (1) dependence between the current returns and current volatility, $corr(r_t, h_t|\mathcal{F}_{t-1})$, (2) the potential influence of current returns on future volatility, $corr(r_t, h_{t+k}|\mathcal{F}_{t-1})$, and (c) the influence

of past volatility on current returns, $corr(r_t, h_{t-k}|\mathcal{F}_{t-1})$. Though empirical estimation of such quantities is not uncommon, e.g., in Bollerslev *et al.* (2006), our aim is to find closed analytical expression for these descriptive measures under $SVM_{\rho\mu}$ specification.

Since $var(r_t|\mathcal{F}_{t-1})$ is given by (7) and $var(h_t|\mathcal{F}_{t-1}) = \sigma^2/(1-\phi^2)$, we only need to find the expressions for the conditional covariances. First, we recall that under the proposed model, the conditional means are $E(r_t|\mathcal{F}_{t-1}) = 0$ and $E(h_t|\mathcal{F}_{t-1}) = \alpha$. Now, if we assume that $corr(r_t, h_t|\mathcal{F}_{t-1}) = \sigma_{rh}$, then

$$cov(r_t, h_{t+1}) = E[r_t(h_t - \alpha)] = E[r_t(\phi(h_t - \alpha) + \sigma\eta_{t+1})] = \phi E[r_t(h_t - \alpha)] = \phi\sigma_{rh},$$

which further implies that $cov(r_t, h_{t+k}) = \phi^k\sigma_{rh}$ for $k \geq 1$. By applying the key mathematical techniques (i.e., properties of expectation, normal mgf and the expansion of $h_t = \alpha + \sigma\sum_{j=1}^{\infty}\eta_{t-j}\phi^j$) used in proving results of Section 2.1, one can easily show that

$$\begin{aligned} \sigma_{rh} = cov(r_t, h_t|\mathcal{F}_{t-1}) &= \rho\sigma \exp\left\{\frac{\alpha}{2} + \frac{\sigma^2}{8(1-\phi^2)}\right\} \times \left\{1 + \frac{\sigma^2}{4(1-\phi^2)}\right\}, \\ cov(r_t, h_{t-k}|\mathcal{F}_{t-1}) &= \sigma_{rh} \cdot \phi^k \cdot \left[\frac{\sigma^2}{4(1-\phi^2)}\right] \Big/ \left[1 + \frac{\sigma^2}{4(1-\phi^2)}\right]. \end{aligned}$$

Clearly, both the lead ($cov(r_t, h_{t+k})$) and lag ($cov(r_t, h_{t-k})$) covariances are smaller than the contemporaneous covariance $cov(r_t, h_t|\mathcal{F}_{t-1})$. The contemporaneous correlation can be interpreted as almost instantaneous *feedback effect* of volatility change on returns, whereas the impact of return change on future volatility is termed as *leverage effect*. Bekaert & Wu (2000) found that volatility feedback effect is stronger than leverage effect. The closed form expressions we have derived above provide a theoretical proof of the mentioned findings under $SVM_{\rho\mu}$ specification. Moreover, Bollerslev *et al.* (2006) have empirically observed that the lag-correlation with lag h is smaller than lead correlation with lead h which we have established theoretically. Further note that all these covariances and hence correlations vanish if $\rho = corr(\epsilon_t, \eta_t) = 0$. Next, we compare the goodness of fit of the three stochastic volatility models, SVM_0 (classical - with zero correlation), SVM_ρ (with correlation ρ) and $SVM_{\rho\mu}$ (mean-corrected with correlation ρ), for a real data on returns.

3. Example: S&P 500 NYSE

In this paper, we compare the performance of the three models ($SVM_0, SVM_\rho, SVM_{\rho\mu}$) on the index returns of Standard and Poor 500 index (S&P500) obtained from New York Stock Exchange during April 01, 2002 – March 30, 2006. We selected this period to avoid extreme behaviour during "2000 – 2002 dot-com bubble" and "2008 Lehman Brothers' crash". Figure 1 displays the time-plot of the returns of 1008 trading days (less than the total number of calendar days).

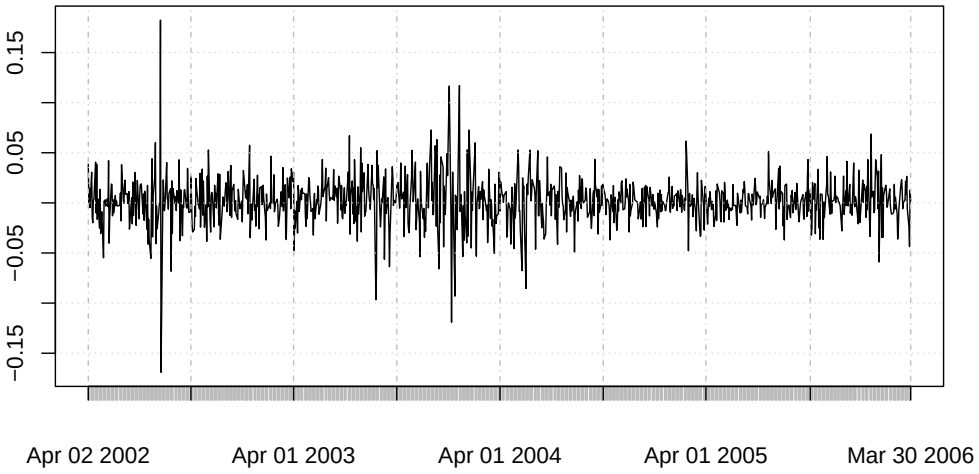


Figure 1. Time plot of S&P500 returns during April 01, 2002 – March 30, 2006

From Figure 1 one can infer that the volatility is relatively high during September 2003 and June 2004, whereas during October 2004 to April 2005, the volatility is relatively lower than usual. A few descriptive statistics of the observed returns are as follows:

$$\begin{array}{ll} \text{mean} = 0.0014, & \text{variance} = 0.0005, \\ \text{skewness} = 0.0329, & \text{kurtosis} = 10.9813. \end{array}$$

We follow Meyer & Yu (2000), and use the same Markov Chain Monte Carlo (MCMC) algorithm implemented in Just Another Gibbs Sampler (JAGS) for fitting the classical model SVM_0. For fitting the other two models, SVM_ρ and $SVM_{\rho\mu}$, we slightly modify the JAGS code to include the $corr(\epsilon_t, \eta_t) = \rho$ and μ (derived in Theorem 1). For implementing SVM_0 in JAGS, the hierarchical model structure is characterized by

$$r_t \mid (h_t, h_{t-1}, \ldots h_1, h_0; \alpha, \phi, \sigma) \sim N\left(0, \exp\{h_t\}\right),$$

$$\text{and} \quad h_t \mid (h_{t-1}, \ldots h_1, h_0; \alpha, \phi, \sigma) \sim N\left(\alpha + \phi(h_{t-1} - \alpha), \sigma^2\right).$$

For SVM_ρ, the mean and variance of the conditional distribution of r_t changes to

$$r_t \mid (h_t, \ldots, h_0; \alpha, \phi, \sigma) \sim N\left(\frac{\rho\, e^{h_t/2}}{\sigma}(h_t - \alpha - \phi(h_{t-1} - \alpha)), e^{h_t}(1 - \rho^2)\right),$$

and the conditional distribution of h_t remains the same. Similarly, the implementation of the mean-corrected model $SVM_{\rho\mu}$ is characterized by updating the mean and variance of the conditional distribution of r_t to

$$r_t \mid (h_t, \ldots, h_0; \alpha, \phi, \sigma) \sim N\left(\mu + \frac{\rho\, e^{h_t/2}}{\sigma}(h_t - \alpha - \phi(h_{t-1} - \alpha)), e^{h_t}(1 - \rho^2)\right).$$

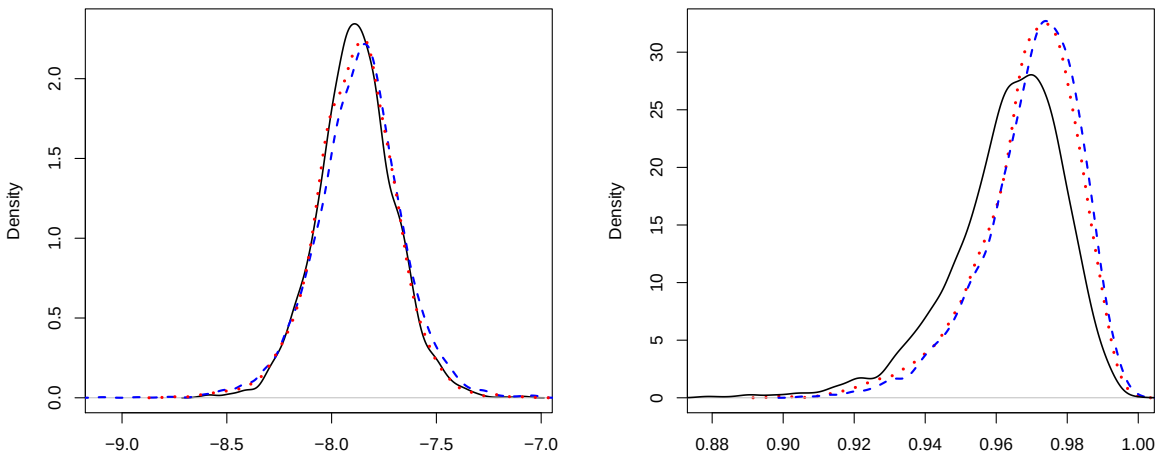
The parameters of interest are $(\alpha, \phi, \rho, \sigma) = \Theta$ (say). We use the same prior (including the hyperparameters) for α, ϕ and σ as in Meyer & Yu (2000), and a non-informative $Unif(-1, 1)$ prior for the correlation parameter. The posterior of Θ and $\mathcal{H} = \{h_t, h_{t-1}, ...\}$ given the data $\{r_t, r_{t-1}, ...\}$ is obtained via JAGS. We set the total length of chains to be 180,000, out of which 30,000 was the burn-in, and from the remaining 150,000 posterior realizations (with the thinning of every 50^{th} realization) were used (i.e., 3000 realizations in total) to obtain the plug-in estimates of the parameters. The thinning process facilitates a safeguard against the chain dependency in the sampling process. Figure 2 shows the density plots of the posterior distribution of Θ for the three models, SVM_0, SVM_ρ and $SVM_{\rho\mu}$. We have not included the traceplots, as all parameters converge nicely and the plots do not reveal anything extra. The plug-in estimates of the parameters are obtained via posterior mean and variance (summarized in Table 1).

Table 1. Plug-in estimators of $\Theta = (\alpha, \phi, \sigma, \rho)$ for the three models. The numbers in parentheses show the standard deviation of the posterior realizations.

Parameter	SVM_0	SVM_ρ	$SVM_{\rho\mu}$
α	-7.88	-7.87	-7.88
	(0.1837)	(0.2077)	(0.192)
ϕ	0.96	0.97	0.96
	(0.016)	(0.014)	(0.014)
σ	0.2	0.177	0.18
	(0.04)	(0.034)	(0.038)
ρ		0.1185	0.105
		(0.1362)	(0.1278)

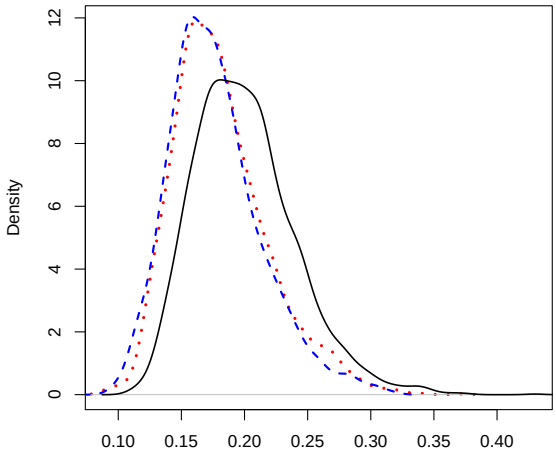
Table 1 shows that the posterior estimates of the parameters in SVM_0, SVM_ρ and $SVM_{\rho\mu}$ are similar. Further, the near-unity estimate of ϕ indicates presence of strong volatility clustering. The estimate of the correlation parameter ρ is small yet positive, which is similar to the findings of French *et al.* (1987) and Campbell & Hentschel (1992). This may be taken as an indication of no significant effect of current return on future volatility.

Figure 2 shows that the posterior distributions of the parameters for SVM_0, SVM_ρ and $SVM_{\rho\mu}$ are different in their kurtosis. A general pattern that can be noticed is that posterior distributions of the parameters under SVM_0 are more leptokurtic compared to their counter parts under other two models except for α. Importantly, too strong volatility clustering is more

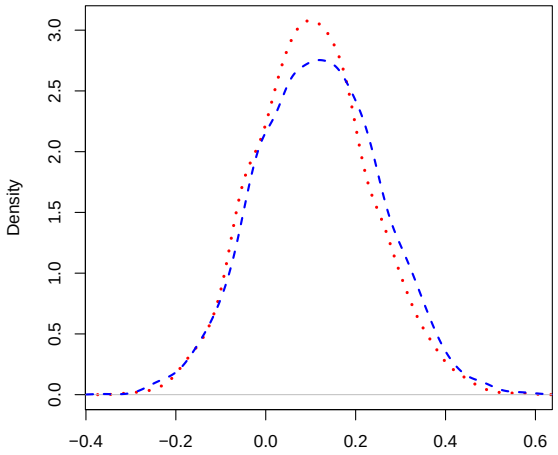


(a) α (expected volatility)

(b) ϕ (stationarity parameter)



(c) σ (variability in volatility)



(d) $\rho = corr(\epsilon_t, \eta_t)$

Figure 2. Posterior distribution of Θ for the three models. The black solid curves represent SVM_0, blue dashed curves are from SVM_ρ model, and the red dotted curves are obtained from the proposed model $SVM_{\rho\mu}$.

probable under SVM_ρ and $SVM_{\rho\mu}$ compared to SVM_0. In case of variance of volatility, posterior distribution under SVM_0 indicates higher values compared to the other two models. Comparing the posterior distributions of ρ under SVM_ρ and $SVM_{\rho\mu}$, the former shows higher probability of being positive valued relative to the latter.

As per Figure 2(d), ρ is very small (close to zero), and thus, it is expected that the proposed model would not provide significant additional strength in modelling the returns data.

We now compare the three models using the descriptive measures (mean, variance, skewness and kurtosis), three lead-lag correlations, mean deviance over the posterior distribution, and the mean square prediction error (MSPE): $\sum_{t=1}^{T} \hat{r}_t^2/T$. The deviance function, suggested by Dempster (1974), is

$$D(\Theta) = -2 \log f(r \mid \Theta, \mathcal{H}) + 2 \log g(r),$$

where $f(r \mid \Theta, \mathcal{H})$ is the likelihood for a given realization of Θ and $\mathcal{H}$, and $g(r)$ is the normalizing constant. Table 2 presents the plug-in values of these "goodness of fit" measures for the three models.

Since $\rho \approx 0.1$ (very small), the estimated mean is also small $\mu = -4.05 \cdot 10^{-6}$. Thus all three models would behave very

Table 2. Goodness of fit measures for the true data and the three models.

GOF measure	True data	SVM_0	SVM_ρ	$SVM_{\rho\mu}$
Mean	0.0014	0	–	-4.05×10^-6
Variance	0.0005	0.0005	0.0005	0.0005
Skewness	0.0329	0	0.0856	0.0769
Kurtosis	10.981	5.196	5.105	5.076
$corr(r_t, h_t)$			0.0305	0.0276
$corr(r_t, h_{t-10})$				0.0053
Deviance		-5019	-5033	-5043
MSPE ($\times 10^{-7}$)		0.178	12.94	9.298

similarly (which is reflected in the estimated moments under the three models). Surprisingly plug-in estimates of kurtosis obtained from all three models under-estimates the kurtosis measured from the data. Deviance values indicate that $SVM_{\rho\mu}$ provides a slightly better fit compared to the other two models. On the other hand, MSPE values indicate that the basic SVM provides better prediction among the three models. Though the numerical results presented through the S& P 500 NYSE example do not provide sufficient evidence for $SVM_{\rho\mu}$ giving additional information than SVM_0, it certainly is the generalization of SVM_0 and an example with large $\rho = corr(\epsilon_t, \eta_t)$ might have given more convincing evidence.

4. Concluding Remarks

In this paper, we have proposed a mean-correction for the SVM with correlation between ϵ_t and η_t. This mean-correction step enables the conditional expected return to be zero, which is a necessary condition for a good SVM (i.e., a model that adhere to the EMH). We have also found the closed form analytical expressions for the higher moments of returns and lead-lag correlation between the return and volatility.

From S&P500 example, we see that most of the empirical observations on statistical properties of returns are reflected through all the three models. However, $SVM_{\rho\mu}$ gives a slightly better fit to the data (in terms of average deviance) compared to the classical model SVM_0 as well as SVM_ρ. A close look at this research endeavour generates several interesting and challenging research problems.

First, the estimated error correlation ρ turns out to be positive despite the fact that return and its volatility move in opposite directions (Nelson (1991)). Glosten *et al.* (1993) attributed this discrepancy due to mis-specification in the underlying SVM, which is caused by not accounting for the size discrepancy in volatility change due to up or down movement of price. The authors have shown that if the size discrepancy is accounted for then ρ becomes negative. This result demonstrates that ρ alone can not explain the asymmetric response of return to its volatility sufficiently. As we have pointed out in the introduction that this size discrepancy can be interpreted as different conditional variances (or volatility) for positive and negative returns, which leads to skewed return distribution instead of a Gaussian one, a new model can be developed by extending $SVM_{\rho\mu}$ in the line of Abanto-Valle *et al.* (2010).

Second, the observed kurtosis from the data is not completely explained by the model based estimates of kurtosis. Indeed, the significant difference between empirical kurtosis and the model based estimates again suggests non-normality of the return error distribution. The problem can be tackled in two ways- (1) introducing jumps in returns or (2) allowing the return error to be heavy-tailed (*e.g.* Student's t). Notice, adding a jump to the return only explains transient changes (as seen on 8^{th} & 9^{th} August, 2002) and does not cause the return distribution to change permanently whereas jump in both return and volatility explains persistent effects of extreme values (*e.g.* September, 2003 – June, 2004). $SVM_{\rho\mu}$ can further be generalized by including jumps in return and volatility (Eraker *et al.* (2003)) following the 1^{st} line of argument and using skew Student's-t distributions following the 2^{nd} line of argument (Abanto-Valle *et al.* (2015)).

Although continuous time stochastic volatility has been studied extensively in the literature, discrete-time SVM brings out new features such as leverage effect and feedback effect which occurs due to lagged reaction between return and its volatility. In this paper we have established that the empirically observed pattern of leverage effect and lagged correlations (Bollerslev *et al.* (2006)) are explained by $SVM_{\rho\mu}$. In particular, we have shown that the correlation between current return and future volatility is maximum in magnitude at lead 0 (or contemporaneously) and the future leverage effects disappear exponentially with the lead time. Indeed, strong volatility clustering effect indicates more persistent leverage effect. It may also be noted that the existing practice of assuming $h_{t+1} = \alpha + \phi(h_t - \alpha) + \sigma\eta_t$ (instead of h_t) and $corr(\epsilon_t, \eta_t) = \rho$ for a correct SVM specification would not support the empirical observation on contemporaneous correlation.

Mean-correction to the contemporaneously correlated SVM has another very important application, which is extensively researched in continuous-time scenario, but barely investigated in the discrete-time domain. An *arbitrage* opportunity is

created in a market if one can borrow an amount P_0 to purchase a stock at time 0, and sell it at time T at a price P_T, making no loss with probability 1 and getting a profit with positive probability. An efficient market would never want an arbitrage opportunity to be created. It can be shown that, under mild assumptions, a necessary and sufficient condition for the existence of no arbitrage is equivalent to the existence of a risk-neutral density. In fact, in continuous time, P_t is a martingale under this density ((Williams, 2006, Ch. 5)). A discrete time analogue of the no-arbitrage condition is to say that $r_t = log\left(\frac{P_t}{P_{t-1}}\right)$ is a martingale difference or equivalently has zero mean. Arbitrage free option prices are derived from SVMs using the risk-neutral density, and may not be unique (e.g., Hull & White (1987)). It can also be shown that if r_t is not a martingale difference process then arbitrage opportunities can be created. As a result, arbitrage free option prices will not exist. In this paper we have proposed mean-correction of the SVM by adding an appropriate non-zero drift term which facilitates no-arbitrage and hence the existence of a risk neutral density. Thus the proposed model opens up the way to compute option prices under the type of SVMs proposed by (Jacquier *et al.* (2004)). This relationship with option pricing requires a detailed investigation which we believe can lead to important results.

Acknowledgements

We would like to thank the reviewers for helpful comments that led to significant improvement in the article.

References

Abanto-Valle, C. A., Bandyopadhyay, D., Lachos, V. H., & Enriquez, I. 2010. Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. *Computational Statistics and Data Analysis*, **54**(12), 2883–2898. http://dx.doi.org/10.1016/j.csda.2009.06.011.

Abanto-Valle, CA, Lachos, VH, & Dey, Dipak K. 2015. Bayesian estimation of a skew-student-t stochastic volatility model. *Methodology and Computing in Applied Probability*, **17**(3), 721–738. http://dx.doi.org/10.1007/s11009-013-9389-9.

Bekaert, G., & Wu, G. 2000. Asymmetric volatility and risk in equity markets. *The review of financial studies*, **13**, 1–42.

Berg, A., Meyer, R., & Yu, J. 2004. Deviance Information Criterion for Comparing Stochastic Volatility Models. *Journal of Business and Economic Statistics*, **22**(1), 107–120.

Black, F. 1976. The pricing of commodity contracts. *Journal of Financial Economics*, **3**, 167–179.

Bollerslev, T, Litvinova, J, & Tauchen, G. 2006. Leverage and feedback effects in high frequency data. *Journal of Financial Econometrics*, **4**, 353–384. http://dx.doi.org/10.1093/jjfinec/nbj014.

Campbell, John Y., & Hentschel, Ludger. 1992. No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, **31**(3), 281 – 318. http://dx.doi.org/10.1016/0304-405X(92)90037-X.

Dempster, A. P. 1974. The direct use of likelihood for significance testing. *Proceedings of conference on foundational questions in statistical inference*, **University of Aarhus**, 335–352.

Engle, Robert F. 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, **50**(4), 987–1007. http://dx.doi.org/10.2307/1912773.

Eraker, B., Johanners, M., & Polson, N. G. 2003. The impact of jumps in volatility and returns. *Journal of Finance*, **53(3**, 1269–1300. http://dx.doi.org/10.1111/1540-6261.00566.

French, R. Kenneth, Schwert, G. William, & Stambaugh, F. Robert. 1987. Expected Stock Returns and Volatility. *Journal of Financial Economics*, **19**, 3–29. http://dx.doi.org/10.1016/0304-405X(87)90026-2.

Ghysels, Eric, Harvey, Andrew C., & Renault, Eric. 1996. Stochastic volatility. *Pages 119 – 191 of:* Maddala, G. S., & Rao, C. R. (eds), *Statistical Methods in Finance*. Handbook of Statistics, vol. 14. New York: Elsevier.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. 1993. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, **48**(5), 1779–1801. http://dx.doi.org/10.1111/j.1540-6261.1993.tb05128.x.

Harvey, C.R., & Siddique, A. 1999. Autoregressive conditional skewness. *Journal of Finance and Quantitative Analysis*, **34**, 465–487. http://dx.doi.org/10.2307/2676230.

Hull, John, & White, Alan. 1987. The pricing of options on assets with stochastic volatilities. *The journal of finance*, **42**(2), 281–300.

Jacquier, E., Polson, N. G., & Rossi, P. E. 2004. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, **122**(1), 185–212. http://dx.doi.org/10.1016/j.jeconom.2003.09.001.

Meyer, R., & Yu, J. 2000. BUGS for a Bayesian analysis of stochastic volatility models. *Econometrics Journal*, **3**(2), 198–215. http://dx.doi.org/10.1111/1368-423X.00046.

Nelson, B. D. 1991. Conditional heteroskedasticity in asset pricing: a new approach. *Econometrica*, **59**, 347–370.

Shephard, Neil (ed). 2005. *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.

Taylor, S. J. 1982. Financial returns modeled by the product of two stochastic processes a study of daily sugar prices 1961-79. *Pages 203–226 of:* Anderson, O. D. (ed), *Time Series Analysis: Theory and Practice 1*. Amsterdam: North-Holland.

Wang, J. J. J., Chan, J. S. K., & Choy, S. T. B. 2011. Stochastic volatility models with leverage and heavy-tailed distributions: A Bayesian approach using scale mixtures. *Computational Statistics and Data Analysis*, **55**(1), 852–862. http://dx.doi.org/10.1016/j.csda.2010.07.008.

Williams, Ruth J. 2006. *Introduction to the Mathematics of Finance*. Vol. 72. American Mathematical Society.

Yu, J. 2005. On leverage in a stochastic volatility model. *Journal of Econometrics*, **127**(2), 165–178. http://dx.doi.org/10.1016/j.jeconom.2004.08.002.

International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

The Effect of Age on Road Traffic Fatality Index in Ghana

Christian A. Hesse[1], John B. Ofosu[1], Samuel K. Darkwah[1]

[1] Department of Mathematical Sciences, Faculty of Informatics and Mathematical Sciences, Methodist University College, Ghana

Correspondence: C. A. Hesse, Department of Mathematical Sciences, Methodist University College, P. O. Box DC 940, Dansoman – Accra, Ghana. E-mail: akrongh@yahoo.com

Received: April 22, 2016 Accepted: May 30, 2016 Online Published: June 22, 2016

doi:10.5539/ijsp.v5n4p111 URL: http://dx.doi.org/10.5539/ijsp.v5n4p111

Abstract

In this paper, data on road traffic casualties by age groups, from 2009 to 2013, will be used. Using published road traffic casualty statistics from the National Road Safety Commission of Ghana, a 2×8 contingency table is used to determine whether road traffic casualty and age group are independent. A one factor analysis of variance tests shall be used to conduct a comparative analysis of the rate of road traffic fatalities per 100 casualties across the various age groups in Ghana. A multiple comparison test, using the Fisher least significance difference (LSD) method, shall be conducted to determine which pairs of age groups are significantly different.

The study will show that road traffic casualty is not independent of age group. The analysis of variance will show that there are significant differences in road traffic fatality indices (fatality per 100 casualties) among various age groups in Ghana. The risks of dying in a road traffic accident among children under 6 years and older population who are over 65 years are both significantly higher than those of other age groups. This points to the fact that, although smaller number of children under 6 years and older population who are over 65 years die in road traffic accidents each year, more and more people as a proportion of the recorded number of casualties, are being killed through road traffic accidents among these two categories of age groups. Thus, the probability of being killed in a fatal road traffic accident is significantly high in each of these two age groups.

Keywords: Contingency, analysis of variance, Traffic fatalities, casualties, injuries

1. Introduction

The European Economic Commission (EEC) and the World Health Organization (1979) have recommended a definition for a road traffic accident fatality. This includes only deaths which occur within 30 days following a road traffic accident while road traffic casualties refer to road traffic accident victims injured or killed within 30 days of the accidents. A number of countries have not yet adopted this definition. For example, in some countries, a road traffic fatality is recorded only if the victim dies at the site or is dead upon arrival at a hospital. In order to make comparison of accident statistics between countries reasonable, figures obtained from countries which have not adopted the 30-day fatality definition, should be properly adjusted. No adjustment is required for figures from countries such as Ghana, U.S.A and Great Britain, which have adopted the standard fatality definition.

Casualties of road traffic accidents in Ghana by age groups, from 2009 – 2013, are given in Table 1. Unlike many fatal diseases, road traffic accidents kill people from all age groups, including young and middle-aged people in their active years. A cumulative total of 10 555 fatalities is recorded over the 5-year period. The highest fatalities during the period were in the 26 – 35 year old. Table 1 also shows that the active age group, 16 – 45 years, was the most vulnerable in road traffic fatalities, representing 63.2% of the total fatalities in the 5-year period.

According to the National Road Safety Commission (NRSC) of Ghana 2013 annual report, one key national Road Traffic Fatality index (F. I.) required for characterization and comparison of the extent and risk of road traffic fatality is fatalities per 100 casualties (see Hesse and Ofosu, 2015). In Table 1, the distribution of the rate of road traffic fatalities per 100 accidents by age groups from 2009 – 2013 are also computed.

Table 1. Age distributions of fatalities and injuries from road traffic accidents from 2010 to 2013

		Casualties											
		Persons Killed						Persons Injured					
		2013	2012	2011	2010	2009	Total	2013	2012	2011	2010	2009	Total
Age groups (Years)	0 – 5	97	113	126	136	130	602	214	241	276	389	401	1521
	6 – 15	148	170	212	217	250	997	529	789	846	962	1112	4238
	16 – 25	315	335	365	269	388	1672	2172	2509	2723	3110	3245	13759
	26 – 35	531	661	658	577	609	3036	3871	4458	5070	5297	5861	24557
	36 – 45	359	441	400	379	383	1962	2162	2753	3009	2932	3138	13994
	46 – 55	188	236	209	184	222	1039	1001	1334	1374	1399	1512	6620
	56 – 65	149	159	126	129	141	704	472	621	493	563	618	2767
	Over 65	111	125	103	95	109	543	190	296	229	266	246	1227
	Total	**1898**	**2240**	**2199**	**1986**	**2232**	**10555**	**10611**	**13001**	**14020**	**14918**	**16133**	**3994**

It can be seen, from Table 2, that the F. I. increased from 24.5 to 31.2 among children under 6 years from year 2009 to 2013, whilst that of the 'over 65' age groups increased marginally from 30.7 to 36.9 over the same period. In very simple terms, these changes imply that the chance of at least one casualty dying as a result of road traffic accident has increased over the period. It can be observed that, over the 5 year period, the 'over 65' continues to be the age group with the highest national fatality rate. For instance, in 2013, about 37% of all road traffic casualties who were over 65 years lost their lives while 31% of casualties who were 5 years old or less died as a result of road traffic accidents.

Table 2. Rate of fatalities per 100 casualties (fatality indices)

		0 – 5	6 – 15	16 – 25	26 – 35	36 – 45	46 – 55	56 – 65	Over 65
		1	2	3	4	5	6	7	8
2013	**1**	31.2	21.9	12.7	12.1	14.2	15.8	24.0	36.9
2012	**2**	31.9	17.7	11.8	12.9	13.8	15.0	20.4	29.7
2011	**3**	31.3	20.0	11.8	11.5	11.7	13.2	20.4	31.0
2010	**4**	25.9	18.4	8.0	9.8	11.4	11.6	18.6	26.3
2009	**5**	24.5	18.4	10.7	9.4	10.9	12.8	18.6	30.7
mean		**29.0**	**19.3**	**11.0**	**11.1**	**12.4**	**13.7**	**20.4**	**30.9**

The number of road traffic fatality victims in Ghana can be classified according to two criteria, of a set of entities, namely casualty and age group. Casualty has 2 levels (i.e. fatalities and injured) while age group has 8 levels. These form a 2×8 contingency table as shown in Table 3.

Table 3. Road traffic accidents victims from 2010 to 2013

		Age Group								
		0 – 5	6 – 15	16 – 25	26 – 35	36 – 45	46 – 55	56 – 65	Over 65	Total
casualty	**Fatalities**	602	997	1672	3036	1962	1039	704	543	10555
	Injured	1521	4238	5759	24557	13994	38551	2767	1227	92614
	Total	2123	5235	7431	27593	15956	39590	3471	1770	103169

In this study, we wish to know whether road traffic casualty and age group are independent. If they are independent, then we would expect to find the same proportion of fatalities across various age groups. We also propose the use of the

completely randomized single factor experiment to determine if there are significant differences in road traffic fatality index rates among the various age groups.

2. Method

Table 4 shows an $r \times c$ contingency table where O_{ij} is the observed frequency for level i of the first method of classification and level j of the second method of classification, where $R_i = \sum_{j=1}^{c} O_{ij}$ is the *marginal total* for row i and $C_j = \sum_{i=1}^{r} O_{ij}$ is the *marginal total* for column j. Note that $\sum_{i=1}^{r} R_i = \sum_{j=1}^{c} C_j = n$, where n is the total sample size.

Table 4. An $r \times c$ contingency table

		Columns				
		1	2	…	c	Total
Rows	1	O_{11}	O_{12}	…	O_{1c}	R_1
	2	O_{21}	O_{22}	…	O_{2c}	R_2
	.	.	.	.	.	.
	.	.	.	.	.	.
	.	.	.	.	.	.
	r	O_{r1}	O_{r2}	…	O_{rc}	R_r
Total		C_1	C_2	…	C_c	n

We are interested in testing the null hypothesis

H_0: the row-and-column methods of classification are independent

against the alternative hypothesis

H_1: the row-and-column methods of classification are not independent.

The test statistic is given by (see Cramér (1946) and Birch (1964)).

$$H = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{1}$$

were E_{ij} is the expected cell frequency for the $(ij)^{\text{th}}$ cell. It can be shown that, if H_0 is true, then:

$$E_{ij} = \frac{R_i \times C_j}{n} = \frac{(\text{column total}) \times (\text{row total})}{\text{grand total}}. \tag{2}$$

It can also be shown that, for large n, the statistic H has an approximate chi-square distribution with $(r - 1)(c - 1)$ degrees of freedom if H_0 is true (see Ofosu and Hesse (2011)). Therefore, we would reject the hypothesis of independence if the observed value of the test statistic H is greater than the critical value $\chi^2_{\alpha,(r-1)(c-1)}$, where α is the size of the test. An extensive treatment of the chi-square distribution can be found in the book by Lancaster (1969).

If we reject the null hypothesis, we conclude that there is some interaction between the two criteria of classification.

3. Results

3.1 Test of Independence

The null and the alternative hypotheses are:

H_0: Casualty is independent of age group.

H_1: Casualty is not independent of age group.

We first find the expected cell frequencies. These are calculated by using Equation (2). Table 5 shows the expected cell frequencies of Table 3 using Equation (2). For example, $E_{11} = \frac{10555 \times 2123}{103169} = 217.200.$

Table 5. Expected cell frequencies of Table 3

		Age Group								
		0 – 5	**6 – 15**	**16 – 25**	**26 – 35**	**36 – 45**	**46 – 55**	**56 – 65**	**Over 65**	**Total**
casualty	**Fatalities**	217.200	535.582	760.250	2822.981	1632.424	4050.368	355.111	181.085	10555
	Injured	1905.800	4699.418	6670.750	24770.019	14323.576	35539.632	3115.889	1588.915	92614
	Total	2123	5235	7431	27593	15956	39590	3471	1770	103169

Note that the expected frequencies in any row or column add up to the appropriate marginal total. The test statistic is

$$H = \sum_{i=1}^{2} \sum_{j=1}^{8} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}.$$

When H_0 is true, H has the chi-square distribution with 7 [i.e. (2 – 1)(8 – 1)] degrees of freedom. We reject H_0 at 0.05 level of significance when the computed value of the test statistic is greater than $\chi^2_{0.05,7} = 14.07.$ Substituting both the observed values in Table 3 and their corresponding expected values in Table 5 into $\chi_{ij} = \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, we obtain the cells in Table 6.

Table 6. Calculations of the observed test statistic

	1	2	3	4	5	6	7	8	Total
χ_{1j}	245.97	213.55	497.18	14.95	55.36	8727.95	172.90	241.22	10169.08
χ_{2j}	97.35	50.24	144.35	1.85	7.76	235.23	43.99	106.75	687.52
Total	343.32	263.79	641.53	16.79	63.12	8963.18	216.89	347.97	10856.59

Thus, the observed value of the test statistic is

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{8} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 10856.59.$$

Since $10856.59 > 14.07$, we reject the hypothesis of independence and conclude that casualty is not independent of age group.

3.2 Completely Randomized Single Factor Experiment

Table 2 is a typical data of a single-factor experiment with 8 levels (age groups) of the factor, where the factor is the effect of age on F. I. We wish to determine if there are significant differences between the average F. I. across the 8 age

groups. In Table 2, let y_{ij} represent the i^{th} observation taken under the j^{th} age group and

$$y_{.j} = \sum_{i=1}^{5} y_{ij}, \quad \bar{y}_{.j} = y_{.j}/21, \quad (j = 1,\ 2,\ ...,\ 8), \quad y_{..} = \sum_{j=1}^{8}\sum_{i=1}^{5} y_{ij}, \quad \bar{y}_{..} = y_{..}/40.$$

Let μ_j represent the true mean of the j^{th} age group and ε_{ij} the experimental error. The model for the completely randomized single factor experiment is

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad (j = 1,\ 2,\ ...,\ 8,\ i = 1,\ 2,\ ...,\ 5). \tag{3}$$

The one-way analysis of variance model assumes that the observations are normally and independently distributed with the same variance for each region or factor level (see Ofosu et al. (2014)).

3.2.1 Validation of Normality and Homogeneity of Variances Assumptions

We check the normality assumption, using the Shapiro-Wilk *W* test. The null hypothesis is

H_0: observations under each region are normally distributed

against the alternative hypothesis

H_1: observations under each region are not from a normally distributed population

The value of the Shapiro-Wilk *W* test statistic for each of the eight age groups is given in Table 7 below.

Table 7. Observed values of the *W* test statistic

Test Statistic	**0 – 5**	**6 – 15**	**16 – 25**	**26 – 35**	**36 – 45**	**46 – 55**	**56 – 65**	**Over 65**
W_{o}	0.802	0.883	0.864	0.930	0.871	0.951	0.836	0.925

H_0 is rejected at the 5% level of significance if the computed value of *W* is less than 0.762, the tabulated 5% point of the distribution of the Shapiro-Wilk test statistic. For each of the 8 age groups, we fail to reject H_0 and therefore conclude that there is not enough evidence of non-normality of these samples.

Levene's test (Levene 1960) is used to test if 8 samples have equal variances. We wish to test

H_0: $\sigma_1^2 = \sigma_2^2 = ... = \sigma_8^2$ against

H_0: $\sigma_i^2 \neq \sigma_j^2$ for at least one pair (*i*, *j*).

In Table 2, let y_{ij} represent the i^{th} observation taken under the j^{th} age group and

$$y_{.j} = \sum_{i=1}^{5} y_{ij}, \quad \bar{y}_{.j} = y_{.j}/5, \quad (j = 1,\ 2,\ ...,\ 8), \quad y_{..} = \sum_{j=1}^{5}\sum_{i=1}^{8} y_{ij}, \quad \bar{y}_{..} = y_{..}/40.$$

t = number of treatments = 8

n_i = number of observations from treatment (region) i

$N = n_1 + n_2 + ... + n_9$ = overall size of combined samples = 40,

$D_{ij} = \left| y_{ij} - \bar{y}_i \right|$ = absolute deviation of observation j from treatment i mean

$\bar{D}_i =$ average of the n_i absolute deviations from treatment i

$\bar{D} =$ average of all N absolute deviations

The Levene's test statistic is given by

$$F_{Levene} = \frac{\sum_{i=1}^{8} n_i\left(\bar{D}_i - \bar{D}\right)^2}{7} \Big/ \frac{\sum_{i=1}^{8} n_i\left(D_{ij} - \bar{D}_i\right)^2}{32}. \tag{4}$$

When H_0 is true, F_{Levene} has the *F*-distribution with 4 and 40 degrees of freedom. H_0 is rejected at significance level 0.05 when the observed value of F_{Levene} is greater than $F_{0.05,\,7,\,32} = 2.33$. Since the observed *F*-ratio, 1.332, is less than the critical *F*-value, 2.33, we fail to reject the null hypothesis at the 0.05 level of significance and conclude that there are no significant differences among the ten variances.

3.2.2 One-way Analysis of Variance

Since the normality and homogeneity of variances assumptions are validated, we can use the one-way analysis of variance to determine if the fatality indices across age groups vary significantly. We wish to test the hypothesis

H_0: The mean fatality indices are the same across the 8 categories of age groups,

against the alternative hypothesis

H_1: The mean fatality indices are not the same for at least 2 of age groups.

The total corrected sum of squares is given by

$$SST = \sum_{j=1}^{8}\sum_{i=1}^{5} y_{ij}^2 - \frac{y_{..}^2}{40} = 2374.360. \quad (5)$$

The sum of squares among treatments is

$$SSA = \sum_{j=1}^{8} \frac{y_{.j}^2}{5} - \frac{y_{..}^2}{40} = 2193.712. \quad (6)$$

The within treatment sum of squares, SSW, can be obtained from the equation

$$SSW = SST - SSA = 180.648. \quad (7)$$

The analysis of variance results, based on the data in Table 2, are summarized in Table 8 below.

Table 8. Analysis of variance table

Source of variation	Sum of squares	Degrees of freedom	Mean square	F-ratio
Among treatments	2193.712	7	313.387	
Within treatments	180.648	32	5.645	55.513
Total	2374.360	39		

The test statistic is

$$F = \frac{\text{among treatments mean square}}{\text{within treatments mean square}}.$$

When H_0 is true, F has the F-distribution with 7 and 32 degrees of freedom. We reject H_0 at significance level 0.05 when the observed value of F is greater than $F_{0.05,\,7,\,32} = 2.33$. From Table 8, the computed value of F is 55.513. Since the observed F-ratio, 55.513, is greater than the critical F-value, 2.33, we reject the null hypothesis at the 0.05 level of significance and conclude that there are significant differences among the fatality indices across the 8 age groups.

4. Discussion

4.1 Multiple Comparison Method

Since the analysis of variance indicates that the null hypothesis should be rejected, it means that there are differences among the 8 treatment means. But as to which of the means are significantly different, the analysis does not specify. Obviously, in such a situation, we need a different method for comparing individual treatment means. One such methods is the multiple comparison test.

Over the years, several methods for making multiple comparison tests have been suggested. Duncan (1951, 1952, 1955) has contributed a considerable amount of research to the subject of multiple comparisons. Other multiple comparison methods in use are those proposed by Tukey (1949, 1953), Newman (1939), Keuls (1952), and Scheff é (1953, 1959). The advantages and disadvantages of the various multiple comparison methods are discussed by Bancroft (1968), O'Neill and Wetherill (1971), Daniel and Coogler (1975), Winer (1971) and Ofosu et al. (2014). Daniel (1980) has prepared a bibliography on multiple comparison procedures.

The oldest multiple comparison method, and perhaps the most widely used, is the least significant difference method of Fisher, who first discussed it in the 1935 edition of his book "The design of experiments" (see Ofosu et al. (2014)). To use this method, we first calculate the least significant difference, (LSD), for the given data. This is given by

$$LSD = t_{\frac{1}{2}\alpha,\ N-k}\sqrt{\frac{2MSW}{n}}, \tag{8}$$

where the level of significance $\alpha = 0.05$, $N = 40$, $n = 5$, $k = 8$ and $MSW = 5.645$. This gives $LSD = 3.068$.

The observed difference between each pair of means is compared to the *LSD*. If the observed numerical difference is greater than 3.068, then the road traffic fatality indices of the two age groups are significantly different. The values of the observed numerical differences between pairs of means of the 8 age groups are given in Table 9. Pairs of age groups with fatality indices not significantly different are highlighted in Table 9.

Table 10. Observed numerical differences between pair of means of road user classes

		0 – 5	6 – 15	16 – 25	26 – 35	36 – 45	46 – 55	56 – 65	Over 65
		29.0	19.3	11.0	11.1	12.4	13.7	20.4	30.9
0 – 5	29.0		9.7	18.0	17.9	16.6	15.3	8.6	**1.9**
6 – 15	19.3			8.3	8.2	6.9	5.6	**1.1**	11.6
16 – 25	11.0				**0.1**	**1.4**	**2.7**	9.4	19.9
26 – 35	11.1					**1.3**	**2.6**	9.3	19.8
36 – 45	12.4						**1.3**	8.0	18.5
46 – 55	13.7							6.7	17.2
56 – 65	20.4								10.5
Over 65	30.9								

For example, from Table 9, it can be seen that, the observed numerical difference between the mean fatality indices for the age groups '0 – 5' and '26 – 35' is 17.9. Since 17.9 is greater than 3.068, it follows that there is a significant difference between the two age groups with respect to F. I. It is obvious that the road traffic fatality index for '0 – 5' age group is significantly higher than that of other age groups except for 'Over 65'. This means that, the risk of dying in a road traffic accident among '0 – 5' and 'Over 65' are both significantly higher than those of other age groups, recording an average rate of 29.0 and 30.9 deaths per 100 casualties, respectively.

5. Conclusion

We've shown that road traffic casualty level depends on age group of victims involved using a 2 × 8 contingency analysis.

The analysis of variance revealed that there are significant differences in road traffic fatality indices (fatality per 100 casualties) among various age groups in Ghana. The risks of dying in a road traffic accident among children under 6 years and older population who are over 65 years are both significantly higher than those of other age groups. This points to the fact that, although smaller number of children under 6 years and older population who are over 65 years die in road traffic accidents each year, more and more people as a proportion of the recorded number of casualties, are being killed through road traffic accidents among these two categories of age groups. Thus, the probability of being killed in a fatal road traffic accident is significantly high in each of these two age groups. This may be due to higher fragility of children and older population of road users.

These findings are consistent with a related study by Loughran et al. (2007), in which they reported that older drivers are more than twice as likely as middle-aged drivers to cause an accident. The research revealed that drivers and passengers riding in cars driven by older drivers are nearly seven times likelier to die in an auto accident than are passengers and drivers riding in cars driven by middle-aged drivers. This statistic suggests that older individuals are much likelier than middle-aged individuals to die in a car accident. Given these trends, the research suggests that public policy should focus more on improving the safety of automobile travel for older drivers and less on screening out older drivers whose driving abilities have deteriorated unacceptably.

References

Bancroft, T. A. (1968). Topics in intermediate statistical methods. *Iowa State university Press, Ames, Iowa*.

Birch, M. W. (1964). A new proof of the Fisher-Pearson theorem. *Annals of Mathematical Statistics*, *35*, 817 – 824. http://dx.doi.org/10.1214/aoms/1177703581

Cramér, H. (1946). Mathematical Methods of Statistics. Princeton University Press, Princeton, N. J.

Daniel, W. W., & Coegler, C. E. (1975), Beyond Analysis of variance: A comparison of some multiple comparison procedures. *Physical Therapy*, *55*, 114 – 150.

Daniel, W. W. (1980). Multiple comparison procedures. A selected Bibiography. Vance Bibliographies, Monticello, III, 1980.

Duncan, D. B. (1951). A significance test for difference between ranked treatments in an Analysis of variance. *Verginal Journal of science*, *2*, 171 – 189.

Duncan, D. B. (1952). On the properties of the multiple comparison tests. *Verginal Journal of Science, 3*, 50 – 67.

Duncan, D. B. (1955). Multiple Range and Multiple F-Tests. *Biometrics*, *11*, 1 – 42. http://dx.doi.org/10.2307/3001478

Hesse, C. A., & Ofosu, J. B. (2015). The Effect of Road Traffic Fatality Rate on Road Users in Ghana. *Research Journal of Mathematics and Statistics*, *7*(4), 53-59. http://dx.doi.org/10.19026/rjms.7.2207

Keuls, M. (1952). The use of the studentized range in connection with an Analysis of variance. *Euphytica, 1*, 112 – 122. http://dx.doi.org/10.1007/BF01908269

Lancaster, H. O. (1969). The chi-squared distribution. John Wiley and Sons, New York.

Levene, H. (1960). In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin et al. eds., Stanford University Press, pp. 278-292.

Loughran, D. S., Seabury, S. A., & Zakaras, L (2007). *Regulating Older Drivers: Are New Policies Needed?* Santa Monica, Calif.: RAND Corporation, OP-189-ICJ.

Newman, D. (1939). The distribution of the range in samples from a normal population in terms of an independent estimate of standard deviation. *Biometrica*, *31*, 20 – 30. http://dx.doi.org/10.1093/biomet/31.1-2.20

Ofosu, J. B., & Hesse, C. A. (2011). Elementary Statistical Methods. *EPP Books Services, Accra*.

Ofosu, J. B., Hesse, C. A., & Otchere, F. (2014). Intermediate Statistical Methods. *EPP Books Services, Accra*.

O'Neill, R., & Wetherill, G. B. (1971). The Present State of Multiple Comparison Methods. *Journal of the Royal Statistical Society*, B, 33, 218 – 241.

Scheffé, H. (1959). Analysis of variance. *John Wiley and Sons. New York*.

Scheffé, H. (1953). A method for judging all contrasts in the Scheffé, H. (1950) Analysis of variance. *Biometrika, 40*, 87 – 104.

Tukey, J. W. (1949). Comparing Individual Mean in the Analysis of variance. *Biometrics, 5*, 99 – 114. http://dx.doi.org/10.2307/3001913

Tukey, J. W. (1953). The problem of multiple comparisons. *Ditto*, Princeton University.

Winer, B. J. (1971). Statistical Principles in Experimental Design. 2nd edition, McGraw-Hill, New York.

World Health Organization. (1997). Road Traffic Accidents Statistics. Copenhagen, Regional Office for Europe.

International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Computation of the Survival Probability of Brownian Motion with Drift When the Absorbing Boundary is a Piecewise Affine or Piecewise Exponential Function of Time

Tristan Guillaume [1]

[1] Université de Cergy-Pontoise, Laboratoire Thema, Cergy, France

Correspondence: Tristan Guillaume, Université de Cergy-Pontoise, Laboratoire Thema, 33 boulevard du port, F-95011 Cergy-Pontoise Cedex, France. Tel: 33-6-12-22-45-88. E-mail: tristan.guillaume@u-cergy.fr

Received: May 12, 2016 Accepted: May 31, 2016 Online Published: June 27, 2016

doi:10.5539/ijsp.v5n4p119 URL: http://dx.doi.org/10.5539/ijsp.v5n4p119

Abstract

A closed form formula is provided for the probability, in a closed time interval, that an arithmetic Brownian motion remains under or above a sequence of three affine, one-sided boundaries (equivalently, for the probability that a geometric Brownian motion remains under or above a sequence of three exponential, one-sided boundaries). The numerical evaluation of this formula can be done instantly and with the accuracy required for all practical purposes. The method followed can be extended to sequences of absorbing boundaries of higher dimension. It is also applied to sequences of two-sided boundaries.

Keywords: boundary crossing probability; survival probability; probability of absorption; first passage time; hitting time; Brownian motion; affine boundary; exponential boundary

1. Introduction

The question of the crossing of a non-constant boundary by a diffusion process is of central importance in many mathematical sciences. As mentioned in Wang and Pötzelberger (2007), it arises in biology, economics, engineering reliability, epidemiology, finance, genetics, seismology and sequential statistical analysis. The probability that a diffusion process will remain under or above some critical threshold over a given time interval can be referred to as a survival probability or probability of non-absorption. The vast majority of the research articles published on this topic either focus on numerical algorithms for general classes of processes or boundaries, usually involving recursive multidimensional quadrature, or they seek to obtain approximate solutions, typically substituting the initial boundary with another one for which computations are easier and then deriving a bound for the error entailed by using the approximating boundary. Much attention has also been paid to asymptotic estimates. However, known closed form results are scarce. By closed form results, we mean fully explicit formulae involving functions whose numerical evaluation can be carried out with the accuracy and the efficiency required for all practical purposes, in contrast to approximate analytical solutions that are quickly computed but inaccurate, and to numerical algorithms that can only produce the required standard of precision through heavy computational burden. The most classical of these closed form results is the so-called Bachelier-Levy formula (Levy, 1948), which provides the first-passage time density of Brownian motion to a linear boundary. This result is extended to a two-sided linear boundary by Doob (1949), but only in infinite time. The generalisation to a closed time interval is given by Anderson (1960), who is also able to integrate the density. The first passage time density of Brownian motion to a quadratic boundary is obtained independently by Salminen (1988) and Groeneboom (1989), while Novikov et al. (1999) manage to derive the hitting time density of Brownian motion to a square root boundary, but the numerical evaluation is quite involved in both cases, requiring infinite series of roots of combinations of Airy functions or confluent hypergeometric functions. By integrating these first passage time densities, the corresponding survival probabilities can be derived, though the integration is not actually performed by the mentioned authors and is far from trivial. Scheike (1992) provides a closed form solution for the survival probability of Brownian motion in infinite time when the boundary consists of two successive linear functions of time but cannot explicitly compute the corresponding integral in finite time. There are also a few closed form results for a Brownian motion (Daniels, 1996; Wang and Pötzelberger, 2007), an Ornstein-Uhlenbeck process (Choi and Nam, 2003; Wang and Pötzelberger, 2007) and a growth process (Wang and Pötzelberger, 2007), that involve very specific forms of the boundary and thus have limited use in practice, although they are quite valuable to test numerical algorithms.

This paper provides new results for the survival probability of Brownian motion. The problem raised by Scheike (1992)

is reformulated, extended and analytically solved. The extension with regard to the existing literature can be summarized as follows :

- cumulative distribution functions are provided, i.e. the integration of the first passage time densities is performed

- results are provided for generalised Brownian motion (whether arithmetic or geometric Brownian motion), i.e. the underlying stochastic dynamics include drift and volatility coefficients

- sequences of up to three general affine boundaries (in the case of arithmetic Brownian motion) or exponential boundaries (in the case of geometric Brownian motion) are handled

- sequences of two-sided piecewise affine or exponential boundaries are also tackled, under the assumption that the growth rate of the boundary is identical on the downside and on the upside, i.e. the upper and the lower sides of the boundary are parallel curves

Only distributions in finite time are considered, as they are the ones used in practice in the various mathematical sciences. The choice of affine and exponential boundaries is because they allow to model a reasonably large variety of time-dependent conditions for real life problems, while preserving analytical tractability. There are potentially many applications, for example in the valuation and risk management of various path dependent financial options or insurance contracts as well as in structural models of credit risk (see, e.g., Jeanblanc et al., 2009).

Section 2 of this article states a closed form formula for the survival probability of an arithmetic or a geometric Brownian under or above a sequence of three different one-sided affine or exponential boundaries over a finite time interval and provides a few numerical results, then outlines a proof omitting cumbersome computations, and finally discusses generalization to higher-dimensional boundaries. Section 3 of this article states a closed form formula for the survival probability of an arithmetic or a geometric Brownian motion under and above a sequence of two different two-sided, parallel, affine or exponential boundaries over a finite time interval, provides a few numerical results and outlines the proof.

2. Survival Probability of an Arithmetic or a Geometric Brownian Motion under or Above a Sequence of One-sided Affine or Exponential Boundaries over a Finite Time Interval

2.1 Definitions

Let μ be a real constant, σ be a positive real constant, and $\{B(t), t \geq 0\}$ be a standard Brownian motion defined on a probability space with measure $\mathbb{P}$. Let $\{X_1(t), t \geq 0\}$ be an arithmetic Brownian motion driven, under $\mathbb{P}$, by :

$$dX_1(t) = \mu\, dt + \sigma\, dB(t) \tag{2.1}$$

Let $\{X_2(t), t \geq 0\}$ be a geometric Brownian motion driven, under $\mathbb{P}$, by :

$$dX_2(t) = \mu X_2(t) dt + \sigma X_2(t) dB(t) \tag{2.2}$$

A finite time interval $[0, T]$ is considered and divided into a partition Π of n subintervals $[t_0 = 0, t_1], [t_1, t_2]$, ... $[t_{n-1}, t_n = T]$, which are not necessarily of equal length, with $t_n \geq t_{n-1} \geq ... \geq t_1 \geq t_0$. Let $\mathbb{I}$ denote the indicator function. For a given $n \in \mathbb{N}$, two piecewise affine absorbing boundaries $g_1(t)$ and $g_2(t)$ are defined as follows :

$$g_1(t) = \sum_{i=1}^{n} \left(a_i + b_i (t - t_{i-1})\right) \mathbb{I}_{[t_{i-1}, t_i]}(t), a_i \in \mathbb{R}, b_i \in \mathbb{R}, \; i \in \{1, 2, ..., n\} \tag{2.3}$$

$$g_2(t) = \sum_{i=1}^{n} (a_i + b_i t)\, \mathbb{I}_{[t_{i-1}, t_i]}(t) \quad , \quad a_i \in \mathbb{R} \quad , \quad b_i \in \mathbb{R} \, , \quad i \in \{1,2,...,n\} \tag{2.4}$$

The difference between $g_1(t)$ and $g_2(t)$ is that $g_1(t)$ is time-homogeneous.

Similarly, we have the two following piecewise exponential boundaries :

$$h_1(t) = \sum_{i=1}^{n} X(0) \exp\left(a_i + b_i (t - t_{i-1})\right) \mathbb{I}_{[t_{i-1}, t_i]}(t), \quad a_i \in \mathbb{R} \quad , \quad b_i \in \mathbb{R} \, , \quad i \in \{1,2,...,n\} \tag{2.5}$$

$$h_2(t) = \sum_{i=1}^{n} X(0) \exp(a_i + b_i t)\, \mathbb{I}_{[t_{i-1}, t_i]}(t), \quad a_i \in \mathbb{R} \quad , \quad b_i \in \mathbb{R} \, , \quad i \in \{1,2,...,n\} \tag{2.6}$$

Consider the cumulative distribution function of a sequence of n maxima or n minima and n endpoints in Π in the two following cases :

- the absorbing boundary is defined either by $g_1(t)$ or $g_2(t)$ and the process under consideration is X_1

- the absorbing boundary is defined either by $h_1(t)$ or $h_2(t)$ and the process under consideration is X_2

Such a function is often referred to as a survival probability. As shown by Wang and Pötzelberger (1997), its value can be approximated by a Monte Carlo simulation scheme drawing on the Markovian nature of X_1 and X_2 in the following manner : the endpoint values of X_1 and X_2 in each time subinterval $[t_{i-1}, t_i]$ are randomly drawn at each performed simulation; if the relevant conditions at each t_i are met, then a cumulative variable records the product of the conditional probabilities that the boundary has not been crossed in each (t_{i-1}, t_i), which admit simple analytical formulae (Siegmund, 1986). This is obviously much more efficient and accurate than discretizing the whole path of the process at each run. For $n > 1$, the survival probability under consideration does not admit any known closed form formula. Although it does not seem possible to come up with an explicit and compact formula for any $n \in \mathbb{N}$, one can actually solve the problem analytically in "moderate" dimension. In this paper, the case $n = 3$ is tackled. More specifically, let $\mathrm{P}_{[...]}(\mu, \sigma, a_1, a_2, a_3, b_1, b_2, b_3, k_1, k_2, k_3, t_1, t_2, t_3)$ be defined as one of the following eight cumulative distribution functions :

$$\mathrm{P}_{[\mathrm{AU1}]}(\mu, \sigma, a_1, a_2, a_3, b_1, b_2, b_3, k_1, k_2, k_3, t_1, t_2, t_3),\ a_1 \in \mathbb{R}_+, (a_2, a_3, b_1, b_2, b_3, k_1, k_2, k_3) \in \mathbb{R}^8 \tag{2.7}$$

$$= \mathbb{P}\begin{pmatrix} (X_1(t) < a_1 + b_1 t, \forall 0 \le t \le t_1) \cap X_1(t_1) < k_1 \cap (X_1(t) < a_2 + b_2 (t - t_1), \forall t_1 \le t \le t_2) \\ \cap X_1(t_2) < k_2 \cap (X_1(t) < a_3 + b_3 (t - t_2), \forall t_2 \le t \le t_3) \cap X_1(t_3) < k_3 \end{pmatrix}$$

$$\mathrm{P}_{[\mathrm{AU2}]}(\mu, \sigma, a_1, a_2, a_3, b_1, b_2, b_3, k_1, k_2, k_3, t_1, t_2, t_3),\ a_1 \in \mathbb{R}_+, (a_2, a_3, b_1, b_2, b_3, k_1, k_2, k_3) \in \mathbb{R}^8 \tag{2.8}$$

$$= \mathbb{P}\begin{pmatrix} (X_1(t) < a_1 + b_1 t, \forall 0 \le t \le t_1) \cap X_1(t_1) < k_1 \cap (X_1(t) < a_2 + b_2 t, \forall t_1 \le t \le t_2) \\ \cap X_1(t_2) < k_2 \cap (X_1(t) < a_3 + b_3 t, \forall t_2 \le t \le t_3) \cap X_1(t_3) < k_3 \end{pmatrix}$$

$$\mathrm{P}_{[\mathrm{GU1}]}\left(\mu,\sigma,a_1,a_2,a_3,b_1,b_2,b_3,k_1,k_2,k_3,t_1,t_2,t_3\right),\left(a_1,k_1,k_2,k_3\right)\in\mathbb{R}_+^4,\left(a_2,a_3,b_1,b_2,b_3\right)\in\mathbb{R}^5 \tag{2.9}$$

$$=\mathbb{P}\begin{pmatrix}\left(X_2\left(t\right)<X_2\left(0\right)\exp\left(a_1+b_1t\right),\forall 0\le t\le t_1\right)\cap X_2\left(t_1\right)<k_1\\ \cap\left(X_2\left(t\right)<X_2\left(0\right)\exp\left(a_2+b_2\left(t-t_1\right)\right),\forall t_1\le t\le t_2\right)\\ \cap X_2\left(t_2\right)<k_2\cap\left(X_2\left(t\right)<X_2\left(0\right)\exp\left(a_3+b_3\left(t-t_2\right)\right),\forall t_2\le t\le t_3\right)\cap X_2\left(t_3\right)<k_3\end{pmatrix}$$

$$\mathrm{P}_{[\mathrm{GU2}]}\left(\mu,\sigma,a_1,a_2,a_3,b_1,b_2,b_3,k_1,k_2,k_3,t_1,t_2,t_3\right),\left(a_1,k_1,k_2,k_3\right)\in\mathbb{R}_+^4,\left(a_2,a_3,b_1,b_2,b_3\right)\in\mathbb{R}^5 \tag{2.10}$$

$$=\mathbb{P}\begin{pmatrix}\left(X_2\left(t\right)<X_2\left(0\right)\exp\left(a_1+b_1t\right),\forall 0\le t\le t_1\right)\cap X_2\left(t_1\right)<k_1\\ \cap\left(X_2\left(t\right)<X_2\left(0\right)\exp\left(a_2+b_2t\right),\forall t_1\le t\le t_2\right)\\ \cap X_2\left(t_2\right)<k_2\cap\left(X_2\left(t\right)<X_2\left(0\right)\exp\left(a_3+b_3t\right),\forall t_2\le t\le t_3\right)\cap X_2\left(t_3\right)<k_3\end{pmatrix}$$

$$\mathrm{P}_{[\mathrm{AL1}]}\left(\mu,\sigma,a_1,a_2,a_3,b_1,b_2,b_3,k_1,k_2,k_3,t_1,t_2,t_3\right),\ a_1\in\mathbb{R}_-,\left(a_2,a_3,b_1,b_2,b_3,k_1,k_2,k_3\right)\in\mathbb{R}^8 \tag{2.11}$$

$$=\mathbb{P}\begin{pmatrix}\left(X_1\left(t\right)>a_1+b_1t,\forall 0\le t\le t_1\right)\cap X_1\left(t_1\right)>k_1\cap\left(X_1\left(t\right)>a_2+b_2\left(t-t_1\right),\forall t_1\le t\le t_2\right)\\ \cap X_1\left(t_2\right)>k_2\cap\left(X_1\left(t\right)>a_3+b_3\left(t-t_2\right),\forall t_2\le t\le t_3\right)\cap X_1\left(t_3\right)>k_3\end{pmatrix}$$

$$\mathrm{P}_{[\mathrm{AL2}]}\left(\mu,\sigma,a_1,a_2,a_3,b_1,b_2,b_3,k_1,k_2,k_3,t_1,t_2,t_3\right),\ a_1\in\mathbb{R}_-,\left(a_2,a_3,b_1,b_2,b_3,k_1,k_2,k_3\right)\in\mathbb{R}^8 \tag{2.12}$$

$$=\mathbb{P}\begin{pmatrix}\left(X_1\left(t\right)>a_1+b_1t,\forall 0\le t\le t_1\right)\cap X_1\left(t_1\right)>k_1\cap\left(X_1\left(t\right)>a_2+b_2t,\forall t_1\le t\le t_2\right)\\ \cap X_1\left(t_2\right)>k_2\cap\left(X_1\left(t\right)>a_3+b_3t,\forall t_2\le t\le t_3\right)\cap X_1\left(t_3\right)>k_3\end{pmatrix}$$

$$\mathrm{P}_{[\mathrm{GL1}]}\left(\mu,\sigma,a_1,a_2,a_3,b_1,b_2,b_3,k_1,k_2,k_3,t_1,t_2,t_3\right),\ a_1\in\mathbb{R}_-,\left(k_1,k_2,k_3\right)\in\mathbb{R}_+^3,\left(a_2,a_3,b_1,b_2,b_3\right)\in\mathbb{R}^5 \tag{2.13}$$

$$=\mathbb{P}\begin{pmatrix}\left(X_2\left(t\right)>X_2\left(0\right)\exp\left(a_1+b_1t\right),\forall 0\le t\le t_1\right)\cap X_2\left(t_1\right)>k_1\\ \cap\left(X_2\left(t\right)>X_2\left(0\right)\exp\left(a_2+b_2\left(t-t_1\right)\right),\forall t_1\le t\le t_2\right)\\ \cap X_2\left(t_2\right)>k_2\cap\left(X_2\left(t\right)>X_2\left(0\right)\exp\left(a_3+b_3\left(t-t_2\right)\right),\forall t_2\le t\le t_3\right)\cap X_2\left(t_3\right)>k_3\end{pmatrix}$$

$$\mathrm{P}_{[\mathrm{GL2}]}\left(\mu,\sigma,a_1,a_2,a_3,b_1,b_2,b_3,k_1,k_2,k_3,t_1,t_2,t_3\right),\ a_1\in\mathbb{R}_-,\left(k_1,k_2,k_3\right)\in\mathbb{R}_+^3,\left(a_2,a_3,b_1,b_2,b_3\right)\in\mathbb{R}^5 \tag{2.14}$$

$$=\mathbb{P}\begin{pmatrix}\left(X_2\left(t\right)>X_2\left(0\right)\exp\left(a_1+b_1t\right),\forall 0\le t\le t_1\right)\cap X_2\left(t_1\right)>k_1\\ \cap\left(X_2\left(t\right)>X_2\left(0\right)\exp\left(a_2+b_2t\right),\forall t_1\le t\le t_2\right)\\ \cap X_2\left(t_2\right)>k_2\cap\left(X_2\left(t\right)>X_2\left(0\right)\exp\left(a_3+b_3t\right),\forall t_2\le t\le t_3\right)\cap X_2\left(t_3\right)>k_3\end{pmatrix}$$

In other words, taking $n=3$,

- $\mathrm{P}_{[\mathrm{AU1}]}$ is the probability that an arithmetic Brownian motion will remain under the piecewise affine time-homogeneous boundary $g_1\left(t\right)$ defined by (2.3) and under the successive endpoints k_1,k_2,k_3

- $\mathrm{P}_{[\mathrm{AU2}]}$ is the probability that an arithmetic Brownian motion will remain under the piecewise affine time-inhomogeneous boundary $g_2\left(t\right)$ defined by (2.4) and under the successive endpoints k_1,k_2,k_3

- $\mathrm{P}_{[\mathrm{GU1}]}$ is the probability that a geometric Brownian motion will remain under the piecewise exponential time-homogeneous boundary $h_1(t)$ defined by (2.5) and under the successive endpoints k_1, k_2, k_3

- $\mathrm{P}_{[\mathrm{GU2}]}$ is the probability that a geometric Brownian motion will remain under the piecewise exponential time-inhomogeneous boundary $h_2(t)$ defined by (2.6) and under the successive endpoints k_1, k_2, k_3

- $\mathrm{P}_{[\mathrm{AL1}]}$ is the probability that an arithmetic Brownian motion will remain above the piecewise affine boundary $g_1(t)$ defined by (2.3) and above the successive endpoints k_1, k_2, k_3

- $\mathrm{P}_{[\mathrm{AL2}]}$ is the probability that an arithmetic Brownian motion will remain above the piecewise affine boundary $g_2(t)$ defined by (2.4) and above the successive endpoints k_1, k_2, k_3

- $\mathrm{P}_{[\mathrm{GL1}]}$ is the probability that a geometric Brownian motion will remain above the piecewise exponential boundary $h_1(t)$ defined by (2.5) and above the successive endpoints k_1, k_2, k_3

- $\mathrm{P}_{[\mathrm{GL2}]}$ is the probability that a geometric Brownian motion will remain above the piecewise exponential boundary $h_2(t)$ defined by (2.6) and above the successive endpoints k_1, k_2, k_3

2.2 Statement of Formula 1

Formula 1 *Let* $\mathrm{P}_{[\ldots]}(\mu,\sigma,a_1,a_2,a_3,b_1,b_2,b_3,k_1,k_2,k_3,t_1,t_2,t_3)$ *be defined as in Subsection 2.1. Then,*

$$\mathrm{P}_{[\ldots]}(\mu,\sigma,a_1,a_2,a_3,b_1,b_2,b_3,k_1,k_2,k_3,t_1,t_2,t_3) \tag{2.15}$$

$$= \Phi_3\left[\begin{matrix}\theta\dfrac{z_1-b_1t_1-\mu_1t_1}{\sigma\sqrt{t_1}}, \theta\dfrac{z_2-b_2t_2-\mu_1t_1-\mu_2(t_2-t_1)}{\sigma\sqrt{t_2}}, \theta\dfrac{z_3-b_3t_3-\mu_1t_1-\mu_2(t_2-t_1)-\mu_3(t_3-t_2)}{\sigma\sqrt{t_3}}; \\ \sqrt{t_1\,/\,t_2}, \sqrt{t_2\,/\,t_3}\end{matrix}\right]$$

$$-\exp\left(\frac{\lambda_1}{\sigma^2}\right)\Phi_3\left[\begin{matrix}\theta\dfrac{z_1-b_1t_1-2a_1-\mu_1t_1}{\sigma\sqrt{t_1}}, \theta\dfrac{z_2-b_2t_2-2a_1-\mu_1t_1-\mu_2(t_2-t_1)}{\sigma\sqrt{t_2}}, \\ \theta\dfrac{z_3-b_3t_3-2a_1-\mu_1t_1-\mu_2(t_2-t_1)-\mu_3(t_3-t_2)}{\sigma\sqrt{t_3}}; \sqrt{t_1\,/\,t_2}, \sqrt{t_2\,/\,t_3}\end{matrix}\right]$$

$$-\exp\left(\frac{\lambda_2}{\sigma^2}\right)\Phi_3\left[\begin{matrix}\theta\dfrac{z_1-b_1t_1-\mu_1t_1+2\mu_2t_1}{\sigma\sqrt{t_1}}, \theta\dfrac{z_2-b_2t_2-2\alpha_2+\mu_1t_1-\mu_2(t_1+t_2)}{\sigma\sqrt{t_2}}, \\ \theta\dfrac{z_3-b_3t_3-2\alpha_2+\mu_1t_1-\mu_2(t_1+t_2)-\mu_3(t_3-t_2)}{\sigma\sqrt{t_3}}; -\sqrt{t_1\,/\,t_2}, \sqrt{t_2\,/\,t_3}\end{matrix}\right]$$

$$+\exp\left(\frac{\lambda_3}{\sigma^2}\right)\Phi_3\left[\begin{matrix}\theta\dfrac{z_1-b_1t_1-2a_1-\mu_1t_1+2\mu_2t_1}{\sigma\sqrt{t_1}},\theta\dfrac{z_2-b_2t_2-2\alpha_2+2a_1+\mu_1t_1-\mu_2\left(t_1+t_2\right)}{\sigma\sqrt{t_2}}\\ \theta\dfrac{z_3-b_3t_3-2\alpha_2+2a_1+\mu_1t_1-\mu_2\left(t_1+t_2\right)-\mu_3\left(t_3-t_2\right)}{\sigma\sqrt{t_3}};-\sqrt{t_1\ /\ t_2},\sqrt{t_2\ /\ t_3}\end{matrix}\right]$$

$$-\exp\left(\frac{\lambda_4}{\sigma^2}\right)\Phi_3\left[\begin{matrix}\theta\dfrac{z_1-b_1t_1-\mu_1t_1+2\mu_3t_1}{\sigma\sqrt{t_1}},\theta\dfrac{z_2-b_2t_2-\mu_1t_1-\mu_2\left(t_2-t_1\right)+2\mu_3t_2}{\sigma\sqrt{t_2}},\\ \theta\dfrac{z_3-b_3t_3-2\alpha_3+\mu_1t_1+\mu_2\left(t_2-t_1\right)-\mu_3\left(t_2+t_3\right)}{\sigma\sqrt{t_3}};\sqrt{t_1\ /\ t_2},-\sqrt{t_2\ /\ t_3}\end{matrix}\right]$$

$$+\exp\left(\frac{\lambda_5}{\sigma^2}\right)\Phi_3\left[\begin{matrix}\theta\dfrac{z_1-b_1t_1-2a_1+2\mu_3t_1-\mu_1t_1}{\sigma\sqrt{t_1}},\theta\dfrac{z_2-b_2t_2-2a_1-\mu_1t_1-\mu_2\left(t_2-t_1\right)+2\mu_3t_2}{\sigma\sqrt{t_2}},\\ \theta\dfrac{z_3-b_3t_3-2\alpha_3+2a_1+\mu_1t_1+\mu_2\left(t_2-t_1\right)-\mu_3\left(t_2+t_3\right)}{\sigma\sqrt{t_3}};\sqrt{t_1\ /\ t_2},-\sqrt{t_2\ /\ t_3}\end{matrix}\right]$$

$$+\exp\left(\frac{\lambda_6}{\sigma^2}\right)\Phi_3\left[\begin{matrix}\theta\dfrac{z_1-b_1t_1-2\mu_3t_1+2\mu_2t_1-\mu_1t_1}{\sigma\sqrt{t_1}},\theta\dfrac{z_2-b_2t_2-2\alpha_2+\mu_1t_1-\mu_2\left(t_1+t_2\right)+2\mu_3t_2}{\sigma\sqrt{t_2}},\\ \theta\dfrac{z_3-b_3t_3-2\alpha_3+2\alpha_2-\mu_1t_1+\mu_2\left(t_1+t_2\right)-\mu_3\left(t_2+t_3\right)}{\sigma\sqrt{t_3}};-\sqrt{t_1\ /\ t_2},-\sqrt{t_2\ /\ t_3}\end{matrix}\right]$$

$$-\exp\left(\frac{\lambda_7}{\sigma^2}\right)\Phi_3\left[\begin{matrix}\theta\dfrac{z_1-b_1t_1-2a_1-2\mu_3t_1+2\mu_2t_1-\mu_1t_1}{\sigma\sqrt{t_1}},\\ \theta\dfrac{z_2-b_2t_2-2\alpha_2+2a_1+\mu_1t_1-\mu_2\left(t_1+t_2\right)+2\mu_3t_2}{\sigma\sqrt{t_2}},\\ \theta\dfrac{z_3-b_3t_3-2\alpha_3+2\alpha_2-2a_1-\mu_1t_1+\mu_2\left(t_1+t_2\right)-\mu_3\left(t_2+t_3\right)}{\sigma\sqrt{t_3}};\\ -\sqrt{t_1\ /\ t_2},-\sqrt{t_2\ /\ t_3}\end{matrix}\right]$$

where the function Φ_n *is a convolution of gaussian densities defined, for any* $n\in\mathbb{N}$, *by :*

$$\Phi_n\left[x_1,...,x_n;\rho_1,...,\rho_{n-1}\right]$$

$$=\int_{D^n}\frac{\exp\left(-\dfrac{y_1^2}{2}-\displaystyle\sum_{i=1}^{n-1}\dfrac{\left(y_{i+1}-\rho_iy_i\right)^2}{2\left(1-\rho_i^2\right)}\right)}{\left(2\pi\right)^{n/2}\displaystyle\prod_{i=1}^{n-1}\sqrt{1-\rho_i^2}}dy_n...dy_1 \tag{2.16}$$

$$D^n=\left]-\infty,x_1\right]\times\left]-\infty,x_2\right]...\times\left]-\infty,x_n\right],\ \ x_i\in\mathbb{R},\ \ \rho_i\in\left]-1,1\right[\ \ ,\ \ i\in\left\{1,...,n\right\}$$

The α_i *terms,* $i\in\left\{2,3\right\}$, *in (2.15) are given by :*

$$\alpha_2=a_2\left(\mathbb{I}_{\left\{\mathrm{P}_{[...]}=\mathrm{P}_{[\mathrm{AU2}]}\right\}}+\mathbb{I}_{\left\{\mathrm{P}_{[...]}=\mathrm{P}_{[\mathrm{GU2}]}\right\}}+\mathbb{I}_{\left\{\mathrm{P}_{[...]}=\mathrm{P}_{[\mathrm{AL2}]}\right\}}+\mathbb{I}_{\left\{\mathrm{P}_{[...]}=\mathrm{P}_{[\mathrm{GL2}]}\right\}}\right)$$

$$+\left(a_2 - b_2 t_1\right)\left(\mathbb{I}_{\left\{P_{[...]}=P_{[AU1]}\right\}} + \mathbb{I}_{\left\{P_{[...]}=P_{[GU1]}\right\}} + \mathbb{I}_{\left\{P_{[...]}=P_{[AL1]}\right\}} + \mathbb{I}_{\left\{P_{[...]}=P_{[GL1]}\right\}}\right)$$

$$\alpha_3 = a_3\left(\mathbb{I}_{\left\{P_{[...]}=P_{[AU2]}\right\}} + \mathbb{I}_{\left\{P_{[...]}=P_{[GU2]}\right\}} + \mathbb{I}_{\left\{P_{[...]}=P_{[AL2]}\right\}} + \mathbb{I}_{\left\{P_{[...]}=P_{[GL2]}\right\}}\right)$$

$$+\left(a_3 - b_3 t_2\right)\left(\mathbb{I}_{\left\{P_{[...]}=P_{[AU1]}\right\}} + \mathbb{I}_{\left\{P_{[...]}=P_{[GU1]}\right\}} + \mathbb{I}_{\left\{P_{[...]}=P_{[AL1]}\right\}} + \mathbb{I}_{\left\{P_{[...]}=P_{[GL1]}\right\}}\right)$$

The λ_i *terms,* $i \in \left\{1,2,3,4,5,6,7\right\}$ *, in (2.15) are given by :*

$$\lambda_1 = 2\mu_1 a_1$$

$$\lambda_2 = 2\mu_2\alpha_2 - 2\mu_1\mu_2 t_1 + 2\mu_2^2 t_1$$

$$\lambda_3 = 2\mu_1 a_1 + 2\mu_2\alpha_2 - 4\mu_2 a_1 - 2\mu_1\mu_2 t_1 + 2\mu_2^2 t_1$$

$$\lambda_4 = 2\mu_3\alpha_3 + 2\mu_3^2 t_2 - 2\mu_1\mu_3 t_1 - 2\mu_2\mu_3\left(t_2 - t_1\right)$$

$$\lambda_5 = 2\mu_3\alpha_3 + 2\mu_1 a_1 - 4\mu_3 a_1 + 2\mu_3^2 t_2 - 2\mu_1\mu_3 t_1 - 2\mu_2\mu_3\left(t_2 - t_1\right)$$

$$\lambda_6 = 2\mu_3\alpha_3 + 2\mu_2\alpha_2 - 4\mu_3\alpha_2 + 2\left(\mu_3 - \mu_2\right)^2 t_1 + 2\mu_1\left(\mu_3 - \mu_2\right)t_1 + 2\mu_3^2\left(t_2 - t_1\right) - 2\mu_2\mu_3\left(t_2 - t_1\right)$$

$$\lambda_7 = 2\mu_1 a_1 + 2\mu_2\alpha_2 - 4\mu_3\alpha_2 + 2\mu_3\alpha_3 + 2\left(\mu_3 - \mu_2\right)^2 t_1 + 2\mu_3^2\left(t_2 - t_1\right)$$

$$-2\mu_2\mu_3\left(t_2 - t_1\right) + 2\left(\mu_3 - \mu_2\right)\left(2a_1 + \mu_1 t_1\right)$$

The z_i *terms,* $i \in \left\{1,2,3\right\}$ *, in (2.15) are given by :*

$$z_1 = \min\left(a_1 + b_1 t_1, k_1, a_2\right)\mathbb{I}_{\left\{P_{[...]}=P_{[AU1]}\right\}} + \min\left(a_1 + b_1 t_1, k_1, a_2 + b_2 t_1\right)\mathbb{I}_{\left\{P_{[...]}=P_{[AU2]}\right\}}$$

$$+\min\left(a_1 + b_1 t_1, \ln\left(k_1 \,/\, X_2\left(0\right)\right), a_2\right)\mathbb{I}_{\left\{P_{[...]}=P_{[GU1]}\right\}} + \min\left(a_1 + b_1 t_1, \ln\left(k_1 \,/\, X_2\left(0\right)\right), a_2 + b_2 t_1\right)\mathbb{I}_{\left\{P_{[...]}=P_{[GU2]}\right\}}$$

$$+\max\left(a_1 + b_1 t_1, k_1, a_2\right)\mathbb{I}_{\left\{P_{[...]}=P_{[AL1]}\right\}} + \max\left(a_1 + b_1 t_1, k_1, a_2 + b_2 t_1\right)\mathbb{I}_{\left\{P_{[...]}=P_{[AL2]}\right\}}$$

$$+\max\left(a_1 + b_1 t_1, \ln\left(k_1 \,/\, X_2\left(0\right)\right), a_2\right)\mathbb{I}_{\left\{P_{[...]}=P_{[GL1]}\right\}} + \max\left(a_1 + b_1 t_1, \ln\left(k_1 \,/\, X_2\left(0\right)\right), a_2 + b_2 t_1\right)\mathbb{I}_{\left\{P_{[...]}=P_{[GL2]}\right\}}$$

$$z_2 = \min\left(a_2 + b_2\left(t_2 - t_1\right), k_2, a_3\right)\mathbb{I}_{\left\{P_{[...]}=P_{[AU1]}\right\}} + \min\left(a_2 + b_2\left(t_2 - t_1\right), \ln\left(k_2 \,/\, X_2\left(0\right)\right), a_3\right)\mathbb{I}_{\left\{P_{[...]}=P_{[GU1]}\right\}}$$

$$+\min\left(a_2 + b_2 t_2, k_2, a_3 + b_3 t_2\right)\mathbb{I}_{\left\{P_{[...]}=P_{[AU2]}\right\}} + \min\left(a_2 + b_2 t_2, \ln\left(k_2 \,/\, X_2\left(0\right)\right), a_3 + b_3 t_2\right)\mathbb{I}_{\left\{P_{[...]}=P_{[GU2]}\right\}}$$

$$+\max\left(a_2 + b_2\left(t_2 - t_1\right), k_2, a_3\right)\mathbb{I}_{\left\{P_{[...]}=P_{[AL1]}\right\}} + \max\left(a_2 + b_2\left(t_2 - t_1\right), \ln\left(k_2 \,/\, X_2\left(0\right)\right), a_3\right)\mathbb{I}_{\left\{P_{[...]}=P_{[GL1]}\right\}}$$

$$+\max\left(a_2 + b_2 t_2, k_2, a_3 + b_3 t_2\right)\mathbb{I}_{\left\{P_{[...]}=P_{[AL2]}\right\}} + \max\left(a_2 + b_2 t_2, \ln\left(k_2 \,/\, X_2\left(0\right)\right), a_3 + b_3 t_2\right)\mathbb{I}_{\left\{P_{[...]}=P_{[GL2]}\right\}}$$

$$z_3 = \min\left(a_3 + b_3\left(t_3 - t_2\right), k_3\right)\mathbb{I}_{\left\{P_{[\ldots]}=P_{[AU1]}\right\}} + \min\left(a_3 + b_3\left(t_3 - t_2\right), \ln\left(k_3 \,/\, X_2\left(0\right)\right)\right)\mathbb{I}_{\left\{P_{[\ldots]}=P_{[GU1]}\right\}}$$

$$+\min\left(a_3 + b_3 t_3, k_3\right)\mathbb{I}_{\left\{P_{[\ldots]}=P_{[AU2]}\right\}} + \min\left(a_3 + b_3 t_3, \ln\left(k_3 \,/\, X_2\left(0\right)\right)\right)\mathbb{I}_{\left\{P_{[\ldots]}=P_{[GU2]}\right\}}$$

$$+\max\left(a_3 + b_3\left(t_3 - t_2\right), k_3\right)\mathbb{I}_{\left\{P_{[\ldots]}=P_{[AL1]}\right\}} + \max\left(a_3 + b_3\left(t_3 - t_2\right), \ln\left(k_3 \,/\, X_2\left(0\right)\right)\right)\mathbb{I}_{\left\{P_{[\ldots]}=P_{[GL1]}\right\}}$$

$$+\max\left(a_3 + b_3 t_3, k_3\right)\mathbb{I}_{\left\{P_{[\ldots]}=P_{[AL2]}\right\}} + \max\left(a_3 + b_3 t_3, \ln\left(k_3 \,/\, X_2\left(0\right)\right)\right)\mathbb{I}_{\left\{P_{[\ldots]}=P_{[GL2]}\right\}}$$

The μ_i *terms,* $i \in \left\{1,2,3\right\}$*, in (2.15) are given by :*

$$\mu_1 = \left(\mu - b_1\right)\left(\mathbb{I}_{\left\{P_{[\ldots]}=P_{[AU1]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[AL1]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[AU2]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[AL2]}\right\}}\right)$$

$$+\left(\mu - \frac{\sigma^2}{2} - b_1\right)\left(\mathbb{I}_{\left\{P_{[\ldots]}=P_{[GU1]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[GL1]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[GU2]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[GL2]}\right\}}\right)$$

$$\mu_2 = \left(\mu - b_2\right)\left(\mathbb{I}_{\left\{P_{[\ldots]}=P_{[AU1]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[AL1]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[AU2]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[AL2]}\right\}}\right)$$

$$+\left(\mu - \frac{\sigma^2}{2} - b_2\right)\left(\mathbb{I}_{\left\{P_{[\ldots]}=P_{[GU1]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[GL1]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[GU2]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[GL2]}\right\}}\right)$$

$$\mu_3 = \left(\mu - b_3\right)\left(\mathbb{I}_{\left\{P_{[\ldots]}=P_{[AU1]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[AL1]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[AU2]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[AL2]}\right\}}\right)$$

$$+\left(\mu - \frac{\sigma^2}{2} - b_3\right)\left(\mathbb{I}_{\left\{P_{[\ldots]}=P_{[GU1]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[GL1]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[GU2]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[GL2]}\right\}}\right)$$

θ is given by :

$$\theta = 1\left(\mathbb{I}_{\left\{P_{[\ldots]}=P_{[AU1]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[AU2]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[GU1]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[GU2]}\right\}}\right)$$

$$-1\left(\mathbb{I}_{\left\{P_{[\ldots]}=P_{[AL1]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[AL2]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[GL1]}\right\}} + \mathbb{I}_{\left\{P_{[\ldots]}=P_{[GL2]}\right\}}\right)$$

End of Formula 1.

From a numerical perspective, Formula 1 raises the question of the evaluation of the function Φ_3. A straightforward calculation yields the following integration rule :

$$\Phi_3\left[x_1, x_2, x_3; \rho_1, \rho_2\right] = \int_{y_2=-\infty}^{x_2} \frac{\exp\left(-y_2^2 \,/\, 2\right)}{\sqrt{2\pi}} N\left[\frac{x_1 - \rho_1 y_2}{\sqrt{1-\rho_1^2}}\right] N\left[\frac{x_3 - \rho_2 y_2}{\sqrt{1-\rho_2^2}}\right] dy_2 \tag{2.17}$$

where the function $N\left[.\right]$ is the univariate standard normal cumulative distribution function.

Using (2.17), the numerical evaluation of the function Φ_3 is easy by means of a classical adaptive Gauss-Legendre quadrature. Alternatively, the following identities can be verified :

$$\Phi_3\left[x_1, x_2, x_3; \sqrt{t_1 \,/\, t_2}, \sqrt{t_2 \,/\, t_3}\right] = N_3\left[x_1, x_2, x_3; \sqrt{t_1 \,/\, t_2}, \sqrt{t_1 \,/\, t_3}, \sqrt{t_2 \,/\, t_3}\right] \tag{2.18}$$

$$\Phi_3\left[x_1,x_2,x_3;-\sqrt{t_1/t_2},-\sqrt{t_2/t_3}\right]=N_3\left[x_1,x_2,x_3;-\sqrt{t_1/t_2},\sqrt{t_1/t_3},-\sqrt{t_2/t_3}\right] \tag{2.19}$$

$$\Phi_3\left[x_1,x_2,x_3;-\sqrt{t_1/t_2},\sqrt{t_2/t_3}\right]=N_3\left[x_1,x_2,x_3;-\sqrt{t_1/t_2},-\sqrt{t_1/t_3},\sqrt{t_2/t_3}\right] \tag{2.20}$$

$$\Phi_3\left[x_1,x_2,x_3;\sqrt{t_1/t_2},-\sqrt{t_2/t_3}\right]=N_3\left[x_1,x_2,x_3;\sqrt{t_1/t_2},-\sqrt{t_1/t_3},-\sqrt{t_2/t_3}\right] \tag{2.21}$$

where the function $N_3\left[.,.,.;.,.,.\right]$ is the trivariate standard normal cumulative distribution function, the numerical evaluation of which can be performed with double precision and computational time of approximately 0.01 second using the algorithm by Genz (2004).

A few numerical results are reported in Table 1, in which the $\mathrm{P}_{[\mathrm{AU1}]}$ survival probability is computed for increasing levels of the volatility coefficient σ and other parameters fixed as follows : $\mu=0.01$, $t_1=0.25$, $t_2=0.5$, $t_3=1$, $k_1=0$, $k_2=0.02$, $k_3=0.03$, $a_1=0.22$, $b_1=-0.12$, $a_2=a_1+b_1t_1$, $b_2=0.16$, $a_3=a_2+b_2\left(t_2-t_1\right)$, $b_3=-0.24$. Notice that the absorbing boundary here is continuous at times t_1 and t_2, but non-continuous boundaries can be handled just as easily. Formula 1 is implemented using the algorithm by Genz (2004) for the computation of the trivariate standard normal cumulative distribution function. The results are compared with those obtained using the semi-analytical Monte Carlo algorithm devised by Wang and Pötzelberger (1997), denoted by WP simulation algorithm, that enables to draw only the endpoints of the time subintervals at each run, which is dramatically more efficient and accurate than a basic Monte Carlo simulation. Random numbers are drawn by the Mersenne Twister generator.

For all computed values, a 5-digit convergence can be observed between Formula 1 and the WP algorithm, on condition that a total of 100,000,000 stochastic simulations are performed. The latter method requires a computational time of 411 seconds on an i-7 4GHz personal computer. This is cut to 42 seconds when only 10,000,000 simulations are performed, which achieves 5-digit convergence in 2 cases out of 3 and 4-digit convergence in one case. The numerical computation of Formula 1 takes approximately 0.2 second. The efficiency of the implementation of the WP algorithm could probably be improved, for instance by resorting to low discrepancy sequences instead of a pseudo random number generator, but this is not the subject of this article.

Table 1. Numerical evaluation of the survival probability of an arithmetic Brownian under a one-sided piecewise affine, time-homogeneous, absorbing boundary, as a function of volatility

	Formula 1	WP simulation algorithm 10,000,000 runs	WP simulation algorithm 100,000,000 runs
Volatility = 20%	0.275332974	0.275318288	0.275387164
Volatility = 50%	0.257810712	0.257763484	0.257885116
Volatility = 80%	0.191728749	0.191718319	0.191716445

2.3 Proof of Formula 1

Only sequences of upper boundaries are tackled, since the results for sequences of lower boundaries ensue by symmetry of Brownian paths.

Let us deal with process X_1 first. Let us denote by p the sought probability when the boundary is defined by $g_1\left(t\right)$

in (2.3). The random variables $X_1(t_1)$, $X_1(t_2)$ and $X_1(t_3)$ are absolutely continuous random variables that admit known Gaussian density functions. At time t_1, X_1 must be located below $a_1 + b_1 t_1, k_1$ and a_2, in order not to be absorbed; at time t_2, it must stand underneath the points $a_2 + b_2(t_2 - t_1), k_2$ and a_3; at time t_3, it must end below $a_3 + b_3(t_3 - t_2)$ and k_3. Hence, by conditioning with respect to $X_1(t_1)$, $X_1(t_2)$ and $X_1(t_3)$, and by using the weak Markov property of $\{X_1(t), t \geq 0\}$, one can come up with the following integral formulation of the problem :

$$p = \int_{x_1=-\infty}^{z_1} \int_{x_2=-\infty}^{z_2} \int_{x_3=-\infty}^{z_3} \mathbb{P}\big(\big(X_1(t_1) \in dx_1\big) \cap \big(X_1(t) < a_1 + b_1 t, \forall 0 \leq t \leq t_1\big)\big) \tag{2.22}$$

$$\mathbb{P}\big(\big(X_1(t_2) \in dx_2\big) \cap \big(X_1(t) < a_2 + b_2(t - t_1), \forall t_1 \leq t \leq t_2\big) \big| X_1(t_1) \in dx_1\big)$$

$$\mathbb{P}\big(\big(X_1(t_3) \in dx_3\big) \cap \big(X_1(t) < a_3 + b_3(t - t_2), \forall t_2 \leq t \leq t_3\big) \big| X_1(t_2) \in dx_2\big) dx_3 dx_2 dx_1$$

$$= \int_{x_1=-\infty}^{z_1 - b_1 t_1} \int_{x_2=-\infty}^{z_2 - b_2 t_2} \int_{x_3=-\infty}^{z_3 - b_3 t_3} f_1(x_1) f_2(x_1, x_2) f_3(x_2, x_3) dx_3 dx_2 dx_1 \tag{2.23}$$

where the functions $f_1(x_1)$, $f_2(x_1, x_2)$ and $f_3(x_2, x_3)$ are defined by :

$$f_1(x_1) = \mathbb{P}\left(Y(t_1) \in dx_1, \sup_{0 \leq t \leq t_1} Y(t) < a_1\right) \tag{2.24}$$

$$f_2(x_1, x_2) = \mathbb{P}\left(Y(t_2) \in dx_2, \sup_{t_1 \leq t \leq t_2} Y(t) < a_2 - b_2 t_1 \big| Y(t_1) \in dx_1\right) \tag{2.25}$$

$$f_3(x_2, x_3) = \mathbb{P}\left(Y(t_3) \in dx_3, \sup_{t_2 \leq t \leq t_3} Y(t) < a_3 - b_3 t_2 \big| Y(t_2) \in dx_2\right) \tag{2.26}$$

and the process $\{Y(t), t \geq 0\}$ is defined by :

$$dY(t) = \begin{cases} \mu_1 dt + \sigma dB(t), \forall 0 \leq t < t_1 \\ \mu_2 dt + \sigma dB(t), \forall t_1 \leq t \leq t_2 \\ \mu_3 dt + \sigma dB(t), \forall t_2 \leq t \leq t_3 \end{cases} \tag{2.27}$$

$$\mu_i = \mu - b_i, i \in \{1, 2, 3\}$$

The function $f_1(x_1)$ is obtained by differentiating the classical formula for the joint distribution of the maximum of Brownian motion with drift and its endpoint over the closed time interval $[0, t_1]$ (see, e.g., Karatzas and Shreve, 1991).

To obtain the functions $f_2(x_1,x_2)$ and $f_3(x_2,x_3)$, the following lemma is introduced.

Lemma 1 *Let* $\{Y(t),t\geq 0\}$ *be an arithmetic Brownian motion with constant drift* $\mu\in\mathbb{R}$ *and volatility* $\sigma\in\mathbb{R}_+$ *under a given probability measure* $\mathbb{P}$. *Let* t_i *and* t_j *be two non-random times such that* $t_j>t_i>t_0=0$.

Then, if x_i, x_j *and* h *are real constants with* $x_i<h$ *and* $x_j<h$, *we have, at time* t_0 :

$$\mathbb{P}\left(Y(t_i)\leq x_i, Y(t_j)\leq x_j, \sup_{t_i\leq t\leq t_j} Y(t)\leq h\right) \tag{2.28}$$

$$= N_2\left[\frac{x_i-\mu t_i}{\sigma\sqrt{t_i}},\frac{x_j-\mu t_j}{\sigma\sqrt{t_j}};\sqrt{\frac{t_i}{t_j}}\right]-\exp\left(\frac{2\mu h}{\sigma^2}\right)N_2\left[\frac{x_i+\mu t_i}{\sigma\sqrt{t_i}},\frac{x_j-2h-\mu t_j}{\sigma\sqrt{t_j}};-\sqrt{\frac{t_i}{t_j}}\right]$$

where the function $N_2[x_1,x_2;\rho]$ *is the bivariate standard normal cumulative distribution function with upper bounds* x_1 *and* x_2 *and correlation coefficient* ρ

Proof of lemma 1

$$\mathbb{P}\left(Y(t_i)\leq x_i, Y(t_j)\leq x_j, \sup_{t_i\leq t\leq t_j} Y(t)\leq h\right)$$

$$= \int_{-\infty}^{x_i}\int_{-\infty}^{x_j}\mathbb{P}\big(Y(t_i)\in dy, Y(t_j)\in dz\big)\mathbb{P}\left(\sup_{t_i\leq t\leq t_j} Y(t)\leq h\,\middle|\,Y(t_i)\in dy, Y(t_j)\in dz\right)dydz \tag{2.29}$$

The pair $\big(Y(t_i),Y(t_j)\big)$ is bivariate normal with correlation coefficient equal to $\sqrt{t_i/t_j}$. The conditional cumulative distribution function of $\sup_{t_i\leq t\leq t_j} Y(t)$ is given by Wang and Pötzelberger (1997) and can be written as follows :

$$\mathbb{P}\left(\sup_{t_i\leq t\leq t_j} Y(t)\leq h\,\middle|\,Y(t_i)\in dy, Y(t_j)\in dz\right)=1-\exp\left(\frac{2(h-y)(z-h)}{\sigma^2(t_j-t_i)}\right) \tag{2.30}$$

One can then solve the integration problem in (2.29) to obtain (2.28).

□

Differentiating the right-hand side of (2.29) and dividing by the density function of $Y(t_i)$, one can obtain :

$$\phi(x_i,x_j,h,\mu,\sigma,t_i,t_j)=\mathbb{P}\left(Y(t_j)\in dx_j, \sup_{t_i\leq t\leq t_j} Y(t)\leq h\,\middle|\,Y(t_i)\in dx_i\right)$$

$$= \exp\left(-\frac{1}{2}\left(\frac{x_j - x_i - \mu\left(t_j - t_i\right)}{\sigma\sqrt{\left(t_j - t_i\right)}}\right)^2\right) / \left(\sigma\sqrt{2\pi\left(t_j - t_i\right)}\right) \tag{2.31}$$

$$-\exp\left(\frac{2\mu\left(h - x_i\right)}{\sigma^2}\right)\left(\exp\left(-\frac{1}{2}\left(\frac{x_j - 2h + x_i - \mu\left(t_j - t_i\right)}{\sigma\sqrt{\left(t_j - t_i\right)}}\right)^2\right) / \left(\sigma\sqrt{2\pi\left(t_j - t_i\right)}\right)\right)$$

Plugging :

$$f_2\left(x_1, x_2\right) = \phi\left(x_1, x_2, a_2 - b_2 t_1, \mu_2, \sigma, t_1, t_2\right) \tag{2.32}$$

$$f_3\left(x_2, x_3\right) = \phi\left(x_2, x_3, a_3 - b_3 t_2, \mu_3, \sigma, t_2, t_3\right) \tag{2.33}$$

into (2.23), the rest of the proof, whose details are omitted, then consists in performing the necessary calculations to solve the triple integral in (2.22) and obtain the linear combination of eight trivariate cumulative distribution functions given by Formula 1. Elementary modifications provide the survival probability when the boundary is defined by the function $g_2\left(t\right)$ in (2.4). A basic application of Ito's lemma to $\ln\left(X_2\left(t\right) / X_2\left(0\right)\right)$ shows that the survival probability of the process X_2 is given by the formula for the survival probability of the process X_1 with the two following adjustments : the drift coefficients become $\mu_i = \mu - b_i - \sigma^2 / 2$, $i \in \left\{1,2,3\right\}$ and k_i becomes $\ln\left(k_i / X_2\left(0\right)\right)$.

□

2.4 Generalization to Higher Dimension

Similar exact formulae can be derived for $n > 3$ but they become more and more cumbersome. In general, for any $n \in \mathbb{N}$, they will involve a number 2^n of the $n-$ variate cumulative distribution functions of Gaussian type given by (2.16). For an arithmetic Brownian motion subject to the absorbing boundary $g_1\left(t\right)$, the integration problem to solve is the following :

$$\int_{D^n} \prod_{i=0}^{n-1} \phi\left(x_i, x_{i+1}, a_{i+1} - b_{i+1} t_i, \mu_{i+1}, \sigma, t_i, t_{i+1}\right) dx_n dx_{n-1} ... dx_1 \tag{2.35}$$

where $x_0 = 0$ and

$$D^n = \left]-\infty, \min\left(a_1 + b_1 t_1, k_1, a_2\right)\right] \times \left]-\infty, \min\left(a_2 + b_2 t_2, k_2, a_3\right)\right] \times ... \times \left]-\infty, \min\left[a_{n-1} + b_{n-1} t_{n-1}, k_n\right]\right]$$

The main issue is numerical rather than analytical : evaluating the Gaussian integral given by (2.16) in high dimension is not easy. Rewriting it in terms of the standard normal cumulative distribution function of order n, as was done in (2.18) – (2.21) for $n = 3$, does not solve the numerical issue, as there does not exist an algorithm capable of evaluating the

$n-$variate standard normal cumulative distribution function with arbitrary precision in "reasonable" time as soon as $n=4$. For more background on this topic, the reader may refer to Genz and Bretz (2009).

However, for $n=4$, it can be verified that the following integration rule holds :

$$\Phi_4\left[x_1,x_2,x_3,x_4;\rho_1,\rho_2,\rho_3\right] \tag{2.36}$$

$$=\int_{y_2=-\infty}^{x_2}\int_{y_3=-\infty}^{\frac{x_3-\rho_2 y_2}{\sqrt{1-\rho_2^2}}}\frac{1}{2\pi}\exp\left(-\frac{\left(y_2^2+y_3^2\right)}{2}\right)N\left[\frac{x_1-\rho_1 y_2}{\sqrt{1-\rho_1^2}}\right]N\left[\frac{x_4-\rho_3\sqrt{1-\rho_2^2}y_3-\rho_3\rho_2 y_2}{\sqrt{1-\rho_3^2}}\right]dy_2dy_3$$

More generally, the actual numerical dimension of the function Φ_n can always be reduced by a factor of 2 by using :

$$\Phi_n\left[x_1,x_2,...,x_{n-1},x_n;\rho_1,...,\rho_{n-2},\rho_{n-1}\right] \tag{2.37}$$

$$=\int_{y_2=-\infty}^{x_2}\int_{y_3=-\infty}^{x_3}\cdots\int_{y_{n-1}=-\infty}^{x_{n-1}}N\left[\frac{x_1-\rho_1 y_2}{\sqrt{1-\rho_1^2}}\right]N\left[\frac{x_n-\rho_{n-1}y_{n-1}}{\sqrt{1-\rho_{n-1}^2}}\right]$$

$$\frac{\exp\left(-\frac{y_2^2}{2}-\frac{1}{2\left(1-\rho_2^2\right)}\left(y_3-\rho_2 y_2\right)^2...-\frac{1}{2\left(1-\rho_{n-2}^2\right)}\left(y_{n-1}-\rho_{n-2}y_{n-2}\right)^2\right)}{\prod_{i=2}^{n-2}\sqrt{\left(1-\rho_i^2\right)}\left(2\pi\right)^{\frac{n-2}{2}}}dy_2dy_3...dy_{n-1}$$

Given the smoothness of the integrand in (2.37), it should be possible to attain a combination of accuracy and efficiency that would be satisfactory for all practical purposes in "moderate" dimension, roughly speaking, by applying adaptive Gauss-Legendre quadrature combined with a Kronrod rule (Kronrod, 1964; Calvetti et al., 2000) to reduce the number of required iterations. These are standard numerical techniques and it is easy to find available code or built-in functions in the usual scientific computing software. The dimension n at which the use of a closed form formula analogous to Formula 1 ceases to be "competitive" with regard to a conditional Monte Carlo scheme should be numerically investigated. It must be emphasized that, even in "high" dimension, where Monte Carlo simulation becomes the method of last resort, exact formulae valid in lower dimension remain useful in two ways : they provide benchmarks with respect to which the accuracy of the numerical algorithms can be checked, and they can be used as control variates that substantially reduce the variance of the Monte Carlo estimates.

3. Survival Probability of an Arithmetic or a Geometric Brownian Motion under and above a Sequence of Two-sided affine or Exponential Boundaries over a Finite Time Interval

3.1 Definitions

Let us consider a finite time interval $\left[t_0,t_2\right]$ divided in two subintervals $\left[t_0,t_1\right]$ and $\left[t_1,t_2\right]$, $t_2\geq t_1\geq t_0=0$. The absorbing boundary now consists of two parallel upper and lower curves in each time interval, these curves being line segments when dealing with process X_1, or exponential curves when dealing with process X_2. More specifically, let $\mathrm{P}_{[...]}\left(\mu,\sigma,a_1,a_2,a_3,a_4,b_1,b_2,k_1,k_2,t_1,t_2\right)$ be defined as one of the following four cumulative distribution functions, where k_1 and k_2 are real constants :

$$\mathrm{P}_{[\mathrm{AUL1}]}\left(\mu,\sigma,a_1,a_2,a_3,a_4,b_1,b_2,k_1,k_2,t_1,t_2\right),\ a_1\in\mathbb{R}_+,a_2\in\mathbb{R}_-,\left(a_3,a_4,b_1,b_2,k_1,k_2\right)\in\mathbb{R}^8,\\ a_3>a_4,a_3>a_2+b_1t_1,\ t_2\geq t_1\geq t_0=0 \tag{3.1}$$

$$=\mathbb{P}\begin{pmatrix}\left(X_1\left(t\right)<a_1+b_1t,\forall 0\leq t\leq t_1\right)\cap\left(X_1\left(t\right)>a_2+b_1t,\forall 0\leq t\leq t_1\right)\cap X_1\left(t_1\right)<k_1\\ \cap\left(X_1\left(t\right)<a_3+b_2\left(t-t_1\right),\forall t_1\leq t\leq t_2\right)\cap\left(X_1\left(t\right)>a_4+b_2\left(t-t_1\right),\forall t_1\leq t\leq t_2\right)\cap X_1\left(t_2\right)<k_2\end{pmatrix}$$

$$\mathrm{P}_{[\mathrm{AUL2}]}\left(\mu,\sigma,a_1,a_2,a_3,a_4,b_1,b_2,k_1,k_2,t_1,t_2\right),\ a_1\in\mathbb{R}_+,a_2\in\mathbb{R}_-,\left(a_3,a_4,b_1,b_2,k_1,k_2\right)\in\mathbb{R}^8,\\ a_3>a_4,a_3+b_2t_1>a_2+b_1t_1,\ t_2\geq t_1\geq t_0=0 \tag{3.2}$$

$$=\mathbb{P}\begin{pmatrix}\left(X_1\left(t\right)<a_1+b_1t,\forall 0\leq t\leq t_1\right)\cap\left(X_1\left(t\right)>a_2+b_1t,\forall 0\leq t\leq t_1\right)\cap X_1\left(t_1\right)<k_1\\ \cap\left(X_1\left(t\right)<a_3+b_2t,\forall t_1\leq t\leq t_2\right)\cap\left(X_1\left(t\right)>a_4+b_2t,\forall t_1\leq t\leq t_2\right)\cap X_1\left(t_2\right)<k_2\end{pmatrix}$$

$$\mathrm{P}_{[\mathrm{GUL1}]}\left(\mu,\sigma,a_1,a_2,a_3,a_4,b_1,b_2,k_1,k_2,t_1,t_2\right),\ a_1\in\mathbb{R}_+,a_2\in\mathbb{R}_-,\left(a_3,a_4,b_1,b_2,k_1,k_2\right)\in\mathbb{R}^8,\\ a_3>a_4,a_3>a_2+b_1t_1,\ t_2\geq t_1\geq t_0=0 \tag{3.3}$$

$$=\mathbb{P}\begin{pmatrix}\left(X_2\left(t\right)<X_2\left(0\right)\exp\left(a_1+b_1t\right),\forall 0\leq t\leq t_1\right)\cap\left(X_2\left(t\right)>X_2\left(0\right)\exp\left(a_2+b_1t\right),\forall 0\leq t\leq t_1\right)\\ \cap X_2\left(t_1\right)<k_1\cap\left(X_2\left(t\right)<X_2\left(0\right)\exp\left(a_3+b_2\left(t-t_1\right)\right),\forall t_1\leq t\leq t_2\right)\\ \cap\left(X_2\left(t\right)>X_2\left(0\right)\exp\left(a_4+b_2\left(t-t_1\right)\right),\forall t_1\leq t\leq t_2\right)\cap X_2\left(t_2\right)<k_2\end{pmatrix}$$

$$\mathrm{P}_{[\mathrm{GUL2}]}\left(\mu,\sigma,a_1,a_2,a_3,a_4,b_1,b_2,k_1,k_2,t_1,t_2\right),\ a_1\in\mathbb{R}_+,a_2\in\mathbb{R}_-,\left(a_3,a_4,b_1,b_2,k_1,k_2\right)\in\mathbb{R}^8,\\ a_3>a_4,a_3+b_2t_1>a_2+b_1t_1,\ t_2\geq t_1\geq t_0=0 \tag{3.4}$$

$$=\mathbb{P}\begin{pmatrix}\left(X_2\left(t\right)<X_2\left(0\right)\exp\left(a_1+b_1t\right),\forall 0\leq t\leq t_1\right)\cap\left(X_2\left(t\right)>X_2\left(0\right)\exp\left(a_2+b_1t\right),\forall 0\leq t\leq t_1\right)\\ \cap X_2\left(t_1\right)<k_1\cap\left(X_2\left(t\right)<X_2\left(0\right)\exp\left(a_3+b_2t\right),\forall t_1\leq t\leq t_2\right)\\ \cap\left(X_2\left(t\right)>X_2\left(0\right)\exp\left(a_4+b_2t\right),\forall t_1\leq t\leq t_2\right)\cap X_2\left(t_2\right)<k_2\end{pmatrix}$$

3.2 Statement of Formula 2

Formula 2 *Let* $\mathrm{P}_{[\ldots]}\left(\mu,\sigma,a_1,a_2,a_3,a_4,b_1,b_2,k_1,k_2,t_1,t_2\right)$ *be defined as in Subsection 3.1. Then,*

$$\mathrm{P}_{[\ldots]}\left(\mu,\sigma,a_1,a_2,a_3,a_4,b_1,b_2,k_1,k_2,t_1,t_2\right)$$

$$=\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\exp\left(\frac{2\mu_1}{\sigma^2}m\theta+\frac{2\mu_2}{\sigma^2}n\phi\right) \tag{3.5}$$

$$\left\{\begin{array}{l} N_2\left[\dfrac{\beta_1-2m\theta+\lambda_1}{\sigma\sqrt{t_1}},\dfrac{\beta_3-2m\theta-2n\phi+\lambda_2}{\sigma\sqrt{t_2}};\sqrt{\dfrac{t_1}{t_2}}\right] \\ -N_2\left[\dfrac{\beta_2-2m\theta+\lambda_1}{\sigma\sqrt{t_1}},\dfrac{\beta_3-2m\theta-2n\phi+\lambda_2}{\sigma\sqrt{t_2}};\sqrt{\dfrac{t_1}{t_2}}\right] \\ -N_2\left[\dfrac{\beta_1-2m\theta+\lambda_1}{\sigma\sqrt{t_1}},\dfrac{\beta_4-2m\theta-2n\phi+\lambda_2}{\sigma\sqrt{t_2}};\sqrt{\dfrac{t_1}{t_2}}\right] \\ +N_2\left[\dfrac{\beta_2-2m\theta+\lambda_1}{\sigma\sqrt{t_1}},\dfrac{\beta_4-2m\theta-2n\phi+\lambda_2}{\sigma\sqrt{t_2}};\sqrt{\dfrac{t_1}{t_2}}\right] \end{array}\right\}$$

$$-\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\exp\left(\frac{2\mu_1}{\sigma^2}m\theta+\frac{2\mu_2}{\sigma^2}\big(\beta_4-n\phi-2m\theta\big)+\frac{2}{\sigma^2}\big(\mu_2^2-\mu_1\mu_2\big)t_1\right)$$

$$\left\{\begin{array}{l} N_2\left[\dfrac{\beta_1-2m\theta+\lambda_3}{\sigma\sqrt{t_1}},\dfrac{\beta_3-2\beta_4+2m\theta+2n\phi+\lambda_4}{\sigma\sqrt{t_2}};-\sqrt{\dfrac{t_1}{t_2}}\right] \\ -N_2\left[\dfrac{\beta_2-2m\theta+\lambda_3}{\sigma\sqrt{t_1}},\dfrac{\beta_3-2\beta_4+2m\theta+2n\phi+\lambda_4}{\sigma\sqrt{t_2}};-\sqrt{\dfrac{t_1}{t_2}}\right] \\ -N_2\left[\dfrac{\beta_1-2m\theta+\lambda_3}{\sigma\sqrt{t_1}},\dfrac{-\beta_4+2m\theta+2n\phi+\lambda_4}{\sigma\sqrt{t_2}};-\sqrt{\dfrac{t_1}{t_2}}\right] \\ +N_2\left[\dfrac{\beta_2-2m\theta+\lambda_3}{\sigma\sqrt{t_1}},\dfrac{-\beta_4+2m\theta+2n\phi+\lambda_4}{\sigma\sqrt{t_2}};-\sqrt{\dfrac{t_1}{t_2}}\right] \end{array}\right\}$$

$$-\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\exp\left(\frac{2\mu_1}{\sigma^2}\big[a_2-m\theta\big]+\frac{2\mu_2}{\sigma^2}\big[n\phi\big]\right)$$

$$\left\{\begin{array}{l} N_2\left[\dfrac{\beta_1-2a_2+2m\theta+\lambda_1}{\sigma\sqrt{t_1}},\dfrac{\beta_3-2a_2+2m\theta-2n\phi+\lambda_2}{\sigma\sqrt{t_2}};\sqrt{\dfrac{t_1}{t_2}}\right] \\ -N_2\left[\dfrac{\beta_2-2a_2+2m\theta+\lambda_1}{\sigma\sqrt{t_1}},\dfrac{\beta_3-2a_2+2m\theta-2n\phi+\lambda_2}{\sigma\sqrt{t_2}};\sqrt{\dfrac{t_1}{t_2}}\right] \\ -N_2\left[\dfrac{\beta_1-2a_2+2m\theta+\lambda_1}{\sigma\sqrt{t_1}},\dfrac{\beta_4-2a_2+2m\theta-2n\phi+\lambda_2}{\sigma\sqrt{t_2}};\sqrt{\dfrac{t_1}{t_2}}\right] \\ +N_2\left[\dfrac{\beta_2-2a_2+2m\theta+\lambda_1}{\sigma\sqrt{t_1}},\dfrac{\beta_4-2a_2+2m\theta-2n\phi+\lambda_2}{\sigma\sqrt{t_2}};\sqrt{\dfrac{t_1}{t_2}}\right] \end{array}\right\}$$

$$+\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\exp\left(\frac{2\mu_1}{\sigma^2}\big(a_2-m\theta\big)+\frac{2\mu_2}{\sigma^2}\big(\beta_4-n\phi+2m\theta-2a_2\big)+\frac{2}{\sigma^2}\big(\mu_2^2-\mu_1\mu_2\big)t_1\right)$$

$$\left\{\begin{array}{l} N_2\left[\dfrac{\beta_1-2a_2+2m\theta+\lambda_3}{\sigma\sqrt{t_1}},\dfrac{\beta_3-2\beta_4+2a_2+2n\phi-2m\theta+\lambda_4}{\sigma\sqrt{t_2}};-\sqrt{\dfrac{t_1}{t_2}}\right] \\ -N_2\left[\dfrac{\beta_2-2a_2+2m\theta+\lambda_3}{\sigma\sqrt{t_1}},\dfrac{\beta_3-2\beta_4+2a_2+2n\phi-2m\theta+\lambda_4}{\sigma\sqrt{t_2}};-\sqrt{\dfrac{t_1}{t_2}}\right] \\ -N_2\left[\dfrac{\beta_1-2a_2+2m\theta+\lambda_3}{\sigma\sqrt{t_1}},\dfrac{-\beta_4+2a_2+2n\phi-2m\theta+\lambda_4}{\sigma\sqrt{t_2}};-\sqrt{\dfrac{t_1}{t_2}}\right] \\ +N_2\left[\dfrac{\beta_2-2a_2+2m\theta+\lambda_3}{\sigma\sqrt{t_1}},\dfrac{-\beta_4+2a_2+2n\phi-2m\theta+\lambda_4}{\sigma\sqrt{t_2}};-\sqrt{\dfrac{t_1}{t_2}}\right] \end{array}\right\}$$

where the following notations hold :

$$\mu_1=\left(\mu-b_1\right)\left(\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{AUL1}]}\right\}}+\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{AUL2}]}\right\}}\right)+\left(\mu-\frac{\sigma^2}{2}-b_1\right)\left(\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{GUL1}]}\right\}}+\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{GUL2}]}\right\}}\right)$$

$$\mu_2=\left(\mu-b_2\right)\left(\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{AUL1}]}\right\}}+\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{AUL2}]}\right\}}\right)+\left(\mu-\frac{\sigma^2}{2}-b_2\right)\left(\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{GUL1}]}\right\}}+\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{GUL2}]}\right\}}\right)$$

$$\theta=a_1-a_2$$

$$\phi=a_3-a_4$$

$$\beta_1=\left(\min\left(a_1+b_1t_1,k_1,a_3\right)-b_1t_1\right)\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{AUL1}]}\right\}}+\left(\min\left(a_1+b_1t_1,k_1,a_3+b_2t_1\right)-b_1t_1\right)\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{AUL2}]}\right\}}$$

$$+\left(\min\left(a_1+b_1t_1,\ln\left(k_1\,/\,X_2\left(0\right)\right),a_3\right)-b_1t_1\right)\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{GUL1}]}\right\}}$$

$$+\left(\min\left(a_1+b_1t_1,\ln\left(k_1\,/\,X_2\left(0\right)\right),a_3+b_2t_1\right)-b_1t_1\right)\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{GUL2}]}\right\}}$$

$$\beta_2=\left(\max\left(a_2+b_1t_1,a_4\right)-b_1t_1\right)\left(\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{AUL1}]}\right\}}+\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{GUL1}]}\right\}}\right)$$

$$+\left(\max\left(a_2+b_1t_1,a_4+b_2t_1\right)-b_1t_1\right)\left(\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{AUL2}]}\right\}}+\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{GUL2}]}\right\}}\right)$$

$$\beta_3=\left(\min\left(a_3+b_2\left(t_2-t_1\right),k_2\right)-b_2t_2\right)\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{AUL1}]}\right\}}+\left(\min\left(a_3+b_2t_2,k_2\right)-b_2t_2\right)\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{AUL2}]}\right\}}$$

$$+\left(\min\left(a_3+b_2\left(t_2-t_1\right),\ln\left(k_2\,/\,X_2\left(0\right)\right)\right)-b_2t_2\right)\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{GUL1}]}\right\}}$$

$$+\left(\min\left(a_3+b_2t_2,\ln\left(k_2\,/\,X_2\left(0\right)\right)\right)-b_2t_2\right)\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{GUL2}]}\right\}}$$

$$\beta_4=\left(a_4-b_2t_1\right)\left(\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{AUL1}]}\right\}}+\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{GUL1}]}\right\}}\right)+a_4\left(\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{AUL2}]}\right\}}+\mathbb{I}_{\left\{\mathrm{P}_{[\ldots]}=\mathrm{P}_{[\mathrm{GUL2}]}\right\}}\right)$$

$$\lambda_1=-\mu_1t_1,\ \lambda_2=-\mu_1t_1-\mu_2\left(t_2-t_1\right),\ \lambda_3=-\mu_1t_1+2\mu_2t_1,\ \lambda_4=\mu_1t_1-2\mu_2t_1-\mu_2\left(t_2-t_1\right)$$

End of Formula 2.

A few numerical values are reported in Table 2 for various levels of volatility and other parameters fixed as follows :

$\mu=0.01$, $t_1=0.25$, $t_2=0.5$, $t_3=1$, $k_1=0$, $k_2=0.02$, $a_1=0.36$, $b_1=0.15$, $a_2=-0.42$,

$b_2 = 0.15$, $a_3 = a_1 + b_1 t_1$, $b_3 = -0.12$, $a_4 = a_2 + b_2 t_1$, $b_4 = -0.12$. A comparison is made with results obtained using the algorithm by Pötzelberger and Wang (2001), denoted by PW, specifically designed for two-sided boundaries. The infinite double series in Formula 2 is truncated to summation operators ranging from $m = -4$ to $m = 4$ and from $n = -4$ to $n = 4$, since adding more terms does not modify the obtained numerical results at least up to the 8th digit. Computational time is approximately 0.3 second. In general, the infinite double series can be truncated in a simple manner by setting a convergence threshold such that no further terms are added once the difference between two successive finite sums becomes smaller than that prespecified level.

Table 2. Numerical evaluation of the survival probability of an arithmetic Brownian under a two-sided piecewise affine, time-homogeneous, absorbing boundary, as a function of volatility

	Formula 2	PW simulation algorithm 10,000,000 runs	PW simulation algorithm 100,000,000 runs
Volatility = 20%	0.377958716	0.377923023	0.377971427
Volatility = 50%	0.12468422	0.124786191	0.124625324
Volatility = 80%	0.02140513	0.021276911	0.021382631

3.3 Proof of Formula 2

Let us consider the calculation of $\mathrm{P}_{[\mathrm{AUL1}]}$. Since the upper and the lower sides of the boundary grow at the same rate in each time interval, i.e. at the rate b_1 both from below and from above in $[t_0, t_1]$ and at the rate b_2 both from below and from above in $[t_1, t_2]$, the same technique can be applied as in the beginning of the proof of Formula 1, i.e. the initial boundary crossing problem is turned into one where the boundary and the drift of the process become piecewise constant. Hence, denoting by p the sought probability, the problem can be formulated as follows :

$$p = \int_{\beta_2}^{\beta_1} \int_{\beta_4}^{\beta_3} f_1(x_1) f_2(x_1, x_2) dx_2 dx_1 \tag{3.6}$$

where the functions $f_1(x_1)$ and $f_2(x_1, x_2)$ are defined by :

$$f_1(x_1) = \mathbb{P}\left(\sup_{0 \le t \le t_1} Y(t) < a_1, \inf_{0 \le t \le t_1} Y(t) > a_2 \middle| Y(t_1) \in dx_1 \right) \tag{3.7}$$

$$f_2(x_1, x_2) = \mathbb{P}\left(\sup_{t_1 \le t \le t_2} Y(t) < a_3 - b_2 t_1, \inf_{t_1 \le t \le t_2} Y(t) > a_4 - b_2 t_1, Y(t_2) \le x_2 \middle| Y(t_1) \in dx_1 \right) \tag{3.8}$$

and the process $\{Y(t), t \ge 0\}$ is defined by :

$$dY(t) = \begin{cases} \mu_1 dt + \sigma dB(t), \forall 0 \le t < t_1 \\ \mu_2 dt + \sigma dB(t), \forall t_1 \le t \le t_2 \end{cases} \tag{3.9}$$

$$\mu_i = \mu - b_i,\ i \in \{1,2\}$$

The function $f_1(x_1)$ results from the differentiation of the classical formula for the joint distribution of the maximum, the minimum and the endpoint of a Brownian motion (see, e.g., Cox & Miller, 1965). To obtain $f_2(x_1, x_2)$, the following lemma is introduced.

Lemma 2 *Let $Y(t)$ be an arithmetic Brownian motion with constant drift $\mu \in \mathbb{R}$ and volatility $\sigma \in \mathbb{R}_+$ under a given probability measure $\mathbb{P}$. Let q be the conditional probability defined, at time $t_0 = 0$, by :*

$$q = \mathbb{P}\left(\sup_{t_i \le t \le t_j} Y(t) \le b, \inf_{t_i \le t \le t_j} Y(t) \ge a, Y(t_j) \le x_j \middle| Y(t_i) \in dx_i \right) \tag{3.10}$$

where x_i, x_j, a and b are real constants such that : $b > a$, $b \ge x_i > a$, $b \ge x_j > a$, and t_i and t_j are two non-random times such that : $t_j > t_j \ge 0$. Then,

$$q = \sum_{n=-\infty}^{\infty} \exp\left(\frac{2n\mu(b-a)}{\sigma^2} \right) \left[\begin{array}{l} N\left(\dfrac{x_j - x_i - \mu(t_j - t_i) - 2n(b-a)}{\sigma\sqrt{t_j - t_i}} \right) \\ -N\left(\dfrac{a - x_i - \mu(t_j - t_i) - 2n(b-a)}{\sigma\sqrt{t_j - t_i}} \right) \end{array} \right] \tag{3.11}$$

$$- \sum_{n=-\infty}^{\infty} \exp\left(\frac{2\mu(a - x_i - n(b-a))}{\sigma^2} \right) \left[\begin{array}{l} N\left(\dfrac{x_j - 2a + x_i - \mu(t_j - t_i) + 2n(b-a)}{\sigma\sqrt{t_j - t_i}} \right) \\ -N\left(\dfrac{-a + x_i - \mu(t_j - t_i) + 2n(b-a)}{\sigma\sqrt{t_j - t_i}} \right) \end{array} \right]$$

Proof of lemma 2

$$\mathbb{P}\left(Y(t_i) \le x_i, \sup_{t_i \le t \le t_j} Y(t) \le b, \inf_{t_i \le t \le t_j} Y(t) \ge a, Y(t_j) \le x_j \right) \tag{3.12}$$

$$= \int_{y=a}^{x_i} \int_{z=a}^{x_j} \mathbb{P}(Y(t_i) \in dy, Y(t_j) \in dz) \mathbb{P}\left(\sup_{t_i \le t \le t_j} Y(t) \le b, \inf_{t_i \le t \le t_j} Y(t) \ge a \middle| Y(t_i) \in dy, Y(t_j) \in dz \right) dzdy$$

The following result can be found in Guillaume (2010) :

$$\mathbb{P}\left(\sup_{t_i \le t \le t_j} Y(t) \le b, \inf_{t_i \le t \le t_j} Y(t) \ge a \middle| Y(t_i) \in dy, Y(t_j) \in dz \right) \tag{3.13}$$

$$= \sum_{n=-\infty}^{\infty} \exp\left(\frac{2n(b-a)(z - y - n(b-a))}{\sigma^2(t_j - t_i)} \right) - \exp\left(\frac{2(b - y - n(b-a))(z - b + n(b-a))}{\sigma^2(t_j - t_i)} \right)$$

Plugging (3.13) into (3.12) yields :

$$\mathbb{P}\left(Y\left(t_i\right)\le x_i,\ \sup_{t_i\le t\le t_j}Y\left(t\right)\le b,\ \inf_{t_i\le t\le t_j}Y\left(t\right)\ge a, Y\left(t_j\right)\le x_j\right)$$

$$=\sum_{n=-\infty}^{\infty}\exp\left(\frac{2n\mu\left(b-a\right)}{\sigma^2}\right)\left\{N_2\left[\frac{x_i-\mu t_i}{\sigma\sqrt{t_i}},\frac{x_j-2n\left(b-a\right)-\mu t_j}{\sigma\sqrt{t_j}};\sqrt{\frac{t_i}{t_j}}\right]\right. \tag{3.14}$$

$$-N_2\left[\frac{x_i-\mu t_i}{\sigma\sqrt{t_i}},\frac{a-2n\left(b-a\right)-\mu t_j}{\sigma\sqrt{t_j}};\sqrt{\frac{t_i}{t_j}}\right]-N_2\left[\frac{a-\mu t_i}{\sigma\sqrt{t_i}},\frac{x_j-2n\left(b-a\right)-\mu t_j}{\sigma\sqrt{t_j}};\sqrt{\frac{t_i}{t_j}}\right]$$

$$\left.+N_2\left[\frac{a-\mu t_i}{\sigma\sqrt{t_i}},\frac{a-2n\left(b-a\right)-\mu t_j}{\sigma\sqrt{t_j}};\sqrt{\frac{t_i}{t_j}}\right]\right\}$$

$$-\sum_{n=-\infty}^{\infty}\exp\left(\frac{2\mu\left(a-n\left(b-a\right)\right)}{\sigma^2}\right)\left\{N_2\left[\frac{x_i+\mu t_i}{\sigma\sqrt{t_i}},\frac{x_j-2a+2n\left(b-a\right)-\mu t_j}{\sigma\sqrt{t_j}};-\sqrt{\frac{t_i}{t_j}}\right]\right.$$

$$-N_2\left[\frac{x_j+\mu t_i}{\sigma\sqrt{t_i}},\frac{-a+2n\left(b-a\right)-\mu t_j}{\sigma\sqrt{t_j}};-\sqrt{\frac{t_i}{t_j}}\right]-N_2\left[\frac{a+\mu t_i}{\sigma\sqrt{t_i}},\frac{x_j-2a+2n\left(b-a\right)-\mu t_j}{\sigma\sqrt{t_j}};-\sqrt{\frac{t_i}{t_j}}\right]$$

$$\left.+N_2\left[\frac{a+\mu t_i}{\sigma\sqrt{t_i}},\frac{-a+2n\left(b-a\right)-\mu t_j}{\sigma\sqrt{t_j}};-\sqrt{\frac{t_i}{t_j}}\right]\right\}$$

The interchange between summation and integral is a straightforward application of Tonelli's theorem to non-negative measurable functions, where the measures are the counting measure on $\mathbb{Z}$ and the Lebesgue measure on $\mathbb{R}$. Lemma 2 ensues by differentiating (3.14) and dividing by the density function of $Y\left(t_i\right)$.

□

Applying Lemma 2, the function $f_2\left(x_1,x_2\right)$ can be plugged in (3.6). Then, performing the necessary calculations, Formula 2 can be obtained.

□

4. Conclusion

In this paper, new formulae were obtained for the probability of absorption of generalised Brownian motion through sequences of affine or exponential one-sided or two-sided boundaries. It was shown that the method could be applied to higher numbers of successive one-sided boundaries. However, such an extension may not be commendable in the case of two-sided boundaries, as the resulting analytical formulae will involve a quickly increasing number of summation operators, thus slowing down the process of numerical convergence.

References

Anderson, T. W. (1960). A Modification of the Sequential Probability Ratio Test to Reduce the Sample Size. *Ann. Math. Statist.*, *31*, 165-197. http://dx.doi.org/10.1214/aoms/1177705996

Calvetti, D., Golub, G.H., Gragg, W.B., & Reichel, L. (2000). Computation of Gauss-Kronrod Quadrature Rules. *Math. Comput*., *69*, 1035-1052. http://dx.doi.org/10.1090/S0025-5718-00-01174-1

Choi, C., & Nam, D. (2003). Some Boundary Crossing Results for Linear Diffusion Processes. *Stat. Probab. Letters*, *62*, 281-291. http://dx.doi.org/10.1016/S0167-7152(03)00015-4

Cox, D.R. & Miller, H.D. (1965). *The Theory of Stochastic Processes*. London, Methuen Daniels, H.E. (1996). Approximating the First Crossing Time Density for a Curved Boundary, *Bernoulli*, *2*, 133-143.

http://dx.doi.org/10.2307/3318547

Doob, J. L. (1949). Heuristic Approach to the Kolmogorov-Smirnov Theorem, *Ann. Math. Statist*., *20*, 393-403

Genz, A. (2004). Numerical Computation of Rectangular Bivariate and Trivariate Normal and *t* Probabilities. *Stat. Comput*., *14*, 151-160. http://dx.doi.org/10.1214/aoms/1177729991

Genz, A., & Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*, Berlin Heidelberg: Springer-Verlag

Groeneboom, P. (1989). Brownian Motion with a Parabolic Drift and Airy Functions, *Prob. Theory Rel. Fields*, *81*, 79-109. http://dx.doi.org/10.1007/BF00343738

Guillaume, T. (2010). Step double barrier options. *J. derivatives, 18* (1), 59-79. http://dx.doi.org/10.3905/jod.2010.18.1.059

Jeanblanc, M., Yor, M., & Chesney, M. (2009). *Mathematical Methods for Financial Markets*, London: Springer-Verlag

Karatzas, I., & Shreve, S. (1991). *Brownian Motion and Stochastic Calculus*. New York : Springer-Verlag Kronrod, A.S. (1964), (Russian), *Doklady Akad. Nauk SSSR*, *154*, 283-286.

Levy, P. (1948). *Processus Stochastiques et Mouvement Brownien*, Paris : Gauthier-Villars Novikov, A. Frishling, V. , & Kordzakhia, N. (1999). Approximations of Boundary Crossing Probabilities for a Brownian Motion, *J.Appl.Prob*., *34*, 1019-1030. http://dx.doi.org/10.1239/jap/1032374752

Pötzelberger, K., & Wang, L. (2001). Boundary Crossing Probability for Brownian Motion, *J.Appl.Prob*., *38*, 152-164. http://dx.doi.org/10.1239/jap/996986650

Salminen, P. (1988). On the First Hitting Time and Last Exit Time for a Brownian Motion to/from a Moving Boundary, *Adv.Appl.Prob*., *20*, 411-426. http://dx.doi.org/10.2307/1427397

Scheike, T.H. (1992). A boundary crossing result for Brownian motion. *J.Appl.Prob*., *29*, 448-453. http://dx.doi.org/10.2307/3214581

Siegmund, D. (1986). Boundary Crossing Probabilities and Statistical Applications, *Ann. Statist*., *14*, 361-404. http://dx.doi.org/10.1214/aos/1176349928

Wang, L., & Pötzelberger, K. (1997). Boundary Crossing Probability for Brownian Motion and General Boundaries, *J.Appl.Prob*., *34*, 54-65. http://dx.doi.org/10.2307/3215174

Wang, L., & Pötzelberger, K. (2007). Crossing Probabilities for Diffusion Processes with Piecewise Continuous Boundaries, *Methodol.Comput. Appl.Probab*., *9*, 21-40. http://dx.doi.org/10.1007/s11009-006-9002-6

International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

On Shifted Weibull-Pareto Distribution

Ahmad Alzaghal[1], Indranil Ghosh[2] & Ayman Alzaatreh[3]

[1]Department of Mathematics, State University of New York at Oswego, Oswego, USA

[2] Department of Mathematics and Statistics, University of North Carolina Wilmington, Wilmington, NC

[3] Department of Mathematics, Nazarbayev University, Astana, KZ

Correspondence: Ahmad Alzaghal, Department of Mathematics, State University of New York at Oswego, Oswego, USA. Tel: 1-315-312-2728. E-mail: ahmad.alzaghal@oswego.edu.

Received: April 12, 2016 Accepted: April 24, 2016 Online Published: June 27, 2016

doi:10.5539/ijsp.v5n4p139 URL: http://dx.doi.org/10.5539/ijsp.v5n4p139

Abstract

The Lomax distribution, known as Pareto (type II) distribution, is a heavy tail probability distribution used extensively in business, economics and in actuarial modeling. The Weibull-Pareto distribution defined by Alzaatreh et al. (2013a) has shown high bias and standard error for the ML estimates when the parameter $c >> 1$. In this paper we use the Lomax distribution to construct the Weibull-Lomax distribution. It is observed that the Weibull-Lomax distribution performs significantly better in terms of the ML estimations. Some structural properties of the Weibull-Lomax distribution are discussed.

Keywords: Lomax distribution, Weibull-Lomax distribution, *T-X* family of distributions

1. Introduction

Let $r(t)$ be the probability density function (PDF) of a random variable $T \in [a, b]$ for $-\infty \leq a < b \leq \infty$ and let $G(x)$ be the cumulative distribution function (CDF) of a random variable X such that the link function $W(\cdot) : [0, 1] \longrightarrow [a, b]$ satisfies the following conditions: (i) $W(\cdot)$ is differentiable and monotonically non-decreasing, and (ii) $W(0) \to a$ and $W(1) \to b$.

The CDF of the *T-X family* defined by Alzaatreh et al. (2013b) is

$$F(x) = \int_a^{W(G(x))} r(t)\, dt. \tag{1}$$

If the random variable $T \in (0, \infty)$ and $W(z) = -\log(1 - z)$, then the PDF corresponding to (1) is given by

$$f(x) = h_g(x)\, R\Big(H_g(x)\Big), \tag{2}$$

where $h_g = g/(1 - G)$ and $H_g = -\log(1 - G)$ are the hazard and cumulative hazard rate functions corresponding to the PDF $g(x)$.

Alzaatreh et al. (2013a) studied a generalization of the Pareto distribution by using (2) where T and X follow Weibull and Pareto random variables respectively. The probability density function (PDF) of the Weibull-Pareto distribution (WPD) is defined as

$$f(x) = \frac{\beta c}{x}\left(\beta \log\left(\frac{x}{\theta}\right)\right)^{c-1} \mathrm{e}^{-\left(\beta \log(x/\theta)\right)^c}, \quad x > \theta; c, \beta, \theta > 0. \tag{3}$$

Alzaatreh et al. (2013a) studied some general properties of the Weibull-Pareto distribution in (3). They showed that this distribution can be applied to fit data with different characteristics. It can fit data with long right tail, long left tail and approximately symmetric. However, they pointed out a major problem in estimating the parameters of the Weibull-Pareto distribution (WPD). The estimated values of the parameters c and β using the maximum likelihood estimation (MLE) have high biases and standard errors when $c >> 1$. This occurred because when $c > 1$, the WPD can be left skewed. The left skewness of the WPD affects the ML estimates of the parameters. It was shown that the estimates of c and β are very sensitive to the estimate of the parameter θ. To solve this problem, Alzaatreh et al. (2013a) proposed a modification of the MLE method (MMLE) which improved the results of the MLE. However, they pointed out that MMLE still produce high bias and standard error values in some cases and further research needed to solve the estimation problem for the parameters of WPD. The main objective of this article is to provide an alternative to the Weibull-Pareto distribution namely, the Weibull-Lomax distribution. By merely considering another member of the family of Pareto distribution

(after suitable transformation) a greater improvement in estimation under the MLE (maximum likelihood method) is achieved. The paper is organized as follows: In section 2, the Weibull-Lomax distribution is defined and several properties are studied including quantile function, limit behaviour, unimoadality, Shannon entropy, reliability parameter, moments, mean deviations and order statistics. The asymptotic distributions of the sample minima and maxima are also studied in section 2. Parameter estimation and application are studied in section 3. We provide some concluding remarks in section 4.

2. The Weibull-Lomax Distribution

Let the random variable T and X follow the Weibull and Lomax distributions. Then from (2), the PDF of the Weibull-lomax distribution (WLD) can be written as

$$f(x) = \frac{\beta c}{x+\theta}\left(\beta \log\left(1+\frac{x}{\theta}\right)\right)^{c-1} \mathrm{e}^{-\left(\beta \log\left(1+\frac{x}{\theta}\right)\right)^{c}}, \quad x > 0; c, \beta, \theta > 0. \tag{4}$$

Remark 2.1. Note that the WLD in (4) is only a shift by θ of the WPD in (3). I.e. If $Y \sim WPD(c, \beta, \theta)$ then $X = Y - \theta$ follows $WLD(c, \beta, \theta)$.

When $c = 1$, the WLD reduces to the Lomax distribution with parameters β and θ. From (4), the CDF and hazard rate function of WLD, respectively, are

$$F(x) = 1 - \mathrm{e}^{-\left(\beta \log\left(1+\frac{x}{\theta}\right)\right)^{c}}. \tag{5}$$

$$h(x) = \frac{\beta c}{x+\theta}\left(\beta \log\left(1+\frac{x}{\theta}\right)\right)^{c-1}, \tag{6}$$

In Figures 1 and 2, various graphs of $f(x)$ and $h(x)$ are provided for different parameter values. The plots indicate that the Weibull-Lomax distribution can be reverse J-shaped, right-skewed or left-skewed. Also, the Weibull-Lomax distribution hazard function can be a decreasing failure rate or upside down bathtub shapes.

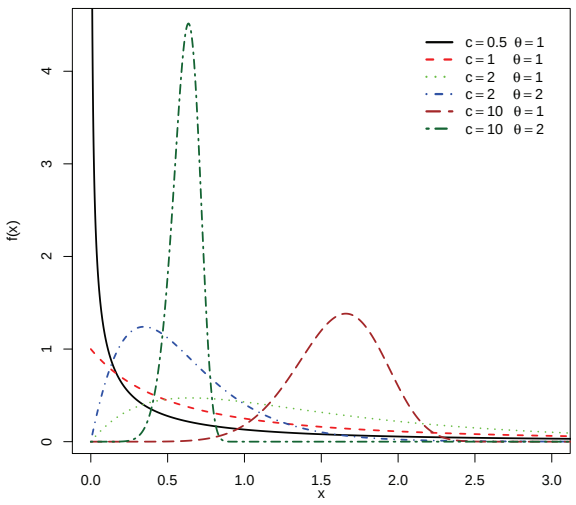


Figure 1. Graphs of the Weibull-Lomax PDF for various choices of c and θ when $\beta = 1$.

Remark 2.2. Based on Remark 2.1 and Alzaatreh et al. (2013a), one can obtain the following properties of the WLD.

(i) If a random variable Y follows the Weibull distribution with parameters c and $1/\beta$, then the random variable $X = \theta(\mathrm{e}^{Y} - 1)$ follows the Weibull-Lomax distribution.

(ii) Let $Q(p)$, $0 < p < 1$ denote the quantile function of WLD. Then $Q(p) = \theta\left\{\exp\left((-\log(1-p))^{1/c}/\beta\right) - 1\right\}$.

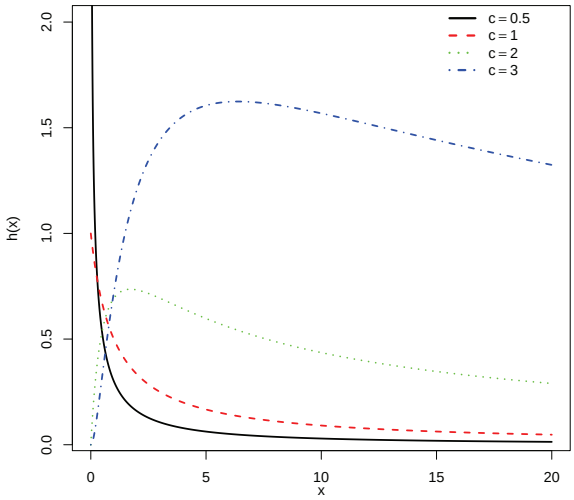


Figure 2. Graphs of the Weibull-Lomax hazard function for various choices of c when $\beta = \theta = 1$

(iii) The limit of the Weibull-Lomax density and the Weibull-Lomax hazard function as $x \to \infty$ is 0, and the limit as $x \to 0$ is given by

$$\lim_{x\to 0} f(x) = \lim_{x\to 0} h(x) = \begin{cases} 0, & c > 1 \\ \beta/\theta, & c = 1. \\ \infty, & c < 1 \end{cases}$$

(iv) The WLD is unimodal at x_0. When $c \leq 1$, the mode is at $x_0 = 0$ and when $c > 1$, the mode is the solution of the equation $k(x_0) = 0$, where

$$k(x) = -\log(1 + x/\theta) - c\left(\beta \log(1 + x/\theta)\right)^c + c - 1.$$

(v) The Shannon entropy (Shannon, 1948) for a random variable X that follows the WLD is

$$\eta_X = -\log\left(\frac{\beta c}{\theta}\right) + \frac{1}{\beta}\Gamma\left(1 + \frac{1}{c}\right) + \left(1 - \frac{1}{c}\right)\delta + 1,$$

where $\delta = -\int_0^\infty e^{-u}\log(u)du = 0.57722$ is the Euler gamma constant.

2.1 Reliability Parameter

The reliability parameter R is defined as $R = P(X > Y)$, where X and Y are independent random variables. Numerous applications of the reliability parameter have appeared in the literature such as the area of classical stress-strength model and the break down of a system having two components. Other applications of the reliability parameter can be found in Hall (1984) and Weerahandi and Johnson (1992). If X and Y are two continuous and independent random variables with the CDFs $F_1(x)$ and $F_2(y)$ and their PDFs $f_1(x)$ and $f_2(y)$ respectively, then the reliability parameter R can be written as

$$R = P(X > Y) = \int_{-\infty}^{\infty} F_2(t) f_1(t) dt.$$

Theorem 2.3. *Suppose that $X \sim WLD(c_1, \beta_1, \theta)$ and $Y \sim WLD(c_2, \beta_2, \theta)$, then*

$$R = 1 - \sum_{k=0}^{\infty} \frac{(-1)^k}{k!}\left(\frac{\beta_2}{\beta_1}\right)^{kc_2}\Gamma\left(\frac{kc_2}{c_1} + 1\right).$$

Proof. Follows from Lemma 2.4 in Alzaatreh and Ghosh (2014). □

2.2 Moments and Mean Deviations

Using Remark 2.1, the s-th moments for the WLD can be written as

$$E(X^s) = (-1)^s \theta^s + \theta^s \sum_{k=1}^{s} \sum_{i=0}^{\infty} \binom{s}{k} \frac{(-1)^{s-k} k^i}{i! \beta^i} \Gamma(1 + i/c). \tag{7}$$

Remark 2.4. From (7), we have the following

(i) If $c > 1$, then the s-th moments of the Weibull-Lomax distribution exist.

(ii) If $c < 1$, then the s-th moments of the Weibull-Lomax distribution do not exist.

(iii) If $c = 1$, then the s-th moments of the Weibull-Lomax distribution exist iff $\beta > s$.

The amount of scatter in a population is evidently measured to some extent by the totality of deviations from the mean and median. If we denote the median by M, then the mean deviation from the mean, $D(\mu)$, and the mean deviation from the median, $D(M)$, can be written as

$$D(\mu) = 2\mu F(\mu) - 2 \int_0^{\mu} x f(x) dx. \tag{8}$$

$$D(M) = \mu - 2 \int_0^{M} x f(x) dx. \tag{9}$$

Now, consider $I_m = \int_0^m x f(x) dx$, where $f(x)$ is defined in (4). On Using the substitution $u = (\beta \log(1 + x/\theta))^c$, we get

$$\begin{aligned} I_m &= \theta \sum_{k=0}^{\infty} \frac{1}{\beta^k k!} \int_0^{a_m} u^{k/c} \mathrm{e}^{-u} du \\ &= \theta \sum_{k=0}^{\infty} \frac{\gamma(a_m, 1 + k/c)}{\beta^k k!}, \quad c > 1, \end{aligned} \tag{10}$$

where $m > 0$, $a_m = (\beta \log(1 + m/\theta))^c$ and $\gamma(x, a) = \int_0^x t^{a-1} e^{-t} dt$ is the lower incomplete gamma function.

From equations (8) and (9), the mean deviation from the mean and the mean deviation from the median are, respectively, given by

$$\begin{aligned} D(\mu) &= 2\mu \left(1 - \exp\left\{-\left(\beta \log(1 + \mu/\theta)\right)^c\right\}\right) - 2I_\mu \\ D(M) &= \mu - 2I_M, \end{aligned}$$

where I_μ and I_M can be calculated numerically from (10).

2.3 Quantile Measures of Tail Behavior

Skewness and kurtosis of a distribution can be measured by $\beta_1 = \mu_3/\sigma^3$ and $\beta_2 = \mu_4/\sigma^4$, respectively. However the expression for the third and fourth moments of WLD(c, β, θ) do not always exist (see Remark 2.4). Consequently, the moment based skewness and kurtosis measures can not be applied when $c < 1$. Instead we consider the quantile based skewness and kurtosis measures which always exist. The quantile function of WLD(c, β, θ) are in closed form, alternatively we can define the measure of skewness and kurtosis based on the quantile function. The Galton' skewness S defined by Galton (1883) and the Moors' kurtosis K defined by Moors (1988) are given by

$$S = \frac{Q(6/8) - 2Q(4/8) + Q(2/8)}{Q(6/8) - Q(2/8)}. \tag{11}$$

$$K = \frac{Q(7/8) - Q(5/8) + Q(3/8) - Q(1/8)}{Q(6/8) - Q(2/8)}. \tag{12}$$

When the distribution is symmetric, $S = 0$ and when the distribution is right (or left) skew, $S > 0$ (or $S < 0$). As K increases the tail of the distribution becomes heavier. To investigate the effect of the two shape parameters c and β on the WLD(c, β, θ) distribution, equation (11) and (12) are used to obtain the Galtons' skewness and Moors' kurtosis where

the quantile function can be found from Remark 2.2. Figure 3 displays 3D graphs for the Galtons' skewness and Moors' kurtosis for the WLD(c,β,θ) when $\theta = 1$ and different values of c and β. From Figure 3, the WLD can be left skewed, right skewed or near symmetric (S=0). Furthermore, for fixed value of $c > 1$, Galton's skewness and Moors' kurtosis are decreasing function of β. Also, for fixed value of β, Galton's skewness and Moors' kurtosis are decreasing function of c. In Figure 4, we determined the regions when the WLD is left skewed, right skewed or nearly symmetric (S=0) occur. A cubic function relating $\log(\beta)$ to $\log(c)$ is obtained for the situation when the distribution is nearly symmetric (see R results in the Appendix). Since the symmetry of WLD does not depend on θ, we assumed $\theta = 1$.

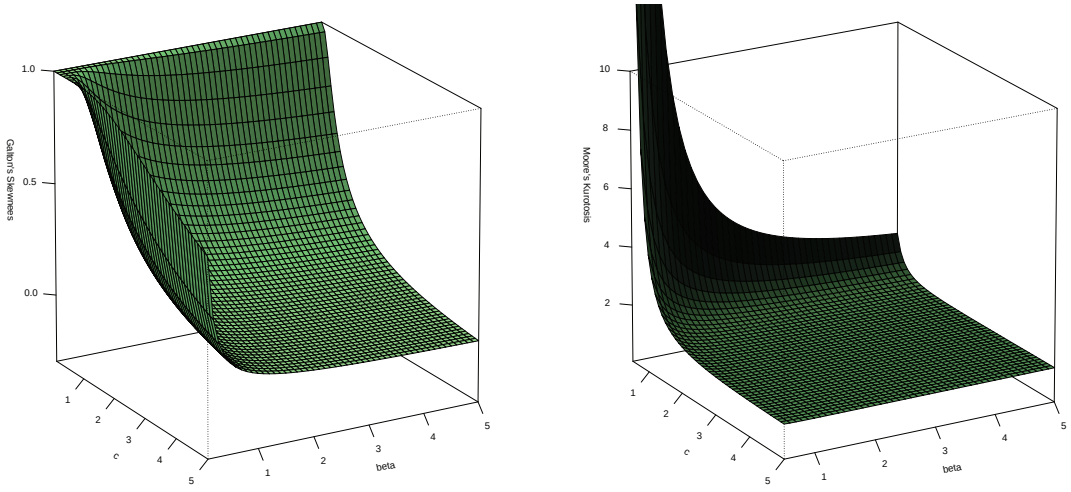


Figure 3. Graphs of Quantile Skewness and Kurtosis for the WL PDF when $\theta = 1$.

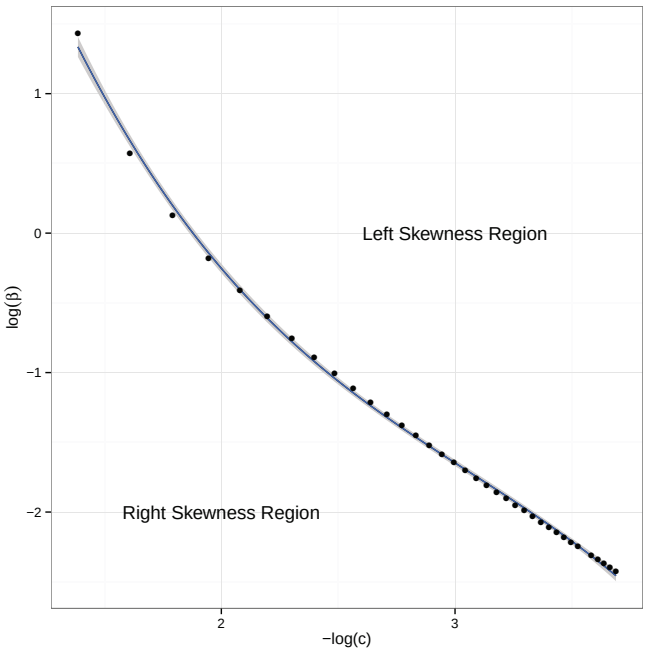


Figure 4. Skewness regions for the WLD.

2.4 Order Statistics for the Weibull Lomax Distribution

In this section, we study the moments of the r-th order statistic and the large sample distribution of the sample minimum and the sample maximum when a random sample of size n are drawn from the WLD(c,β,θ) distribution. The density

function of the r-th order statistic $X_{r:n}$ for a random sample of size n drawn from (4), is given by

$$\begin{aligned} f_{X_{r:n}}(x) &= r\binom{n}{r}(F(x))^{r-1}(1-F(x))^{n-r}f(x) \\ &= r\binom{n}{r}\sum_{k=0}^{r-1}(-1)^k\binom{r-1}{k}\frac{WLD(c,\beta(n-r+k+1)^{1/c},\theta)}{(n-r+k+1)}. \end{aligned} \tag{13}$$

From (13), the PDF of the rth order statistic $X_{r:n}$ can be expressed as a finite sum of the Weibull-Lomax PDFs. From (13), the distribution of the sample maximum $X_{n:n} = \max(X_1, X_2, \cdots, X_n)$, and the sample minimum $X_{1:n} = \min(X_1, X_2, \cdots, X_n)$ are, respectively, given by

$$f_{X_{n:n}}(x) = n\sum_{k=0}^{n-1}\frac{(-1)^k}{k+1}\binom{n-1}{k}WLD(c,\beta(k+1)^{1/c},\theta). \tag{14}$$

$$f_{X_{1:n}}(x) = WLD(c,\beta n^{1/c},\theta). \tag{15}$$

From (15), it is evident that the Weibull-Lomax distribution is closed under minimization. This property is also known as min stable property (See Feller, 1971).

In order to find the large sample distribution of $X_{n:n}$, we will use the sufficient condition for weak convergence due to von Mises (1936) which is stated in the following theorem:

Theorem 2.5. *Let F be an absolutely continuous c.d.f and suppose $h_f(x)$ is nonzero and differentiable function. If*

$$\lim_{x\to F^{-1}(1)}\frac{d}{dx}\left(\frac{1}{h_f(x)}\right) = 0,$$

then $F \in \mathcal{D}(G_1)$, where $G_1(x) = \exp(-\exp(x))$.

In our case $F^{-1}(1) = \infty$ and from (6), we have

$$\lim_{x\to\infty}\frac{d}{dx}\left(\frac{1}{h_f(x)}\right) = \begin{cases} 0, & c \geq 1 \\ \infty, & c < 1. \end{cases} \tag{16}$$

Hence, the large sample distribution of $X_{n:n}$ is of extreme value type provided that $c > 1$. When $c = 1$, the Weibull Lomax distribution reduces to the Lomax distribution. Also, according to Arnold et al. (2008), the large sample distribution of $X_{n:n}$ will be of extreme value type.

In order to derive the asymptotic distribution of the sample minima $X_{1:n}$, we consider Theorem 8.3.6 of Arnold et al.(2008). Observe that, since $F^{-1}(0)$ is finite, it follows from the theorem that the asymptotic distribution of the sample minima $X_{1:n}$ is not of Fréchet type. The asymptotic distribution of $X_{1:n}$ will be of the Weibull type with parameter $\alpha > 0$ if

$$\lim_{\varepsilon\to 0_+}\frac{F(F^{-1}(0)+\varepsilon x)}{F(F^{-1}(0)+\varepsilon)} = x^\alpha, \quad \text{for all} \quad x > 0.$$

Note that

$$\lim_{\varepsilon\to 0_+}\frac{F(\theta+\varepsilon x)}{F(\theta+\varepsilon)} = x\lim_{\varepsilon\to 0_+}\frac{f(\theta+\varepsilon x)}{f(\theta+\varepsilon)} = x^c.$$

Hence, the asymptotic distribution of the sample minima $X_{1:n}$ is of the Weibull type with shape parameter c. Furthermore, since the WLD is only a shift by θ of the WPD, therefore, the asymptotic distributions for the sample maxima when $c \geq 1$ is of extreme type. Also, the asymptotic distribution the sample minima for the WPD is of the Weibull type with shape parameter c.

Next, we consider the s-th moment for $X_{r:n}$, $1 \leq r \leq n$. From (13), if $Y \sim WLD(c,\beta(n-r+k+1)^{1/c},\theta)$, then

$$E(Y^s) = (-1)^s\theta^s + \theta^s\sum_{k=1}^{s}\sum_{i=0}^{\infty}\binom{s}{k}\frac{(-1)^{s-k}k^i}{i!\beta^i(n-r+k+1)^{i/c}}\Gamma(1+i/c). \tag{17}$$

From (13), the sth moment of $X_{r:n}$ can be written as

$$E(X_{r:n}^s) = r\binom{n}{r}\sum_{k=0}^{r-1}\frac{(-1)^k}{n-r+k+1}\binom{r-1}{k}E(Y^s), \tag{18}$$

where $E(Y^s)$ are given in (17). The s-th non central moment of $X_{1:n}$ and $X_{n:n}$ can be obtained by setting $r = 1$ and $r = n$ in equation (18).

Theorem 2.6. *(i) If $c > 1$, then the s-th non central moment of $X_{r:n}$ exist.*

(ii) If $c < 1$, then the s-th non central moment of $X_{r:n}$ do not exist.

(iii) If $c = 1$, then the s-th non central moment of $X_{r:n}$ exist iff $\beta(n-r+1) > s$.

Proof. Follows directly from Remark 2.4. □

3. Parameter Estimation and Application

The parameters of WLD are estimated by using the maximum likelihood. A simulation study is conducted to evaluate the performance of the maximum likelihood method.

3.1 Maximum Likelihood Estimation Method

Let $X_1, X_2, \cdots, X_n$ be a random sample of size n drawn from the density in (4). The log-likelihood function $\ell = \ell(c, \beta, \theta)$ is given by

$$\ell = nc\log\beta + n\log c - \sum_{i=1}^{n}\log(\theta + x_i) + (c-1)\sum_{i=1}^{n}\log(\log(1 + x_i/\theta)) - \sum_{i=1}^{n}(\beta\log(1 + x_i/\theta))^c. \tag{19}$$

The log-likelihood can be maximized numerically to obtain the maximum likelihood estimates. Several routines available for numerical maximization. We used the PROC NLMIXED in SAS to maximize equation (19). The initial estimates for the parameters of WLD are obtained as follows: the initial estimates of c and β are the moment estimates of c and β from the Weibull distribution, which are given by $c_0 = \frac{\pi}{(6s_{\log x_i})}$ and $\beta_0 = \exp\left(-\bar{x}_{\log x_i} - \delta/c_0\right)$, where $s_{\log x_i}$ and $\bar{x}_{\log x_i}$ are the sample standard deviation and the sample mean for $\log x_i$ and δ is the Euler gamma constant (Johnson et al., 1994, pp. 642-643). The initial estimate of θ is taken to be 1.

3.2 Simulation Study to Evaluate the Performance of MLE

In this sub-section, a simulation study is conducted to evaluate the performance of the MLE method of WLD in terms of both bias and standard deviation for various parameter combinations and different sample sizes. For each parameter combination, a random sample from WLD is generated by first generating a random sample $y_1, y_2, \cdots, y_n$ from the Weibull distribution with parameters c and $1/\beta$, then by using Remark 2.2(i), $X_i = \theta(e^{Y_i} - 1)$, $i = 1, 2, \ldots, n$ follows the $WLD(c, \beta, \theta)$.

The parameter combinations for the simulation study are $c = 0.5, 1, 4, 7$, $\beta = 0.5, 1, 3$, and $\theta = 0.5, 1, 3$. Two different sample sizes of $n = 100$ and 500 are used in the simulation. For each sample size and parameter combination, the simulation process is repeated 200 times. The average bias (estimate - actual), and the average standard deviation of the maximum likelihood estimates are presented in Tables 1 and 2.

The results show that the maximum likelihood estimation method performs well. In general, the biases and standard deviations of the parameters are reasonably small. As the sample size increases, the results show that the biases and standard deviations of the estimators decrease. The results from this simulation study, suggest that the maximum likelihood estimates method can be used effectively to estimate the parameters of the Weibull-Lomax distribution. Also, if $c > 1$, the results from Tables 1 and 2 show that the MLE method performs good in estimating the model parameters. This suggests that the WLD can be used an alternative to the WPD in modeling real life data sets.

3.3 Applications

Alzaatreh et al. (2013a) used three data sets from Park et al. (1964) and Park (1954) and fit them to Weibull-Pareto distribution using the MMLE method. In this subsection, we fit these data sets and show that WLD provides similar fit. The data sets are the grouped frequency distributions of adult numbers for Tribolium Confusum and Tribolium Castaneum cultured at 24 C and Tribolium Confusum strain. In particular, Data set 1 represents a random sample of 857 Tribolium

Table 1. Bias and standard deviation of the parameter estimates for $n = 100$

Actual Values			Bias			Standard deviation		
c	β	θ	$\hat{c}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{c}$	$\hat{\beta}$	$\hat{\theta}$
0.5	0.5	0.5	0.0287	−0.0001	0.0218	0.0748	0.1764	0.4158
		1	0.0225	0.0073	0.1774	0.0784	0.1920	0.9007
		3	0.0266	0.0092	0.3276	0.0772	0.1826	2.5553
	1	0.5	0.0232	−0.0011	0.0184	0.0683	0.3520	0.3326
		1	0.0218	0.0086	0.1598	0.0681	0.3814	0.9051
		3	0.0223	0.0165	0.3531	0.0626	0.3643	2.1206
	3	0.5	0.0206	0.0176	0.0258	0.0517	1.1344	0.2790
		1	0.0241	0.1880	0.1011	0.0497	1.1983	0.6267
		3	0.0154	0.1619	0.1682	0.0478	1.1155	1.6073
1	0.5	0.5	0.0836	0.0083	0.0680	0.2328	0.1924	0.4886
		1	0.0500	0.0227	0.2613	0.2244	0.1856	0.9557
		3	0.0773	0.0168	0.5588	0.2333	0.1861	2.9163
	1	0.5	0.0333	0.0658	0.0743	0.1542	0.3902	0.3491
		1	0.0472	0.0258	0.0940	0.1596	0.3618	0.6508
		3	0.0556	0.0495	0.3845	0.1691	0.3917	2.1340
	3	0.5	0.0408	0.0797	0.0193	0.1052	1.1522	0.2372
		1	0.0376	0.0614	0.0085	0.1022	1.1650	0.4817
		3	0.0256	0.1354	0.1569	0.0955	1.1701	1.4879
4	0.5	0.5	0.5229	−0.0268	−0.0256	1.0086	0.1698	0.4036
		1	0.3374	−0.0004	0.0775	0.8892	0.1671	0.8325
		3	0.6047	−0.0413	−0.3706	1.0022	0.1418	1.9807
	1	0.5	0.6234	−0.0505	−0.0246	0.9791	0.4286	0.3377
		1	0.4174	0.0260	0.0676	0.9472	0.4471	0.7082
		3	0.4091	0.0246	0.1803	1.0248	0.4609	2.1711
	3	0.5	0.1886	−0.1328	−0.0265	0.4220	1.1250	0.2178
		1	0.1209	0.1592	0.0614	0.4010	1.2174	0.4728
		3	0.1377	−0.0921	−0.1133	0.4029	1.1831	1.3794
7	0.5	0.5	0.4476	−0.0086	0.0007	1.3689	0.1239	0.2988
		1	0.4629	−0.0004	0.0396	1.3946	0.1250	0.6038
		3	0.6692	−0.0191	−0.1366	1.5165	0.1097	1.5466
	1	0.5	1.8592	−0.1978	−0.1348	1.7858	0.4377	0.3463
		1	1.6115	−0.1439	−0.1945	1.9261	0.4591	0.7135
		3	1.9645	−0.2476	−1.0595	1.7831	0.4022	1.8825
	3	0.5	0.2483	0.0982	0.0197	0.7125	1.1624	0.2273
		1	0.2000	0.0234	0.0083	0.6995	1.2274	0.4811
		3	0.2757	−0.0104	−0.0079	0.6927	1.2099	1.4260

Table 2. Bias and standard deviation of the parameter estimates for $n = 500$

Actual Values			Bias			Standard deviation		
c	β	θ	$\hat{c}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{c}$	$\hat{\beta}$	$\hat{\theta}$
0.5	0.5	0.5	0.0132	−0.0151	−0.0281	0.0348	0.0944	0.1861
		1	0.0116	−0.0030	−0.0058	0.0341	0.0974	0.3857
		3	0.0067	0.0026	0.0893	0.0372	0.0978	1.2534
	1	0.5	0.0079	0.0138	0.0189	0.0356	0.2224	0.2066
		1	0.0077	0.0057	0.0303	0.0359	0.2250	0.4078
		3	0.0138	−0.0107	0.0092	0.0384	0.2345	1.1882
	3	0.5	0.0057	0.1076	0.0213	0.0298	0.9001	0.1922
		1	0.0073	0.0660	0.0442	0.0321	0.8795	0.4162
		3	0.0060	0.0491	0.0766	0.0302	0.8460	1.1589
1	0.5	0.5	0.0098	0.0221	0.0816	0.1273	0.1266	0.3280
		1	0.0171	0.0143	0.1173	0.1333	0.1168	0.5645
		3	0.0253	0.0084	0.2387	0.1228	0.1140	1.7001
	1	0.5	0.0117	0.0198	0.0211	0.0962	0.2375	0.2007
		1	0.0200	0.0162	0.0540	0.1017	0.2642	0.4483
		3	0.0161	0.0260	0.1870	0.0987	0.2624	1.3323
	3	0.5	0.0179	0.1027	0.0238	0.0676	1.0732	0.2246
		1	0.0118	0.1047	0.0476	0.0602	0.9458	0.3995
		3	0.0112	0.1673	0.2464	0.0672	1.0295	1.3051
4	0.5	0.5	0.1038	0.0497	0.1590	0.9285	0.1857	0.4613
		1	0.0944	0.0491	0.3146	0.9194	0.1861	0.9304
		3	0.3544	−0.0079	0.0952	0.9766	0.1537	2.1699
	1	0.5	0.0156	0.1598	0.1371	0.6922	0.3962	0.3200
		1	0.0250	0.1420	0.2442	0.6449	0.3926	0.6303
		3	−0.0064	0.1548	0.7849	0.6285	0.3846	1.8555
	3	0.5	0.0956	0.0416	0.0097	0.2913	1.2129	0.2396
		1	0.0868	0.0659	0.0288	0.2875	1.2269	0.4841
		3	0.1223	0.0032	0.0094	0.3002	1.2342	1.4565
7	0.5	0.5	0.2962	0.0415	0.1375	1.2609	0.1857	0.4586
		1	−0.0383	0.0600	0.3516	1.1002	0.1743	0.8852
		3	−0.0157	0.0398	0.7004	1.1850	0.1412	2.1105
	1	0.5	0.5314	0.0559	0.0571	1.5608	0.4655	0.3683
		1	0.5067	0.0802	0.1527	1.5718	0.4696	0.7455
		3	0.5462	0.0624	0.3714	1.6205	0.4495	2.1272
	3	0.5	0.3919	−0.4290	−0.0834	0.5326	1.1271	0.2209
		1	0.0085	0.2958	0.1175	0.4168	1.0852	0.4266
		3	0.1736	0.0318	0.0439	0.5122	1.2105	1.4283

Table 3. Parameter estimates and K-S values for the fitted data sets

Data sets	Data set 1		Data set 2		Data set 3	
Distributions	*WPD*	*WLD*	*WPD*	*WLD*	*WPD*	*WLD*
Estimates	$\hat{\theta} = 9.107$	$\hat{\theta} = 14.245$	$\hat{\theta} = 33.208$	$\hat{\theta} = 22.122$	$\hat{\theta} = 23.283$	$\hat{\theta} = 12.627$
	$\hat{\beta} = 0.380$	$\hat{\beta} = 0.436$	$\hat{\beta} = 0.747$	$\hat{\beta} = 0.521$	$\hat{\beta} = 0.866$	$\hat{\beta} = 0.517$
	$\hat{c} = 6.953$	$\hat{c} = 6.863$	$\hat{c} = 5.059$	$\hat{c} = 8.719$	$\hat{c} = 6.161$	$\hat{c} = 12.398$
K-S values	0.0482	0.0549	0.0730	0.0905	0.0803	0.0901
AIC values	8904.5	8908.6	9276.41	9311.5	2973.44	2982.1

Castaneum cultured at 24^0C, the Data set 2 represents a random sample of 952 Tribolium Castaneum cultured at 24^0C and Data set 3 represents a random sample of 368 Tribolium Castaneum cultured at 24^0C. The data sets are avialable in Alzaatreh, et al. (2013a). Above we provide the estimates, the Kolmogorov-Smirnov test (K-S) and the Akaike information criterion (AIC) values for the WLD and WPD in Table 3. The estimates of WLD are based on MLE method and the estimates of WPD are based on MMLE method (taken from Alzaatreh et al., 2013a). The results in Table 3 show that WLD and WPD provide similar fits to three data sets. Figure 5 supports the results in Table 3.

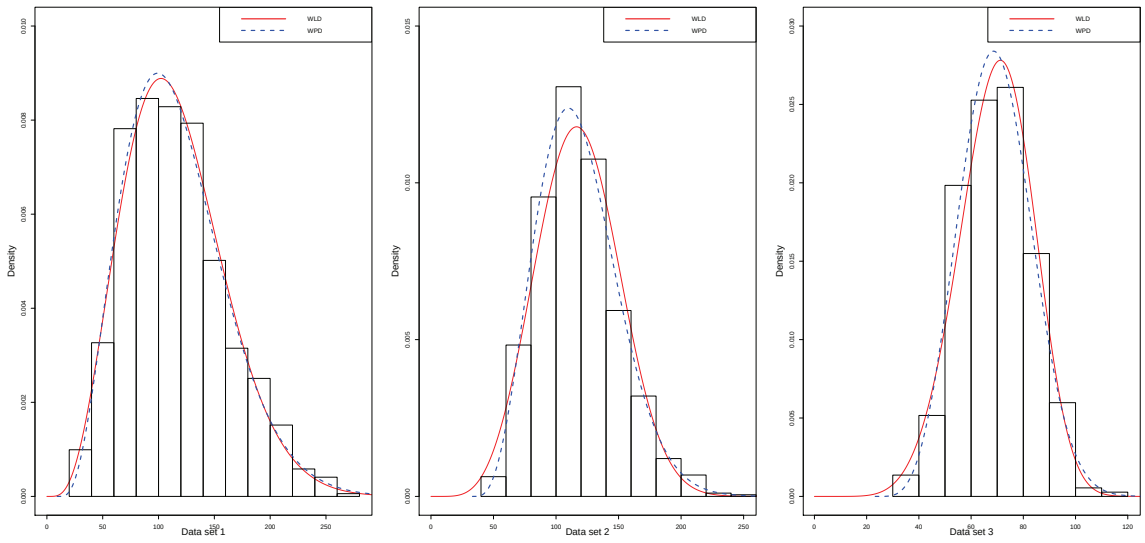


Figure 5. Histograms and fitted PDFs for WLD and WPD.

4. Conclusion

In this paper, a shift by a parameter $\theta > 0$ of the Weibull-Pareto distribution defined by Alzaatreh et al. (2013a) is proposed namely, the Weibull-Lomax distribution. Based on Alzaatreh et al. (2013a), the maximum likelihood estimation produces high biases and standard errors for the WPD parameters and therefore, they proposed a modification of the MLE, MMLE, which can be used only when $c > 1$. This problem in estimatioing the WPD parameters has motivated us to present an alternative to the WPD by using Lomax distribution in replace of Pareto distribution in the $T - X$ family proposed by Alzaatreh et al. (2013b). In this paper we showed that the MLE method can be used effectively to estimate the WLD parameters without any restriction on the parameter c. The results of the simulation study in section 3.2 are compared with the results of the simulation study for the WPD obtained by Alzaatreh et al. (2013a). In most cases, it was observed that the biases and standard errors for the MLEs of WLD parameters are smaller than the MMLEs for the WPD parameters. Furthermore, WLD is applied to the same data sets used in Alzaatreh et al. (2013a). The results in Table 3 and Figure 5 showed that WLD and WPD provide similar fit to the data sets.

Acknowledgments

The authors are grateful for the comments and suggestions by the referees and the Editor-in-Chief. Their comments and suggestions have greatly improved the paper.

References

Alzaatreh, A., Famoye, F. and Lee, C. (2013a). Weibull-Pareto distribution and its applications, *Communications in Statistics:Theory and Methods*, 42, 1673-1691. http://dx.doi.org/10.1080/03610926.2011.599002.

Alzaatreh, A. and Ghosh, I. (2014). On the Weibull-X family of distributions, *Journal of Statistical Theory and Applications*, 14, 169-183.

Alzaatreh, A., Lee, C. and Famoye, F. (2013b). A new method for generating families of continuous distributions, *Metron*, 71, 63-79. http://dx.doi.org/10.1007/s40300-013-0007-y.

Arnold, B.C., Balakrishnan, N. and Nagarajah, H.N. (2008). *A first course in Order Statistics*, John Wiley, New York. http://dx.doi.org/10.1137/1.9780898719062 .

Feller, W. (1971). *An introduction to probability theory and its applications*, John Wiley, N.Y.

Galton, F. (1883). *Enquiries into Human Faculty and its Development*, Macmillan and Company, London.

Hall, I.J. (1984). Approximate one-sided tolerance limits for the difference or sum of two independent normal variates, *Journal of Qualitative Technology*, 16, 15-19.

Johnson, N. L., Kotz, S., Balakrishnan, N. (1994). *Continuous Univariate Distributions*, 2nd ed. Vol. 1. New York: John Wiley & Sons.

Moors, J.J. (1988). A quantile alternative for kurtosis, *Statistician*, 37, 25C32. http://dx.doi.org/10.2307/2348376 .

Park, T. (1954). Experimental studies of interspecies competition II. Temperature, humidity, and competition in two species of Tribolium, *Physiolog. Zool.*, 27, 177-238. http://dx.doi.org/10.1086/physzool.27.3.30152164.

Park, T., Leslie, P. H, Mertz, D. B. (1964). Genetic strains and competition in population of Tribolium, *Physiolog. Zool.*, 37, 97-162. http://dx.doi.org/10.1086/physzool.37.2.30152328.

Shannon, C. E. (1948). A mathematical theory of communication, *Bell Syst. Tech. J.*, 27, 379-432.

von Mises, R. (1936). La distribution de la plus grande de n valeurs, *Review of Mathematics Union Interbalcanique*, 1, 141-160.

Weerahandi, S. and Johnson, R.A. (1992). Testing reliability in a stress-strength model when X and Y are normally distributed, *Technometrics*, 38, 83-91. http://dx.doi.org/10.2307/1269555.

Appendix R outputs for the regression line in Figure 4.

```
Call:
lm(formula = log(b) ~ log(c) + I(log(c)^2) + I(log(c)^3))

Residuals:
      Min        1Q    Median        3Q       Max
-0.089138 -0.017417 -0.003819  0.019843  0.095153

Coefficients:
             Estimate   Std. Error t value  Pr(>|t|)
(Intercept)   9.16963    0.36639   25.03    < 2e-16 ***
log(c)       -8.50759    0.45294  -18.78    < 2e-16 ***
(log(c))^2    2.42676    0.17909   13.55    8.36e-15 ***
(log(c))^3   -0.26426    0.02279  -11.59    5.41e-13 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1

Residual standard error: 0.03302 on 32 degrees of freedom
Multiple R-squared:  0.9988,    Adjusted R-squared:  0.9987
F-statistic:  8651 on 3 and 32 DF,  p-value: < 2.2e-16
```

Copyrights

International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Estimation of Smooth Functions via Convex Programs

Eunji Lim[1] & Mina Attallah[1]

[1] College of Business and Public Administration, Kean University, Union, NJ 07083, USA

Correspondence: Eunji Lim, College of Business and Public Administration, Kean University, Union, NJ 07083, USA. Tel: 1-908-737-4709. E-mail: elim@kean.edu

Received: May 17, 2016 Accepted: June 6, 2016 Online Published: June 28, 2016

doi:10.5539/ijsp.v5n4p150 URL: http://dx.doi.org/10.5539/ijsp.v5n4p150

Abstract

One of the numerically preferred methods for fitting a function to noisy data when the underlying function is known to be smooth is to minimize the roughness of the fit while placing a limit on the sum of squared errors. We show that the fit can be formulated as a solution to a convex program. Since convex programs can be solved by various methods with guaranteed convergence, our formulation enables one to use these methods to compute the fit numerically. Numerical results show that our formulation is successfully applied to the problem of sensitivity estimation of option prices as functions of the underlying stock price.

Keywords: nonparametric regression, smoothing spline, convex programming, penalized least squares

1. Introduction

We consider the problem of fitting a function $g : [a, b) \to \mathbb{R}$ to noisy data when the underlying function $f_* : [a, b) \to \mathbb{R}$ is assumed to satisfy a certain smoothness condition. We presume that we observe m noisy measurements $Y_{i1}, \ldots, Y_{im}$ of f_* at each $x_i \in [a, b)$ for $1 \le i \le n$, and that

$$Y_{ij} = f_*(x_i) + \epsilon_{ij},$$

for $1 \le i \le n$ and $1 \le j \le m$, where $\epsilon_{i1}, \ldots, \epsilon_{im}$ are independent and identically distributed (iid) with a mean of 0 and a variance of $\sigma_i^2 < \infty$. We are particularly interested in the case where the underlying function f_* is known to be k times differentiable (for $k \ge 2$) and its kth derivative is square integrable. This situation naturally arises when the $(k-2)$th derivative of f_* is observed to be smooth, and hence, one tries to find the fit g by minimizing the roughness of the $(k-2)$th derivative of the fit g. Since the "roughness" of a function $f : [a, b) \to \mathbb{R}$ is measured by $\int_a^b \left\{f^{(2)}(x)\right\}^2 dx$, the roughness of the $(k-2)$th derivative of g is measured by $\int_a^b \left\{g^{(k)}(x)\right\}^2 dx$. While minimizing the roughness $\int_a^b \left\{g^{(k)}(x)\right\}^2 dx$, one needs to ensure that the fitted function is close enough to the estimated values of f_*. This can be done by imposing a limit on the sum of squared distances between the fitted function and the estimates of f_* as follows:

$$\sum_{i=1}^{n} (\overline{Y}_i - g(x_i))^2/n \le S$$

for some positive constant $S > 0$, where $\overline{Y}_i = \sum_{j=1}^{m} Y_{ij}/m$ for $1 \le i \le n$. This leads to the following formulation for computing the fit:

$$\text{Minimize } \int_a^b \left\{g^{(k)}(x)\right\}^2 dx \text{ subject to } \sum_{i=1}^{n} \left(\overline{Y}_i - g(x_i)\right)^2/n \le S \tag{1}$$

over $g \in \mathcal{D}^k$, where

$$\mathcal{D}^k = \left\{ f : [a, b) \to \mathbb{R} : f \text{ is } k \text{ times differentiable and } \int_a^b \left\{f^{(k)}(x)\right\}^2 dx < \infty \right\}.$$

Formulation (1) was introduced by Schoenberg (1964) and was studied extensively in the numerical analysis community by a number of authors including Reinsch (1967), Reinsch (1971), Wahba (1975), and Gander (1980). Formulation (1) is often contrasted, in the statistics literature, with the following formulation:

$$\text{Minimize } \sum_{i=1}^{n} \left(\overline{Y}_i - g(x_i)\right)^2/n + \lambda_n \int_a^b \left\{g^{(k)}(x)\right\}^2 dx \tag{2}$$

over $g \in \mathcal{D}^k$ for a sequence of nonnegative real numbers $(\lambda_n : n \geq 1)$. In Formulation (2), the term $\int_a^b \left\{g^{(k)}(x)\right\}^2 dx$ measures the roughness of the fit g and the term $\sum_{i=1}^n \left(\overline{Y}_i - g(x_i)\right)^2 /n$ measures the sum of squared errors. The parameter λ_n controls the trade-off between the roughness and the goodness-of-fit.

One of the advantages of Formulation (1) over Formulation (2) is that a good estimate of S is easily provided in Formulation (1), while the performance of Formulation (2) is highly sensitive to the choice of $(\lambda_n : n \geq 1)$ and selecting $(\lambda_n : n \geq 1)$ is not straightforward. For this reason, Formulation (1) is preferred in the numerical analysis community. A good estimate of S in Formulation (1) can be obtained using the laws of large numbers as follows. By the weak law of large numbers, $\sum_{j=1}^m \epsilon_{ij}/\sqrt{m}$ converges in distribution to $N(0, \sigma_i^2)$ as m increases to infinity, where $N(\mu, \sigma^2)$ denotes a normal random variable with a mean of μ and a variance of σ^2. Hence, $\overline{\epsilon}_i^2$, where $\overline{\epsilon}_i = (1/m)\sum_{j=1}^m \epsilon_{ij}$, can be approximated by $(N(0, \sigma_i^2))^2/m$ for m sufficiently large. By applying the strong law of large numbers in n, $\sum_{i=1}^n (N(0, \sigma_i^2))^2/n$ can be approximated as $\sum_{i=1}^n \sigma_i^2/n$ for n sufficiently large. Therefore, the following approximation is possible

$$\frac{1}{n}\sum_{i=1}^n \left(\overline{Y}_i - f_*(x_i)\right)^2 = \frac{1}{n}\sum_{i=1}^n \overline{\epsilon}_i^2 \approx \frac{1}{nm}\sum_{i=1}^n (N(0, \sigma_i^2))^2 \approx \frac{1}{nm}\sum_{i=1}^n \sigma_i^2$$

for n and m sufficiently large. The symbol $\approx$ is used to informally express "approximate equality." In practice, σ_i^2 is estimated by the sample variance S_i^2 of $Y_{i1}, \ldots, Y_{im}$ for $1 \leq i \leq n$; i.e., $S_i^2 = \sum_{j=1}^m \left(\epsilon_{ij} - \overline{\epsilon}_i\right)^2 /(m-1)$. A good estimate of S is thus $\sum_{i=1}^n S_i^2/(nm)$.

Despite its practical importance, there exist few numerical procedures that compute the solution of Formulation (1) with guaranteed convergence. Traditionally, the solution of (1) is computed as follows. For each $S > 0$, there is a unique $\lambda_n = \lambda_n(S)$ such that the solution of (2) for this λ_n is the solution of (1). Furthermore, the solution of (2) for this λ_n, denoted by g_{λ_n}, satisfies $\sum_{i=1}^n \left(\overline{Y}_i - g_{\lambda_n}(x_i)\right)^2 /n = S$. Since (2) can be solved by solving a set of linear systems (pages 410–412 of Györfi et al., 2002), $\lambda_n(S)$ can be computed iteratively by means of the Newton procedure starting from an initial guess of $\lambda_n(S)$. This procedure does not guarantee global convergence to the solution of (1); see Reinsch (1971).

In this paper, we show that (1) can be reformulated as a convex program (Proposition 1). Convex programs can be solved using various methods that guarantee global convergence to the solution; see the Lagrangian method on page 217 of Zangwill (1969) for an example of methods that solve convex programs. Our formulation thus enables one to compute the solution of (1) with guaranteed convergence by using those methods and powerful software packages that are designed to solve convex programs.

This paper is organized as follows. Section 2 describes the proposed formulation in detail. Numerical results in Section 3 illustrate that our formulation is successfully applied to the problem of sensitivity estimation of option prices as functions of the underlying stock price. Concluding remarks are included in Section 4.

1. Proposed Formulation

In this section, we describe how Formulation (1) can be reformulated as a convex program. We first present some preliminary results.

A spline function with degree $r > 1$ with knots $x_1, \ldots, x_n$, where $a < x_1 < \ldots, < x_n < b$, is a function $s : [a, b) \to \mathbb{R}$ having the following two properties: (a) In each of the intervals $[a, x_1), [x_1, x_2), \ldots, [x_{n-1}, x_n), [x_n, b)$, $s(x)$ is given by some polynomial of degree r or less, and (b) $s(x)$ is $r-1$ times continuously differentiable on $[a, b)$. We denote the set of spline functions with degree r by $\mathcal{S}_r([a, b))$. Even though $\mathcal{S}_r([a, b))$ seems to be infinite dimensional, it turns out to be finite dimensional with the dimension equal to $r + n + 1$. We describe one of the bases for $\mathcal{S}_r([a, b))$, which is the set of B-splines; the B-splines are preferred in numerical studies since they have bounded supports, and hence, produce well–conditioned numerical settings. We introduce additional knots $x_{-r}, \ldots, x_0, x_{n+1}, \ldots, x_{n+r+1}$ so that

$$x_{-r} < x_{-r+1} < \cdots < x_0 < a < x_1 < \cdots < x_n < b < x_{n+1} < \cdots < x_{n+r+1}.$$

The B-spline $B_{i,r}$ of degree r is defined recursively by

$$B_{i,0}(x) = \begin{cases} 1, & \text{if } x_i \leq x < x_{i+1} \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

for $i = -r, \ldots, n + r$ and $x \in \mathbb{R}$ and

$$B_{i,l}(x) = \frac{x - x_i}{x_{i+l} - x_i} B_{i,l-1}(x) + \frac{x_{i+l+1} - x}{x_{i+l+1} - x_{i+1}} B_{i+1,l-1}(x) \tag{4}$$

for $i = -r, \ldots, n+r-l$, $l = 1, \ldots, r$, and $x \in \mathbb{R}$. By Theorem 14.1 on page 262 of Györfi et al. (2002), $\{B_{i,r} : i = -r, \ldots, n\}$ restricted to $[a, b)$ is a basis of $\mathcal{S}_r([a, b))$. We are ready to present the main result of this paper.

Proposition 1 *Assume $2 \le k \le n$. Consider the minimization problem*

$$\begin{aligned} &\textit{minimize} && \int_a^b \left\{g^{(k)}(x)\right\}^2 dx \\ &\textit{subject to} && \frac{1}{n}\sum_{i=1}^{n}\left(\overline{Y}_i - g(x_i)\right)^2 \le S \end{aligned} \tag{5}$$

over $g \in \mathcal{D}^k$ for some constant $S > 0$. Then, there exists a solution $\hat{g}_n \in \mathcal{D}^k$ to Problem (5). Furthermore, $\hat{g}_n$ can be represented as

$$\hat{g}_n(x) = \sum_{i=-(2k-1)}^{n} \hat{c}_i B_{i,2k-1}(x)$$

for $x \in [a, b)$, where the $\hat{c}_i$s are the solution to the following convex program in the decision variables $c_{-(2k-1)}, \ldots, c_n, y_1, \ldots, y_n \in \mathbb{R}$.

$$\begin{aligned} &\textit{minimize} && \int_a^b \left(\sum_{i=-(2k-1)}^{n} c_i B_{i,2k-1}^{(k)}(x)\right)^2 dx = \sum_{i=-(2k-1)}^{n}\sum_{j=-(2k-1)}^{n} c_i c_j \int_a^b B_{i,2k-1}^{(k)}(x) B_{j,2k-1}^{(k)}(x) dx \\ &\textit{subject to} && \sum_{i=1}^{n}\left(\overline{Y}_i - y_i\right)^2 / n \le S, \\ & && \sum_{i=-(2k-1)}^{n} c_i B_{i,2k-1}(x_j) = y_j,\ j = 1, \ldots, n. \end{aligned} \tag{6}$$

Proof. We let $C = \left\{(y_1, \ldots, y_n) \in \mathbb{R}^n : \sum_{i=1}^{n}\left(\overline{Y}_i - y_i\right)^2 / n \le S\right\}$. It should be noted that C is a nonempty, closed and bounded subset of $\mathbb{R}^n$.

By Lemmas 20.2 and 20.3 on pages 415 and 416 of Györfi et al. (2002), for any $y = (y_1, \ldots, y_n)$ in C, there exists a unique spline function h of degree $2k-1$ satisfying $h(x_i) = y_i$ for $i = 1, \ldots, n$ and $h^{(l)}(a) = h^{(l)}(b) = 0$ for $l = k, \ldots, 2k-1$. Furthermore, $\int_a^b \left\{h^{(k)}(x)\right\}^2 dx \le \int_a^b \left\{g^{(k)}(x)\right\}^2 dx$ for any $g \in \mathcal{D}^k$ satisfying $g(x_i) = y_i$ for $1 \le i \le n$. Since $(B_{i,2k-1} : -(2k-1) \le i \le n)$ is a basis of $\mathcal{S}_{2k-1}([a, b))$ (Theorem 14.1 on page 262 of Györfi et al., 2002), there exists a unique $c_{-(2k-1)}, \ldots, c_n$ such that $h(x) = \sum_{i=-(2k-1)}^{n} c_i B_{i,2k-1}(x)$ for $x \in [a, b)$. The c_is are determined by the following linear system with $n + 2k$ variables and $n + 2k$ linear equations:

$$\begin{aligned} \sum_{i=-(2k-1)}^{n} c_i B_{i,2k-1}(x_j) &= y_j \\ \sum_{i=-(2k-1)}^{n} c_i B_{i,2k-1}^{(l)}(a) &= 0 \\ \sum_{i=-(2k-1)}^{n} c_i B_{i,2k-1}^{(l)}(b) &= 0 \end{aligned} \tag{7}$$

for $j = 1, \ldots, n$ and $l = k, \ldots, 2k-1$. By the uniqueness of $c_{-(2k-1)}, \ldots, c_n$, the linear system (7) is nonsingular, and hence, the map from y in $\mathcal{D}^k$ to $\int_a^b \left\{h^{(k)}(x)\right\}^2 dx$ is continuous. Hence, there exists $\hat{y} = (\hat{y}_1, \ldots, \hat{y}_n) \in C$ that minimizes $\int_a^b \left\{g^{(k)}(x)\right\}^2 dx$ over $g \in \mathcal{D}^k$. Let $\hat{h}$ be the unique spline function satisfying $\hat{h}(x_i) = \hat{y}_i$ for $i = 1, \ldots, n$ and $\hat{h}^{(l)}(a) = \hat{h}^{(l)}(b) = 0$ for $l = k, \ldots, 2k-1$. Let $\hat{c}_{-(2k-1)}, \ldots, \hat{c}_n$ be the solution to (7) when $(y_1, \ldots, y_n) = (\hat{y}_1, \ldots, \hat{y}_n)$. Obviously, $\hat{h}$ is a feasible solution to (6). For any feasible solution $c_{-(2k-1)}, \ldots, c_n, y_1, \ldots, y_n$ of (6), $\sum_{i=-(2k-1)}^{n} c_i B_{i,2k-1}(x)$ belongs to $\mathcal{D}^k$, and hence, $\int_a^b \left(\sum_{i=-(2k-1)}^{n} c_i B_{i,2k-1}^{(k)}(x)\right)^2 dx$ is less than or equal to $\int_a^b \left\{\hat{h}^{(k)}(x)\right\}^2 dx$. Thus, the $\hat{c}_i$s and the $\hat{y}_i$s are the solution to (6). □

We close this section by describing how to evaluate the B-splines and the integration of the product of their kth derivatives that appear in (6). The B-splines can be evaluated recursively through Equations (3) and (4). The kth derivative of the B-spline can be evaluated recursively through the following relation: for a B-spline of degree r,

$$dB_{i,r}(x)/dx = (r/(x_{i+r} - x_i))B_{i,r-1}(x) - (r/(x_{i+r+1} - x_{i+1}))B_{i+1,r-1}(x) \tag{8}$$

for $i = -r, \ldots, n$ and $x \in [a, b)$; see Lemma 14.6 on page 265 of Györfi et al. (2002).

There are a couple of ways to compute the integration in (6). First, $\int_a^b B_{i,2k-1}^{(k)}(x) B_{j,2k-1}^{(k)}(x)dx$ can be computed by evaluating $B_{i,2k-1}^{(k)}$ using the recursion in (8) and by numerically evaluating the integration. Second, one can use the closed form formula for the integration of the product of the kth derivatives of the B-splines given by Equation (7) on page 1026 of Vermeulen et al. (1992) and the closed form formulas for the B-splines (Equation (1.20) on page 8 of Dierckx, 1993) to directly compute the kth derivatives and the integration of their products.

3. Numerical Results

3.1 A Motivation from Finance

This paper is motivated by the need to estimate the price of a stock option and its derivatives as functions of the underlying stock price. The first and second derivatives of an option price play an important role when financial institutions manage a portfolio of stocks and stock options in an attempt to hedge the risks associated with the portfolio. For example, consider a call option that gives the holder of the option the right to buy the underlying stock by a certain date for a certain price. Since the price of such a call option depends on the underlying stock price, the option price can be denoted by $f_*(x)$, where x is the underlying stock price per share. The delta (Δ) of the option is defined by the first derivative df_*/dx of the option price with respect to the underlying stock price. It is well known that a portfolio consisting of a short position of the call option and a long position of Δ shares of the underlying stock is expected to grow at a risk-free interest rate. Since the value of delta changes as the underlying stock price changes over time, the number of shares of the underlying stock in the portfolio must be changed periodically to stay in the risk-free position. This step is called rebalancing. When rebalancing a portfolio, the gamma (Γ) of the option, which is the second derivative $d^2 f_*/dx^2$ of the option price with respect to the underlying stock price, is used since the value of gamma tells us how much delta changes, and hence, how many shares of the underlying stock should be sold or bought in order to stay in a delta neutral position.

Recently, financial institutions have been issuing stock options with much more complex payoff structures than that of a call option. For such options, the option price cannot be expressed in a closed-form formula, and hence, one needs to use simulation to estimate the option price. Simulation of the option price consumes a significant amount of time. In order to facilitate quick decisions, traders in financial institutions conduct simulations before they actually need to rebalance a portfolio. Since the underlying stock price in the future cannot be predicted accurately, the traders conduct simulations for all possible underlying stock prices on the day when rebalancing takes place. The question thus boils down to how to estimate the option price f_* and its first and second derivatives, df_*/dx and $d^2 f_*/dx^2$, over a range $[a, b)$ of possible underlying stock prices using simulation.

One simple strategy for estimating $f_*, df_*/dx$ and $d^2 f_*/dx^2$ for $x \in [a, b)$ is to choose various possible values for the stock price, say $x_1, \ldots, x_n$, from $[a, b)$, estimate the option price at each x_i for $1 \le i \le n$ using simulation, use finite differences of the estimated option prices to estimate delta, and use finite differences of the estimated delta values to estimate gamma. A serious drawback of this approach is that the estimated delta and gamma values often lead to noisy curves as functions of the underlying stock price. It is especially frustrating for traders to see gamma values fluctuating around zero because positive gamma values suggest purchasing additional shares of the underlying stock, while negative gamma values suggest selling some of the shares. Figure 1 shows an example of the estimated gamma values plotted against the underlying stock price. The stock prices S_1, S_2, and S_3 are close to one another, but the graph suggests different strategies because the gamma values are positive, negative, and positive at S_1, S_2, and S_3, respectively. This degree of randomness in the gamma curve is not acceptable in practice.

To overcome this drawback, we propose fitting a curve $g : [a, b) \to \mathbb{R}$ to the estimated values of f_* so that the fitted curve has a smooth second derivative. Since the "roughness" of a function $f : [a, b) \to \mathbb{R}$ is measured by $\int_a^b \left\{f^{(2)}(x)\right\}^2 dx$, the roughness of the second derivative of g is measured by $\int_a^b \left\{g^{(4)}(x)\right\}^2 dx$. While minimizing $\int_a^b \left\{g^{(4)}(x)\right\}^2 dx$, we want to make sure that the fitted function is close to the estimated values of f_* by placing a limit on the sum of squared distance between the fitted values and the estimated values. This leads to the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & \int_a^b \left\{g^{(4)}(x)\right\}^2 dx \\ \text{subject to} \quad & \frac{1}{n}\sum_{i=1}^{n} \left(\overline{Y}_i - g(x_i)\right)^2 \le S \end{aligned}$$

over $g \in \mathcal{D}^4$, where Y_{ij} is the jth replication of the estimated value of the option price at x_i for $1 \le i \le n$ and $1 \le j \le m$. The above formulation is a special case of Formulation (1) with $k = 4$.

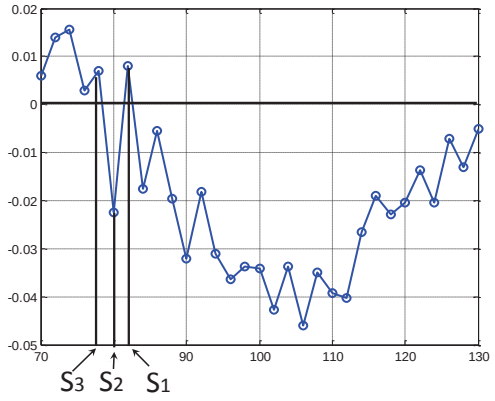


Figure 1. The horizontal axis is the underlying stock price, and the vertical axis is gamma.

3.2 Applying Our Formulation to Sensitivity Estimation of Option Prices

We consider the case where $f_*(x)$ is the expected payoff of a certain equity-linked security (ELS) when the underlying stock price is denoted by x. The payoff function of the ELS has the following structure. Suppose that the ELS is issued at time 0 and matures at time T. We denote the underlying stock price at time $t \in [0, T]$ by S_t. There are q days when early redemption is possible. On each of those days d_i for $1 \le i \le q$, the ELS expires with a payoff of $\$r_i$ if S_{d_i}/S_0 exceeds some threshold b_i. Otherwise, the ELS does not expire until maturity. If there is no early redemption and S_t/S_0 does not drop below a limit b until maturity, then the ELS expires with a payoff of \$1 at maturity. Otherwise, the ELS expires with a payoff of $\$S_T/S_0$ at maturity.

We let $a = 90, b = 110$, and $x_i = 90 + (20)(i/n) - (10/n)$ for $1 \le i \le n$. For each x_i, a sample path of a geometric Brownian motion is generated as a trajectory of the stock price between now and maturity, and the corresponding payoff of the ELS is computed. Y_{ij} is the payoff computed this way in the jth replication of the geometric Brownian motion at x_i. The parameters used for the experiment are $T = 365$ days, $q = 6, d_1 = 61, d_2 = 122, d_3 = 182, d_4 = 243, d_5 = 304, d_6 = 365, b_1 = 0.9, b_2 = 0.9, b_3 = 0.85, b_4 = 0.85, b_5 = 0.8, b_6 = 0.8, r_1 = 1.05, r_2 = 1.10, r_3 = 1.15, r_4 = 1.20, r_5 = 1.25, r_6 = 1.30$, and $b = 0.7$. The remaining time until maturity is 60 days, the annual volatility is 30%, the annual risk–free interest rate is 5%, and the initial stock price at time 0 is \$125.

We set $m = 50$, so 50 sample paths for the geometric Brownian motion are generated at each x_i to compute $Y_{i1}, \ldots, Y_{i50}$ for $1 \le i \le n$. We compute $\overline{Y}_i = \sum_{j=1}^{50} Y_{ij}/50$ for $1 \le i \le n$ and use $(x_1, \overline{Y}_1), \ldots, (x_n, \overline{Y}_n)$ to compute the proposed estimator $\hat{g}_n$ by solving (6) with CVX, a package for specifying and solving convex programs (Grant & Boyd, 2014). The constant S in Formulation (6) is replaced with $\sum_{i=1}^{n} S_i^2$, where S_i^2 is the sample variance of $Y_{i1}, \ldots, Y_{im}$ for $1 \le i \le n$.

To measure the accuracy of the proposed estimator, we compute the following integrated mean square error (IMSE) between the underlying function f_* and $\hat{g}_n$: $\sum_{i=1}^{n} (\hat{g}_n(x_i) - f_*(x_i))^2 / n$, where $f_*(x_i)$ is estimated from the average of 400,000 iid replications of Y_{ij} at each x_i. Table 1 reports the averages and the standard deviation of the IMSE, computed based on 200 iid replications of $\hat{g}_n$, for a variety of n values. The IMSE decreases as n increases, which displays the convergence of $\hat{g}_n$ to f_* as $n \to \infty$.

Table 1. The averages and the standard deviation of the IMSE when $f_*(x)$ is the price of the ELS.

	n	5	10	20	40	80	160
IMSE	Average (10^{-4})	8.6	4.4	2.1	1.1	0.7	0.6
	Standard deviation (10^{-4})	6.1	2.9	1.5	0.8	0.5	0.3

4. Concluding Remarks

In this paper, we study a numerically preferred formulation for fitting a smooth function to noisy data. Numerical results illustrate that our formulation successfully computes the fit. They also suggest that the fit converges to the true function as the number of observations in the data set increases to infinity. Future research topics include studies on asymptotic

properties of the fit when the number of observations in the data set increases to infinity.

References

Dierckx, P. (1993). *Curve and surface fitting with splines.* New York: Oxford University Press Inc. http://dx.doi.org/10.5860/choice.31-2162

Gander, W. (1980). Least squares with a quadratic constraint. *Numer. Math., 36*, 291-307. http://dx.doi.org/10.1007/bf01396656

Grant, M., & Boyd, S. (2014). CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx, March 2014.

Györfi, L., Kohler, M, Krzyżak, A., & Walk, H. (2002). *A distribution-free theory of nonparametric regression.* New York: Springer. http://dx.doi.org/10.1007/b97848

Reinsch, C. H. (1967). Smoothing by spline functions. *Numer. Math., 10*, 177-183. http://dx.doi.org/10.1007/bf02162161

Reinsch, C. H. (1971). Smoothing by spline functions II. *Numer. Math., 16*, 451-454. http://dx.doi.org/10.1007/bf02169154

Schoenberg, I. J. (1964). Spline functions and the problem of graduation. *Proc. Nat. Acad. Sci. (USA), 52*, 947-950. http://dx.doi.org/10.1073/pnas.52.4.947

Vermeulen, I. J., Bartels, R. H., & Heppler, G. R. (1992). Integrating products of B-splines. *Siam J. Sci. Stat. Comput., 13*, 1025-1038. http://dx.doi.org/10.1137/0913060

Wahba, G. (1975). Smoothing noisy data with spline functions. *Numer. Math., 24*, 383-393. http://dx.doi.org/10.1007/bf01437407

Zangwill, W. I. (1969). *Nonlinear programming: a unified approach.* New Jersey: Prentice-Hall Inc.

Copyrights

International Journal of Statistics and Probability; Vol. 5, No. 4; July 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Reviewer Acknowledgements

International Journal of Statistics and Probability wishes to acknowledge the following individuals for their assistance with peer review of manuscripts for this issue. Their help and contributions in maintaining the quality of the journal is greatly appreciated.

Many authors, regardless of whether *International Journal of Statistics and Probability* publishes their work, appreciate the helpful feedback provided by the reviewers.

Reviewers for Volume 5, Number 4

Abdullah A. SMADI

Afsin Sahin

Anwar H Joarder

Bibi Abdelouahab

Chin-Shang Li

Encarnación Alvarez-Verdejo

Farida Kachapova

Gabriel A. Okyere

Gane Samb Lo

Hongsheng Dai

Jiannan Lu

Marcelo Bourguignon

Mirko D'Ovidio

Nicolas MARIE

Philip Westgate

Vyacheslav Abramov

Wei Zhang

Wojciech Gamrot

Yimei Li

Yulei Pang

Wendy Smith

On behalf of,

The Editorial Board of *International Journal of Statistics and Probability*

Canadian Center of Science and Education

Call for Manuscripts

International Journal of Statistics and Probability is a peer-reviewed journal, published by Canadian Center of Science and Education. The journal publishes research papers in all areas of statistics and probability. The journal is available in electronic form in conjunction with its print edition. All articles and issues are available for free download online.

We are seeking submissions for forthcoming issues. All manuscripts should be written in English. Manuscripts from 3000–8000 words in length are preferred. All manuscripts should be prepared in LaTeX or MS-Word format, and submitted online, or sent to: ijsp@ccsenet.org

Paper Selection and Publishing Process

a) Upon receipt of a submission, the editor sends an e-mail of confirmation to the submission's author within one to three working days. If you fail to receive this confirmation, your submission e-mail may have been missed.

b) Peer review. We use a double-blind system for peer review; both reviewers' and authors' identities remain anonymous. The paper will be reviewed by at least two experts: one editorial staff member and at least one external reviewer. The review process may take two to three weeks.

c) Notification of the result of review by e-mail.

d) If the submission is accepted, the authors revise paper and pay the publication fee.

e) After publication, the corresponding author will receive two hard copies of the journal, free of charge. If you want to keep more copies, please contact the editor before making an order.

f) A PDF version of the journal is available for download on the journal's website, free of charge.

Requirements and Copyrights

Submission of an article implies that the work described has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the authorities responsible where the work was carried out, and that, if accepted, the article will not be published elsewhere in the same form, in English or in any other language, without the written consent of the publisher. The editors reserve the right to edit or otherwise alter all contributions, but authors will receive proofs for approval before publication.

Copyrights for articles are retained by the authors, with first publication rights granted to the journal. The journal/publisher is not responsible for subsequent uses of the work. It is the author's responsibility to bring an infringement action if so desired by the author.

More Information

E-mail: ijsp@ccsenet.org

Website: www.ccsenet.org/ijsp

Paper Submission Guide: www.ccsenet.org/submission

Recruitment for Reviewers: www.ccsenet.org/reviewer

The journal is peer-reviewed
The journal is open-access to the full text
The journal is included in:

BASE
EBSCOhost
Google Scholar
JournalTOCs
Library and Archives Canada
LOCKSS
PKP Open Archives Harvester
ProQuest
SHERPA/RoMEO
Standard Periodical Directory
Ulrich's

International Journal of Statistics and Probability

Bimonthly

Publisher	Canadian Center of Science and Education
Address	1120 Finch Avenue West, Suite 701-309, Toronto, ON., M3J 3H7, Canada
Telephone	1-416-642-2606
Fax	1-416-642-2608
E-mail	ijsp@ccsenet.org
Website	www.ccsenet.org/ijsp

ISSN 1927-7032