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

International Journal of Statistics and Probability

Vol. 1, No. 1 May 2012

CCSE

CANADIAN CENTER OF SCIENCE AND EDUCATION

Editorial Board

Contents

Improved Liu Estimators for the Poisson Regression Model

Kristofer Mansson

Department of Economics, Finance and Statistics, Jonkoping University

Jonkoping, Sweden

B. M. Golam Kibria (Corresponding author)

Department of Mathematics and Statistics, Florida International University

Miami, Florida, USA

Par Sjolander

Department of Economics, Finance and Statistics, Jonkoping University

Jonkoping, Sweden

Ghazi Shukur

Department of Economics, Finance and Statistics, Jonkoping University

Jonkoping, Sweden

&

Department of Economics and Statistics, Linnaeus University

Vaxjo, Sweden

Received: February 6, 2012 Accepted: February 16, 2012 Published: May 1, 2012

doi:10.5539/ijsp.v1n1p2 URL: http://dx.doi.org/10.5539/ijsp.v1n1p2

Abstract

A new shrinkage estimator for the Poisson model is introduced in this paper. This method is a generalization of the Liu (1993) estimator originally developed for the linear regression model and will be generalized here to be used instead of the classical maximum likelihood (ML) method in the presence of multicollinearity since the mean squared error (MSE) of ML becomes inflated in that situation. Furthermore, this paper derives the optimal value of the shrinkage parameter and based on this value some methods of how the shrinkage parameter should be estimated are suggested. Using Monte Carlo simulation where the MSE and mean absolute error (MAE) are calculated it is shown that when the Liu estimator is applied with these proposed estimators of the shrinkage parameter it always outperforms the ML.

Keywords: Estimation, MSE, MAE, Multicollinearity, Poisson, Liu, Simulation

AMS Subject classification: Primary 62J07, Secondary 62F10

1. Introduction

In the field of economics, health, social and physical sciences, the dependent variable often comes in the form of a non-negative integers or counts. In that situation one often apply the Poisson regression model which is usually estimated by maximum likelihood (ML) where the solution to a non-linear equation is found by applying iterative weighted least square (IWLS). This method has been shown in Mansson and Shukur (2011) to be sensitive to multicollinearity and it becomes difficult to make a valid statistical inference since the mean squared error (MSE) becomes inflated. In Mansson and Shukur (2011), a ridge regression estimator (RRE) was presented for logistic regression which was a generalization of that proposed for linear regression by Hoerl and Kennard (1970). It has been shown that the RRE outperformed the ML.

The RRE is effective but as Liu (1993) pointed out it has the disadvantage that the estimated parameters are complicated non-linear functions of the ridge parameter k. Therefore, in this paper another shrinkage estimator for the Poisson model will be proposed which is a generalization of the method proposed for linear regression by Liu (1993). The advantage of this method is that the estimators are a linear function of the shrinkage parameter d. For this reason, this shrinkage estimator has become more popular during recent years (see for examples, Akdeneiz & Kaciranlar, 1995; Kaciranlar, 2003; Alheety & Kibria, 2009 among others).

The purpose of this paper is to solve the problem of an inflated MSE of the ML estimator by applying a Liu estimator. Furthermore, we derive the optimal value of the shrinkage parameter and based on this value we suggest some methods of how the shrinkage parameter should be estimated. In a Monte Carlo study we evaluate the performance of the ML and the Liu estimator applied with the suggested estimators of the shrinkage parameter. The performance criteria used in the simulation study is the MSE and mean absolute error (MAE). In our simulation, factors including the degree of correlation, the sample size and the number of explanatory variables are varied. In this application, the effect of the usage of cars and trucks on the number of killed pedestrians in different counties in Sweden is investigated.

This paper is organized as follows: In Section 2, the statistical methodology is described. In Section 3, the design of the Monte Carlo experiment is presented and the result from the simulation study is discussed. An application is presented in Section 4. Finally, a brief summary and conclusions is given in section 5.

2. Methodology

2.1 Poisson Regression

The Poisson regression model is a benchmark model when the dependent variable (y_i) comes in the form of counts data and distributed as $Po(\mu_i)$, where $\mu_i = \exp(x_i\beta)$, x_i is the i*th* row of X which is a $n \times (p+1)$ data matrix with p explanatory variables and β is a $(p+1) \times 1$ vector of coefficients. The log likelihood of this model may be written as:

$$l(\mu; y) = \sum_{i=1}^{n} \exp(x_i\beta) + \sum_{i=1}^{n} y_i \log(\exp(x_i\beta)) + \log\left(\prod_{i=1}^{n} y_i!\right). \tag{1}$$

The most common method to maximize the likelihood function is to apply the IWLS algorithm:

$$\hat{\beta}_{ML} = \left(X'\hat{W}X\right)^{-1}\left(X'\hat{W}\hat{z}\right), \tag{2}$$

where $\hat{W} = diag[\hat{\mu}_i]$ and $\hat{z}$ is a vector where the ith element equals $\hat{z}_i = \log(\hat{\mu}_i) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}$. The MSE of this estimator equals:

$$E\left(L_{ML}^2\right) = E\left(\hat{\beta}_{ML} - \beta\right)'\left(\hat{\beta}_{ML} - \beta\right) = tr\left(X'WX\right)^{-1} = \sum_{j=1}^{J} \frac{1}{\lambda_j}, \tag{3}$$

where λ_j is the jth eigenvalue of the $X'WX$ matrix. When the explanatory variables are highly correlated the weighted matrix of cross-products, $X'WX$, is ill-conditioned which leads to instability and high variance of the ML estimator. In that situation it is very hard to interpret the estimated parameters since the vector of estimated coefficients is on average too long. By noting that the IWLS algorithm approximately minimizes the weighted sum of squared error (WSSE) one may apply a generalization of the Liu (1993) estimator for linear regression instead:

$$\hat{\beta}_d = \left(X'\hat{W}X + I\right)^{-1}\left(X'\hat{W}X + dI\right)\hat{\beta}_{ML} \tag{4}$$

For this estimator we have replaced the matrix of cross-products used in the Liu (1993) estimator with the weighted matrix of cross-products and the ordinary least square estimator (OLS) of β with the ML estimator. The MSE of the Liu estimator equals:

$$\begin{aligned} MSE\left(\hat{\beta}_d\right) &= E\left(L_d^2\right) = E\left(\hat{\beta}_d - \beta\right)'\left(\hat{\beta}_d - \beta\right) \\ &= E\left[\left(\hat{\beta}_{ML} - \beta\right)' Z'Z\left(\hat{\beta}_{ML} - \beta\right)\right] + (Z\beta - \beta)'(Z\beta - \beta) \\ &= tr\left[\left(\hat{\beta}_{ML} - \beta\right)'\left(\hat{\beta}_{ML} - \beta\right)Z'Z\right] + k^2\beta'\left(X'WX + kI\right)^{-2}\beta \\ &= \textstyle\sum_{j=1}^{J} \frac{\left(\lambda_j + d\right)^2}{\lambda_j\left(\lambda_j + 1\right)^2} + (d-1)^2 \sum_{j=1}^{J} \frac{\alpha_j^2}{\left(\lambda_j + 1\right)^2} \end{aligned} \tag{5}$$

where α_j^2 is defined as the jth element of $\gamma\beta$ and γ is the eigenvector defined such that $X'WX = \gamma'\Lambda\gamma$, where Λ equals $diag\left(\lambda_j\right)$. In order to show that there exist a value of d bounded between zero and one so that $MSE\left(\hat{\beta}_d\right) < MSE\left(\hat{\beta}_{ML}\right)$, we start by taking the first derivative of equation (5) with respect to d:

$$g'(d) = 2\sum_{j=1}^{J} \frac{\lambda_j + d}{\lambda_j\left(\lambda_j + 1\right)^2} + 2(d-1)\sum_{j=1}^{J} \frac{\alpha_j^2}{\left(\lambda_j + 1\right)^2} \tag{6}$$

and then by inserting the value one in equation (6) we get:

$$g'(d) = 2\sum_{j=1}^{J} \frac{1}{\lambda_j\left(\lambda_j + 1\right)} \tag{7}$$

which is greater than zero since $\lambda_j > 0$. Hence, there exists a value of d that lies between zero and one so that $MSE\left(\hat{\beta}_d\right) < MSE\left(\hat{\beta}_{ML}\right)$. Furthermore, the optimal value of any individual parameter d_j can be found by setting equation (6) to zero and solve for d_j. Then it may be shown that

$$d_j = \frac{\alpha_j^2 - 1}{\frac{1}{\lambda_j} + \alpha_j^2}, \tag{8}$$

corresponds to the optimal value of the shrinkage parameter. Hence, the optimal value of d_j is negative when α_j^2 is less than one and positive when it is greater than one. However, just as in Liu (1993) the shrinkage parameter will be limited to take on values only between zero and one.

2.2 Estimating the Shrinkage Parameter

The value of d may only take on values between zero and one and there does not exist a definite rule of how to estimate it. However, in this paper some methods will be proposed that are based on the work for linear ridge regression by for instance Hoerl and Kennard (1970), Kibria (2003) and Khalaf and Shukur (2005). As in those papers, the shrinkage parameter, d_j, will be estimated by a single value $\hat{d}$. The first estimator is the following:

$$D1 = \max\left(0, \frac{\hat{\alpha}_{\max}^2 - 1}{\frac{1}{\hat{\lambda}_{\max}} + \hat{\alpha}_{\max}^2}\right),$$

where we define $\hat{\alpha}_{\max}^2$ and $\hat{\lambda}_{\max}$ to be the maximum element of $\hat{\alpha}_j^2$ and $X'\hat{W}X$, respectively. Replacing the values of the unknown parameters with the maximum value of the unbiased estimators is an idea taken from Hoerl and Kennard (1970). However, for the Liu estimator another maximum operator is also used that will ensure that the estimated value of the shrinkage parameter is not negative. Furthermore, the following estimators will be used:

$$D2 = \max\left(0, median\left(\frac{\hat{\alpha}_j^2 - 1}{\frac{1}{\hat{\lambda}_j} + \hat{\alpha}_j^2}\right)\right), D3 = \max\left(0, \frac{1}{p}\sum_{j}^{J} \frac{\hat{\alpha}_j^2 - 1}{\frac{1}{\hat{\lambda}_j} + \hat{\alpha}_j^2}\right),$$

$$D4 = \max\left(0, \max\left(\frac{\hat{\alpha}_j^2 - 1}{\frac{1}{\hat{\lambda}_j} + \hat{\alpha}_j^2}\right)\right), D5 = \max\left(0, \min\left(\frac{\hat{\alpha}_j^2 - 1}{\frac{1}{\hat{\lambda}_j} + \hat{\alpha}_j^2}\right)\right).$$

Using the average value and the median is very common when estimating the shrinkage parameter for ridge parameter and the D2 and D3 estimators has direct counterparts in equation (13) and (15) in Kibria (2003). Using other quantiles such as the maximum value, was successfully applied in Khalaf and Shukur (2005) and the idea behind the D4 and D5 estimator are taken from those papers.

2.3 Judging the Performance of the Estimators

To investigate the performance of the Liu and the ML methods, we calculate the MSE using the following equation:

$$MSE = \frac{\sum_{i=1}^{R}\left(\hat{\beta} - \beta\right)'\left(\hat{\beta} - \beta\right)}{R}, \tag{9}$$

and the MAE as:

$$MAE = \frac{\sum_{i=1}^{R}\left|\hat{\beta} - \beta\right|_i}{R} \tag{10}$$

where $\hat{\beta}$ is the estimator of β obtained from either ML or Liu and R equals 2000 which corresponds to the number of replicates used in the Monte Carlo simulation.

3. The Monte Carlo Simulation

This section consists of a brief description of how the data is generated together with a discussion of our findings.

3.1 The Design of the Experiment

The dependent variable of the Poisson regression model is generated using pseudo-random numbers from the $Po(\mu_i)$ distribution, where

$$\mu_i = \exp\left(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}\right), i = 1, 2, ...n, j = 1, 2, ...p. \tag{11}$$

Following Gibbons (1981), the parameter values in equation (11) are chosen so that $\sum_{j=1}^{p} \beta_j^2 = 1$ and $\beta_1 = \cdots = \beta_p$. To be able to generate data with different degrees of correlation we use the following formula to obtain the regressors:

$$x_{ij} = \left(1 - \rho^2\right)^{(1/2)} z_{ij} + \rho z_{ip}, i = 1, 2, ...n, j = 1, 2, ...p \tag{12}$$

where z_{ij} are pseudo-random numbers generated using the standard normal distribution and ρ^2represents the degree of correlation (see, Kibria, 2003; Muniz & Kibria, 2009 among others). In the design of the experiment three different values of ρ^2 corresponding to 0.85, 0.95 and 0.99 are considered. To reduce eventual start-up value effects we discard the first 200 observations.

In the design of the experiment the factors n and p are also varied. Since the ML estimators are consistent, increasing the sample size is assumed to lower MSE and MAE while p is assumed to increases the instability of $X'WX$ and lead to an increase of both measures of performance. We use sample sizes corresponding to 15, 20, 30, 50 and 100 degrees of freedoms (*df*=*n*-*p*) and number of regressors p equals to 2 and 4.

3.2 Results Discussion

The estimated MSE and MAE for p=2 and 4 are presented in Tables 1 and 2 respectively. It is evident from these tables that the degree of correlation and the number of explanatory variables inflate both the MSE and MAE while increasing the sample size leads to a decrease of both measures of performance. We can also see that the MSE increases more when considering the MSE instead of the MAE criteria. Hence, the gain of applying Liu is larger in terms of MSE than MAE. Furthermore, when looking at both measures of performance we can see that the estimator D5 is always either the shrinkage parameter that minimizes the MSE and MAE or it is very close to the shrinkage parameter that minimizes these loss functions.

4. Conclusions

In this paper, the shrinkage estimator developed by Liu (1993) for the linear regression model has been extended for the Poisson regression model. This estimator is proposed in order to reduce the inflation of the variance of the ML estimator caused by multicollinearity. The Liu and the ML estimators are evaluated by means of Monte Carlo simulations. Both MSE and MAE are used as a performance criteria and factors including the degree of correlation, the sample size and the number of explanatory variables are varied. Both measures of performance show that the proposed Liu estimators are better than ML in the sense of smaller MSE and MAE. We also observed that the estimator D5 is often the shrinkage parameter that minimizes the estimates MSE and MAE.

Acknowledgements

This paper was written while Dr. Kibria was visiting Professor Ghazi Shukur in the Department of Economics and Statistics, Linnaeus University, Vaxjo, and Jonkoping International Business School (JIBS), Jonkoping, Sweden during May-June 2011. He acknowledges the excellent research support provided by Linnaeus University and JIBS, Sweden. The authors are thankful to the referee for the valuable comments and suggestions.

References

Akdeniz, F., & Kaciranlar, S. (1995). On the almost unbiased generalized Liu estimator and unbiased estimation of the bias and MSE. *Communications in Statistics-Theory and Methods, 24*, 1789-1797. http://dx.doi.org/10.1080/03610929508831585

Alheety, M. I., & Kibria, B. M. G. (2009). On the Liu and almost unbiased Liu estimators in the presence of multicollinearity with heteroscedastic or correlated errors. *Surveys in Mathematics and its Applications, 4*, 155-167.

Alkhamisi, M., & Shukur, G. (2007). A Monte-Carlo Study of Recent Ridge Parameters. *Communications in Statistics-Theory and Methods, 36*, 535-547.

Gibbons, D. G. (1981). A Simulation Study of Some Ridge Estimators. *Journal of the American Statistical Association, 76*(373), 131-139. http://dx.doi.org/10.2307/2287058

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: biased estimation for non-orthogonal Problems. *Technometrics, 12*, 55-67.

Kaciranlar, S. (2003). Liu estimator in the general linear regression model. *Journal of Applied Statistical Science, 13*, 229-234.

Khalaf, G., & Shukur, G. (2005). Choosing Ridge Parameter for Regression Problems. *Communications in Statistics-Theory and Methods, 34*(5), 1177-1182. http://dx.doi.org/10.1081/STA-200056836

Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics, Theory and Methods, 32*, 419-435.

Liu, K. (1993). A new class of biased estimate in linear regression. *Communications in Statistics-Theory and Methods, 22*, 393-402. http://dx.doi.org/10.1080/03610929308831027

Muniz, G., & Kibria, B. M. G. (2009). On some ridge regression estimators: An Empirical Comparison. *Communications in Statistics-Simulation and Computation, 38*, 621-630. http://dx.doi.org/10.1080/03610910802592838

Mansson, K., & Shukur, G. (2011). On Ridge Parameters in Logistic Regression. *Communications in Statistics-Theory and Methods, 40*(18), 3366-3381. http://dx.doi.org/10.1080/03610926.2010.500111

Table 1. Estimated MSE and MAE of the estimators when p=2

Estimated MSE							Estimated MAE					
	ML	D1	D2	D3	D4	D5	ML	D1	D2	D3	D4	D5
ρ^2=0.85												
50	0.514	0.244	0.216	0.216	0.251	0.210	0.742	0.517	0.499	0.499	0.520	0.495
75	0.350	0.198	0.183	0.183	0.201	0.179	0.626	0.478	0.468	0.468	0.479	0.465
100	0.181	0.125	0.122	0.122	0.125	0.122	0.459	0.387	0.385	0.385	0.387	0.385
150	0.098	0.080	0.080	0.080	0.080	0.080	0.339	0.308	0.308	0.308	0.308	0.308
200	0.038	0.035	0.035	0.035	0.035	0.035	0.215	0.207	0.207	0.207	0.207	0.207
ρ^2=0.95												
50	1.755	0.581	0.442	0.442	0.656	0.375	1.391	0.730	0.655	0.655	0.761	0.621
75	1.044	0.379	0.302	0.302	0.396	0.279	1.104	0.633	0.589	0.589	0.644	0.575
100	0.621	0.273	0.250	0.250	0.279	0.239	0.847	0.552	0.538	0.538	0.556	0.531
150	0.324	0.186	0.183	0.183	0.187	0.181	0.625	0.477	0.474	0.474	0.477	0.473
200	0.135	0.103	0.103	0.103	0.103	0.103	0.409	0.359	0.359	0.359	0.359	0.359
ρ^2=0.99												
50	10.49	3.366	3.277	3.277	4.887	2.453	3.345	1.439	1.305	1.305	1.748	1.094
75	6.396	1.922	1.764	1.764	2.679	1.356	2.685	1.150	1.009	1.009	1.338	0.887
100	3.443	1.010	0.809	0.809	1.221	0.670	2.001	0.870	0.770	0.770	0.951	0.712
150	1.767	0.541	0.445	0.445	0.582	0.414	1.473	0.718	0.667	0.667	0.742	0.648
200	0.774	0.310	0.298	0.298	0.318	0.290	0.984	0.597	0.590	0.590	0.602	0.585

Table 2. Estimated MSE and MAE of the estimators when p=4

Estimated MSE							Estimated MAE					
	ML	D1	D2	D3	D4	D5	ML	D1	D2	D3	D4	D5
ρ^2=0.85												
50	1.810	0.861	0.645	0.572	1.079	0.465	1.846	1.260	1.109	1.082	1.341	1.049
75	0.916	0.483	0.386	0.374	0.533	0.366	1.383	1.025	0.956	0.948	1.052	0.943
100	0.424	0.286	0.268	0.268	0.290	0.267	0.978	0.815	0.800	0.800	0.818	0.800
150	0.194	0.158	0.157	0.157	0.158	0.157	0.671	0.613	0.612	0.612	0.613	0.612
200	0.072	0.067	0.067	0.067	0.067	0.067	0.417	0.404	0.404	0.404	0.404	0.404
ρ^2=0.95												
50	5.658	2.191	1.627	1.231	3.592	0.695	3.253	1.878	1.521	1.395	2.260	1.225
75	3.165	1.324	0.877	0.753	1.780	0.606	2.571	1.607	1.337	1.276	1.786	1.207
100	1.513	0.756	0.544	0.520	0.832	0.501	1.818	1.270	1.133	1.119	1.312	1.109
150	0.618	0.387	0.349	0.349	0.394	0.348	1.187	0.949	0.920	0.920	0.953	0.920
200	0.226	0.181	0.180	0.180	0.181	0.180	0.733	0.662	0.661	0.661	0.662	0.661
ρ^2=0.99												
50	27.75	9.273	8.294	7.180	20.554	2.621	7.329	3.530	3.052	2.810	5.577	1.553
75	17.44	6.077	4.952	3.854	11.940	1.424	5.880	2.988	2.446	2.168	4.274	1.390
100	7.538	2.631	1.740	1.370	4.320	0.750	4.050	2.173	1.699	1.555	2.725	1.279
150	3.471	1.435	0.870	0.730	1.795	0.616	2.791	1.687	1.362	1.297	1.852	1.236
200	1.321	0.693	0.520	0.510	0.728	0.508	1.739	1.241	1.127	1.122	1.263	1.121

Stochastic Modeling and Estimation of Market Volatilities with Applications in Financial Forecasting

Jin Zheng (Correspondence author)

Department of Mathematical Sciences, University of Liverpool

Liverpool L69 3BX, UK

E-mail: Jin.Zheng@liverpool.ac.uk

Dejun Xie

Department of Mathematical Sciences, Xian Jiaotong Liverpool University

Suzhou 215123, China

E-mail: Dejun.Xie@xjtlu.edu.cn

Received: February 15, 2012 Accepted: February 27, 2012 Published: May 1, 2012

doi:10.5539/ijsp.v1n1p7 URL: http://dx.doi.org/10.5539/ijsp.v1n1p7

Abstract

This paper aims to provide a framework for modeling and estimating the implied volatilities of stock prices using stochastic processes. The CIR (Cox-Ingersoll-Ross) model is proposed to capture the mean reversion characteristic as shown in the movement of the daily implied volatilities of the S&P 500 Index and Nikkei 225 Index. The maximum likelihood procedure is applied to estimate the parameters appearing in the model, where both analytical and approximation techniques are sought to handle the difficulties arising from the corresponding optimization problem. The procedure is validated with varying sampling methods by setting different time horizons and observation intervals. Results from numerical experiments show that the stochastic volatility model proposed in this paper functions well in both the US and Japan markets. As one of the applications, our approach is tested to be effective in detecting jumps in likelihood ratios, hence useful for forecasting stock market shocks and crisis.

Keywords: Stock volatility, CIR model, Maximum likelihood ratio, Financial forecasting

1. Introduction

Implied volatility plays a key role in studies of financial options and derivatives. A classic example is the Black-Schole option pricing method (Black and Scholes, 1973), where the implied volatilities of the same underlying asset with the same expiration date are assumed as constant, even when the strike prices are different. Here one of the most important assumptions of BS model is the constant volatility, or local volatility. However, abundant studies have questioned the validity of the constant volatility assumption. Dumas et al. (1998) analyzes the prices of S&P 500 index options for the period from June 1988 to December 1992, and shows that the volatility is a function of asset price and time to maturity. Derman et al. (1994) argues that the implied volatility is a function of share price, exercise price, time to maturity and a drift function. In the experiment of Peña et al. (1999), moneyness is considered as a determining factor of the implied volatility for options on the Spanish IBEX-35 exchange market. In Cont and Fontseca (2002), a contrast study focusing on out of money options is reported using S&P500 index and FTSE100 index. Cassese and Guidolin (2003) studies whether the implied volatility is a function of time to maturity, moneyness, or the interaction between the two with data of the DOTM options of MIB30 index on Italian market.

Understanding implied volatility is important to both theoretical and practical finance as it offers a reliable barometer to forecast the trend of financial market and largely affects the values of option contracts. Most of the previous literatures in this topic are of empirical in nature. In this work, we intend to address the problem by structural modeling approach where the CIR mean-reverting process is adopted to describe the movement of implied volatilities over time. The exponential average method is used to generate the daily implied volatilities for S&P 500 and Nikkei 225, for the purpose of which Chaikin's volatility function is implemented in finding the spread between the stocks' high price and low price and in determining the percent of changes in moving average of high price against low price for a specified time interval. The maximum likelihood procedure is carried out to estimate the parameters of the CIR model based on discretization of the model and approximation of the corresponding transitional probability density function. Matlab algorithms are proved to be helpful in solving the pertaining optimization problems arising from the maximum likelihood procedure. Likelihood ratio analysis of historical movement of volatilities is carried out for predicting stock market shocks.

The rest of the paper is organized as follows. Section 2 discusses the methodology and procedures of our study, including data description and model derivation. Section 3 contains data analysis and empirical results of our study. Section 4 is to show the applications of our modeling and estimation procedure for financial crisis prediction. Concluding remarks and and possible future directions are provided in Section 5.

2. Methodology and Procedures

2.1 Data for the Study

The data sets used in this paper are S&P 500 index and Nikkei 225 index. The S&P 500 index is chosen for the following considerations. First, since its inception in 1957, S&P 500 index has been a barometer for the American economy, which draws interests from many researchers. The market trend predicted by the many financial analysts is mostly based on observing the movement of S&P 500 index, although such a trend analysis may not always quickly foresee the potential bubbles existed in the market. Second, since there is a large trading volume of S&P 500 index every day, the data set is abundant and representative for the experiment in this model. Third, the data are maintained continuous and consistent in S&P 500 index, which is critically important for model testing and selection. Nikkei 225 is a stock market index for the Tokyo Stock Exchange. As the most widely quoted average of Japanese equities, Nikkei 225 is a benchmark to value the Japanese economy. Previous literatures such as Kunitomo and Sato (2009) have emphasized the importance of Nikkei 225 for Japanese market. The financial systems are relatively mature in the United States and Japan, which is also one reason to choose the indexes of these two countries. The data entry is composed of trading date, high price and low price. The data sets are openly available at yahoofinance. All public holidays are deleted from the raw data entries for considerations of consistency and continuity. The implied volatilities can be computed by the method of exponential average, for instance. The parameters for CIR model can be estimated through maximum likelihood procedure, although technical points such as the singularities arising from the optimization iterations involving Bessel type of functions have to be addressed separately.

2.2 Model Description

The adoption of a mean-reverting stochastic process expressed in terms of, say, the Ornstein Uhlenbeck equation

$$d\sqrt{v(t)} = -\beta\sqrt{v(t)}dt + \delta dZ(t) \tag{1}$$

where $v(t)$ is the variance of the stock, and $Z(t)$ is a Brownian Motion, and β and δ are positive constants, for modeling volatilities can be traced back to Stein and Stein (1991), for instance, where several theoretical favorable consequences ensuing from such a gratification have been discussed, including the mathematical tractability in the case of simple coupon bond pricing. On the other hand, some undesirable implications, mainly the allowing of negative variance, have also been noticed and explained in Stein and Stein (1991). Therefore, it is more theoretically plausible and practically useful to adopt a model with only non-negative trajectories.

To transform equation (1), let $x(t) = \sqrt{v(t)}$, one gets

$$dx = -\beta x dt + \delta dZ(t)$$

Applying Ito′s lemma with $df = dv(t) = dx^2(t)$ (Heston,1993) gives

$$\begin{aligned} dx^2(t) &= \frac{\partial f}{\partial x(t)}dx(t) + \frac{1}{2}\frac{\partial^2 f}{\partial x(t)^2}dx^2(t) \\ &= 2x(t)(-\beta x(t)dt + \delta dZ(t)) + \delta^2 dt \\ &= (\delta^2 - 2\beta x^2(t))dt + 2x(t)\delta dZ(t) \end{aligned}$$

Replacing $x(t)$ by $\sqrt{v(t)}$ yields

$$dv(t) = (\delta^2 - 2\beta v(t))dt + 2\delta\sqrt{v(t)}dZ(t), \tag{2}$$

which is a special case of the usual CIR model:

$$dv(t) = k(\theta - v(t))dt + \sigma\sqrt{v(t)}dZ(t) \tag{3}$$

In this way, one can capture the variance or the volatility of the stock prices described by the CIR model. While Heston (1992) mainly concerns the theoretical benefits of adopting such models in terms of the option pricing, the work contained herein focuses on the validity and robustness of such modelings from statistical analysis perspective. In particular, we are more interested in the consistency and convergence of the parameter estimation via maximum likelihood procedure, for instance. We are also interested in whether the model is applicable to predicting stock plunges and economic cycles, which are more practical and reasonable concerns from practitioner and policy maker's point of view.

2.3 ML Estimation for CIR Process

Under the Euler-Maruyama approximation, we can rewrite equation (3) as

$$\Delta v(t) = k(\theta - v(t))\Delta t + \sigma\sqrt{v(t)}\Delta Z(t)$$

or equivalently, by the definition of $\Delta v(t) = v(t+\Delta t) - v(t)$ and $\varepsilon_t = \Delta Z(t)$

$$v(t+\Delta t) - v(t) = k(\theta - v(t))\Delta t + \sigma\sqrt{v(t)}\varepsilon_t \tag{4}$$

where ε_t is to be minimized in the following maximum likelihood procedure.

To obtain maximum likelihood estimation of the CIR process, we first notice that the transitional probability density for the process is explicitly given as (Cox et al., 1985)

$$p(\Delta, x, y) = c * \exp(-u - v) * (\frac{u}{v})^{\frac{q}{2}} * I_q(2\sqrt{uv})$$

where

$$\begin{aligned} c &= \frac{2k}{\sigma^2(1-\exp(-k\Delta))} \\ q &= \frac{2k\theta}{\sigma^2} - 1 \\ u &= cx * \exp(-k\Delta) \\ v &= cy \end{aligned}$$

from which one can derive the log-likelihood function of CIR process :

$$\begin{aligned} \ln L(m) = (N-1)\ln c + \sum_{i=1}^{N-1}(-u_{t(i)} - v_{t(i+1)} \\ +0.5q\ln\frac{v_{t(i+1)}}{u_{t(i)}} + \ln I_q[2\sqrt{u_{t(i)}v_{t(i+1)}}]) \end{aligned}$$

where $m = (k, \theta, \sigma)$ denotes the estimation of the parameters for the model, and N is the total number of observations. To obtain the estimated parameters $m = (k, \theta, \sigma)$, one needs to maximize the log-likelihood function $\ln L(m)$. The maximum of $\ln L(m)$ can be solved by the system of equations (5) :

$$\begin{cases} \frac{\partial \ln L(m)}{\partial k} = 0 \\ \frac{\partial \ln L(m)}{\partial \theta} = 0 \\ \frac{\partial \ln L(m)}{\partial \sigma} = 0 \end{cases} \tag{5}$$

However, the estimation of $m = (k, \theta, \sigma)$ is not trivial with a modified Bessel function of the first kind $I_q(2\sqrt{uv})$ contained in the expression, which is related one of the two linearly independent solutions to the following modified Bessel's differential equation:

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} - (x^2 + n^2)y = 0$$

Consider the definition of the Modified Bessel's function of the first kind of order n:

$$\begin{aligned} I_n(x) &= \frac{x^n}{2^n\Gamma(n+1)}[1 + \frac{x^2}{2(2n+2)} + \frac{x^4}{2*4(2n+2)+(2n+4)} + \cdots] \\ &= \sum_{k=0}^{\infty}\frac{x^n}{2^n\Gamma(n+1)} * \frac{(\frac{x}{2})^{2k}}{\frac{(n+k)!}{n!}k!} \\ &= \sum_{k=0}^{\infty}\frac{(\frac{x}{2})^{2k+n}}{n!\frac{(n+k)!}{n!}k!} \\ &= \sum_{k=0}^{\infty}\frac{(\frac{x}{2})^{2k+n}}{k!\Gamma(n+k+1)} \end{aligned}$$

Substituting $x = 2\sqrt{uv}$, $n = q$ into the equation $I_x(n)$ yields

$$I_q(2\sqrt{uv}) = \sum_{k=0}^{\infty} \frac{(-1)^k(\frac{2\sqrt{uv}}{2})^{q+2k}}{k!\Gamma(q+k+1)}$$

It is costly to solve the system of the equations (5) for obtaining the estimated parameters $m = (k, \theta, \sigma)$ without leveraging on computing packages provided by the fminsearch function in Matlab, for instance. However, the Matlab routine produces different answers for the estimated parameters with different initial values. In this paper, the ordinary least square procedure proposed by Kladıvkos (2011) is used to obtain the initial points of optimization.

Plug in the observed sample values into equation (4) and rearrange it, we have

$$\sum_{i=1}^{N-1} \varepsilon_t^2 = \sum_{i=1}^{N-1} [\frac{v(t_{i+1}) - v(t_i)}{\sigma\sqrt{v(t_i)}} - \frac{k(\theta - v(t_i))\Delta t}{\sigma\sqrt{v(t_i)}}]^2 \tag{6}$$

where the estimation of initial value of σ is the standard deviation of residuals, after which the initial values of k and θ can be obtained using the following procedure of ordinary least squares:

$$\begin{cases} \dfrac{\partial \sum_{i=1}^{N-1} \varepsilon_t^2}{\partial k} = 0 \\ \dfrac{\partial \sum_{i=1}^{N-1} \varepsilon_t^2}{\partial \theta} = 0 \end{cases} \tag{7}$$

2.4 Procuring Daily Volatilities

As the main objective is to model the implied volatilities of S&P 500 index and Nikkei 225 index as well as to estimate the parameters of the governing model, we need a proper scheme to measure the daily volatility of a specific stock given the historical data, where the model under consideration in this work is given in the equation (3).

Since $v(t)$, the daily variance is unknown, the daily volatility has to be retrieved with implicit method. In this paper the spread between the stock's high price and low price in a specified interval such as a trading day is computed in order to determine the Chaikin's volatility using percentage change in the moving averages for a given period. There are two major steps for the calculation.

First, we calculate an exponential moving average of the difference between the daily high and low prices. Compared to simple average method, the exponential moving average method assumes the latest price contributes more weight than the prices in more distant periods. Generally, a 10-day moving average is recommended (Achelis, 2011). The general form of a series Y is defined as

$$S_t = \alpha * Y_{t-1} + (1-\alpha) * S_{t-1}$$

where S_t is the exponential average value of stock price at time t, Y_t is the observation at time t, α represents the degree of weighting decrease. According to Hunter (1986), we can rewrite as a weighted sums of the following series:

$$\begin{aligned} S_t = \alpha * [Y_{t-1} + (1-\alpha) * Y_{t-2} + (1-\alpha)^2 * Y_{t-3} + \cdots \\ +(1-\alpha)^9 * Y_{t-10}] + (1-\alpha)^{10} * S_{t-10} \end{aligned}$$

where Y_t represents difference between the high price and the low price of a specific stock at time t.

Second, we calculate the percentage of changes over a specified time period according to the formula

$$\sigma = \frac{S_t - S_{t-10}}{S_{t-10}} * 100\%.$$

3. Data Analysis and Numerical Results

3.1 Model Calibration Using Historical Data

We present the results for estimated parameters, k, θ, and σ using daily observations of S&P 500 Index and Nikkei 225 index. Since the observations are daily recorded, the time interval, Δ, is set as 1/365 in the unit of number of years. The Figure 1 and 2 below show the historical behavior of S&P 500 Index and Nikkei 225. The daily volatility against time is plotted in the graph. The data of S&P 500 Index and Nikkei 225 Index both have a duration of 12 years from 1995 to 2006. The results of parameter estimations from the implementation of the maximum likelihood procedure are presented in the following Table 1 and Table 2.

Table 1 shows the results for the estimated parameters k, θ and σ for S&P 500 index. In this experiment, 5 years', 8 years', 10 years' and 12 years' data are used for comparative analysis. As shown in Table 1, the speed that the process comes back to the mean, k, decreases smoothly from 5 years data to 12 years data. The estimated long-term means, θ , are close to the means of historical data for different durations, which means the value for θ approaches to the corresponding value of the mean of $v(t)$. In addition, the values for estimated long-term means, θ, converge to 0.03 in different sub-periods. The estimated values for volatility, σ, for the periods of 5 years, 8 years, 10 years and 12 years are all located in the interval [0.05 0.08]. The estimated value for θ has shown slight increment with the increase of time duration, although the increment is not significant.

Table 2 shows the results for the estimated parameters k, θ and σ for Nikkei 225 index. In this experiment, again, we use 5 years', 8 years', 10 years' and 12 years' data to make comparisons. In Table 2, the drift parameter, k, varies in a wide range for different periods. The sudden change of k is a symbol of the dramatic changes in the implied volatility, which has to do with the economic shocks. For instance, the Asian market saw financial and currency crisis in 1997. Russia experienced debt default and long-term capital management crisis in 1998. Both of the two crises had a significant impact on major Asian markets like Japan's Nikkei 225, which can explain why k jumped from 4 to 26.8. Moreover, the estimated long-term means approach to the means of historical data, the trend of which is the same as S&P 500 index, And the estimated volatility, σ, with the time period of 5 years, 8 years, 10 years and 12 years are approaching to 0.07. Similarly, the values for θ are not exactly the same for different periods, but the difference is insignificant.

Overall, two stock markets have profiled similar trends of volatility. Compared to those for Nikkei 225 Index, the estimated parameters of the volatilities of S&P 500 Index are more consistent. As shown in Table 1, the estimated parameters, k, θ and σ show small difference when using different sampling periods. However, for Nikkei 225, the estimates to the speed of adjustment for the process, k, has dramatic changes from one sampling period to another, which is due to the economic crisis in Asia. The volatility model in this paper, or CIR model, functions well in a stable market.

To make the results more comprehensive, we extend our experiment to include the cases of using short-term data, i.e., 1 year and 2 years of S&P 500 Index and Nikkei 225 Index to obtain the estimates of the model parameters. However, the results are not as consistent as expected. In particular, k, the speed that the process will come back to the mean, decreases dramatically from 1 year's data to 2 years' data for S&P 500 Index. By observing the daily recorded data in 2006, the value for k is 144.2, while when we use 2 years' data in 2005 and 2006, k jumps to 45.8. As shown in Table 1, the estimates of k using medium to long term observation durations approach to 6, implying that the model is more suitable for medium to longterm forecast. In addition, we use 1 year's data to 2 years' data for Nikkei 225 Index to estimate the parameters. The estimations of k, θ and σ are consistent in different samplings, but the estimation of k flutters due to economic shocks.

3.2 Influence of Observation Frequency

3.2.1 Volatility Indicator Based on EMA

Chainkin's volatility is used in this paper to obtain the daily implied volatilities for S&P 500 Index and Nikkei 225 Index. Practically a 10-day exponential moving average is generally used for computing the volatility. However, there is not a generally agreed horizon for risk management. The horizons differ from different derivatives, or financial instruments. The equity and foreign exchange have horizons of 7 to 10 days, while the interest rate instruments have horizons for about 30 days (Diebold et al., 1998). Traders will choose a time period fitting to their investment time frame or preference. Generally, risk is assessed at a short time horizon (Christoffersen et al., 1998). To make the results more conclusive, we repeat the estimation process and compare the volatilities obtained by 5-day and 15-day EMA methods, in comparison with the previous results using 10-day EMA method.

Table 3 and 4 show the results of estimated parameters for S&P 500 Index and Nikkei 225 Index for 5-day moving average, 10-day moving average and 15-day moving average from 1995 to 2006. As shown in Table 3, for S&P 500 Index, k fluctuates wildly when using different time periods for averaging. In addition, k achieves the smallest value when 10 day moving average is used. The values for σ and θ vary insignificantly with different time periods. Table 4 shows the similar results for the values of estimated parameters for Nikkei 225 index.

3.2.2 Influence of Observation Frequency

In this paper, abundant observations are recorded with daily recorded data. It is desirable to investigate the influence of observation frequency on the parameter estimation. For the selected CIR model in this paper, we carry out the estimation process using not only daily observations, but also weekly observations and monthly observations. Table 5 and 6 present the results of estimated parameters for S&P 500 Index and Nikkei 225 Index for daily, weekly and monthly data from 1995 to 2006. The values of the estimated parameters using the data set of S&P 500 Index do not change significantly except for k. The value for k achieves the maximum for monthly recorded data, and the minimum for daily recorded data.

The results of the estimated parameters using Nikkei 225 index are consistent with S&P 500 Index, implying that time increment in sampling or increase of observation frequencies result in different parameter values for stochastic volatility models.

In this study, the parameters of the stochastic volatility model are estimated through the maximum likelihood procedure. We would like to remark that there exist alternative approaches to estimate the parameters for stochastic models, such as the Bayesian estimation method. Compared to maximum likelihood method, Bayesian estimation requires the specification of a prior distribution for unknown parameters. And also, Bayesian method is often realized through sampling simulations with augmented latent points, giving rises of the issue of inconsistency. Thus maximum likelihood estimation is still more preferred for the current work. The results of the maximum likelihood procedure through this work show that the parameters θ and σ are more consistent for different sampling time periods. However, the estimated values for k are not quite consistent for different sampling periods, even for different time horizons of averaging. Similar findings, particularly the difficulty of achieving convergent values for k, have been discussed in Feng and Xie (2011, 2012), for instance.

4. Application in Financial Crisis Prediction

4.1 Economic Cycle

Stock volatility, which represents massiveness of the change of stock prices (Olowe & Ayodeji, 2009), is a major indicator of future economic activity (Romer, 1990). It is generally believed that the stock market fluctuation is closely related to the macroeconomic development. The increased stock volatility usually signals higher uncertainty for future economic activity. The investors or the stock holders will tend to spend less money or wealth on consumption and investment, which will lead to decrease in aggregate demand, even economic slack. On the other hand, the decrease in stock volatility encourages the investors to make consumption or investment, which is often the first move in a cycle towards the economic prosperity.

4.2 Financial Crisis Prediction

Financial crisis, a symbol of economic downturn, or recession, happens when financial institutions or financial markets lose large part of their value. The stock volatility is a powerful indictor of financial crisis because it is closely related to business or economic cycle. By observing the U. S. stock market's implied volatility and quarterly percentage growth of real GDP, Raunig and Scharler (2011) carried out an experiment on the relationship between stock volatility and stock market shocks and found a negative relationship between volatility and GDP growth.

From Figure 7, one can see that the implied volatility fluctuates during the period from 1984 to 2006. As mentioned, stock market shocks are often a strong indicator of economic recessions. From the historical record of the implied volatility, it can be seen that the implied volatility fluctuates intensely in the period from 1986 to 1994. During these years, Japan suffered several major financial crises, including the burst of Japanese real estate bubble. From 1986 to 1991 the property market of Japan had undergone a great inflation, which explains the fluctuant movement of volatilities during the same period. In the following years, the world saw the Japan's ERM crisis in 1992, Mexican peso crisis in 1994, the Asian financial and currency crisis in 1997, Russian debt default and long-term capital crisis in 1998. All these crises had severally and collectively impacted the Japanese economy and its stock market.

4.3 Case Study: Stock Market Crisis in 1987

The U. S. stock market suffered one of its largest daily percentage decline on October 19, 1987, on which day the volatility saw a huge jump. Black (1988) largely attributed the stock market plunges on October 19 to the errors in the perception of mean-reversion expectations of investors, or equivalently, the so called mean reversion illusion (Hillerbrand, 2003). Here the mean reversion speed is the speed that the process will come back to the mean, which corresponds to the parameter, k, in the stochastic volatility model in this paper.

According to Hillebrand (2003), the mean reversion in returns is phenomenal, which is transient but recurring. In Table 7 and Figure 8, one can observe the fact that the value of k, the mean reversion speed, is higher after the stock plunge than before the plunge, which supports the concept of mean-reversion disillusion. In addition, k is higher at about 9 months before the plunge, which is consistent with the experiment carried out by Hillebrand (2003). The underestimation of the mean reversion is the leading reason for the stock plump.

In order to provide a further proof of the mean-reversion disillusion, which assumes the mean-reversion rate is higher after stock fluttering period, we carry out the statistical t-test with the following hypotheses:

$$\begin{aligned} H_0 &: k_1 < k_2 \\ H_1 &: k_1 \geq k_2 \end{aligned}$$

where k_1 refers to the mean of k before the stock plump and k_2 is the mean of k after the plump. We use two sampled t-test to evaluate the null hypothesis and alternative hypothesis:

$$\begin{aligned} t &= \frac{k_2 - k_1}{S_{(k_2-k_1)}} \\ S_{(k_2-k_1)} &= \sqrt{\frac{S_{k_1}^2}{n_1} + \frac{S_{k_2}^2}{n_2}} \end{aligned}$$

where S_t is the unbiased estimator of the variance of the two samples, n_1 = number of participants before the market plunge, n_2 = number of participants after the plunge. The degree of freedom, denoted as d. f., of our approximation is

$$d.f. = \frac{(\frac{S_{k_1}^2}{n_1} + \frac{S_{k_2}^2}{n_2})^2}{(\frac{S_{k_1}^2}{n_1})^2/(n_1-1) + (\frac{S_{k_2}^2}{n_2})^2/(n_2-1)}$$

The calculation shows that the p-value of the t-statistics 0.1695, which means we can not reject H_0 even at 10% significant level.

We apply Hillebrand (2003)'s approach of using likelihood ratio to test whether the mean-reversion speed is higher after the stock market plump. For this purpose, the following two models are compared.

$$\begin{aligned} Model1 &: \; d\ln S_t = (u - \frac{1}{2}\sigma^2)dt + \sigma dW_t \\ Model2 &: \; d\ln S_t = (u - \frac{1}{2})dt + \lambda(\ln V_t - \ln\ S_t)dt + \sigma dW_t \end{aligned}$$

The unconditional distribution of the log-price processes above are given by:

$$\begin{aligned} Model1 &: \; (\ln S_t - \ln S_{t-1}) \backsim N(u - \frac{1}{2}\sigma^2, \sigma^2) \\ Model2 &: \; (\ln S_t - \ln S_{t-1}) \backsim N(u - \frac{1}{2}\sigma^2 + \lambda(\ln V_t - \ln S_t), \sigma^2) \end{aligned}$$

where S_t is the stock price at time t, $V_t = S_0 exp((u - \frac{1}{2}\sigma^2)t)$, u is the long-term mean growth of the market, λ is the mean-reversion speed, σ is the volatility of the stock price, and W_t is the standard Brownian Motion. The hypotheses are

$$\begin{aligned} H_0 &: \; \lambda = 0 \\ H_1 &: \; \lambda \neq 0 \end{aligned}$$

The likelihood-ratio statistics is used to evaluate the null hypothesis and alternative hypothesis, where the likelihood ratio statistics is defined as

$$L = \frac{sup_{u\geqslant,\sigma\geqslant 0}[\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} exp(-\frac{((y_i-x_i)-(u-\frac{1}{2}\sigma^2))^2}{2\sigma^2})]}{sup_{u\geqslant 0,\sigma\geqslant 0,\lambda\geqslant 0}[\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} exp(-\frac{((y_i-x_i)-(u-\frac{1}{2}\sigma^2+\lambda(\ln V_t - \ln S_t)))^2}{2\sigma^2})]}$$

where y_i and x_i represent the logarithms of stock prices on ith and $(i-1)$th day.

We apply the transition probability functions of Model (1) and Model (2) to the likelihood ratio statistics using four different sampling periods. The parameters (u, σ) in Model (1) and (u, σ, λ) in Model (2), are estimated by MLE using daily stock index values. The results are presented in Figure 9.

4.4 Further Applications

The implied volatility model is widely accepted and used in financial market. In risk management and financial mathematics, value-at-risk is a widely used risk measure technique which estimates the risk of loss on a specific portfolio of financial assets. To be more concrete, value-at-risk of a given portfolio demonstrates the expected maximum loss over an aimed horizon within a given confidence interval (Larsen, 2001). Since the implied volatility provides an unbiased and efficient forecast of future market, capturing the volatility process is a main approach for value-at-risk estimation. Hence, the implied volatility models in Section 2 can be used for computing value-at-risk portfolios, which include assets whose payoffs/returns are functions of the S&P 500 index or Nikkei 225 Index (Cassese & Guidolin, 2003), although it is not often to have closed form solution for value-at-risk. The chosen portfolio can be combination of equity shares, assets and other financial derivatives.

5. Concluding Remarks

To estimate whether the implied volatilities are constant during the whole duration of options as Black-Schole model assumed, we use CIR model to describe the implied volatilities and maximum likelihood method to estimate the parameters. The data set in this paper are S&P 500 Indexes and Nikkei Index from 1995 to 2006. According to the statistical analysis results, the selected stochastic volatility model fits well the S&P 500 Index and Nikkei Index. The robustness of our model is calibrated with historical data with varying sampling methods. The estimated parameters are convergent in the stable market.

The implied volatilities can be used as an unbiased and efficient forecast of the performance of financial market. The model of implied volatility can be applied in many fields of financial market, including for calculation of the value-at-risk for portfolios and for stock crisis prediction. The results support the idea that the implied volatility tends to be higher in the recession period. In the meantime, the mean-reverting disillusion, which explains the mean-reverting speed is higher after the stock fluttering period than before, is proved by both the t-statistics and likelihood ratio test in the study. The approaches of financial modeling and the relating parameter estimation procedures contained in this study also provide useful supports to neighboring studies such as characterization of mortgage loans and American put options (Xie et al., 2007, 2011).

One limitation of the current study is the relatively narrow scope of the observations, where only 12 years of data of two mature markets are considered. For future research, the data set period should be chosen to allow more frequent regime switches. Also an improved modeling should be able to achieve more stable and convergent estimation of the mean-reverting speed, k. In addition, the current work only concerns the stochastic modeling for S&P 500 Index and Nikkei Index. A further study is recommended to be carried on for other types of index such as KOPSI 200 in Korea and the Hushen 300 Index in China.

Acknowledgment

The research is partially supported by the Natural Science Foundation of Jiangsu Province of China. The authors would like to thank two anonymous referees for their helpful comments that have improved this manuscript.

References

Achelis, S. B. (1999). Trading Solutions Function Library. [Online] Available: http://www.tradingsolutions.com/functions/ChaikinsVolatility.html (November 7, 2011).

Black, F., & Scholes, M. (1973). The pricing of Options and Corporate Liabilities. *Journal of Political Economy, 81,* 637-654. http://dx.doi.org/10.1086/260062

Black, F. (1988). An Equilibrium Model of the Crash. *NBER Macroeconomics Annual, 3,* 269-275. http://dx.doi.org/10.2307/3584955

Cassese, G., & Guidolin, M. (2003). Modeling the Implied Volatility Surface: Does Market Efficiency Matter? *University of Virginia Working Paper.*

Christoffersen, P., Diebold, F., & Schuermann, T. (1998). Horizon Problems and Extreme Events in Financial Risk Management. *Economic Policy Review, Federal Reserve Bank of New York*, 109-118.

Cont, R., & Fonseca, J. D. (2002). Dynamics of Implied Volatility Surfaces. *Quantitative Finance, 2,* 40-56. http://dx.doi.org/10.1088/1469-7688/2/1/304

Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica, 53,* 385-408. http://dx.doi.org/10.2307/1911242

Dumas, B., Fleming, J., & Whaley, R. E. (1998). Implied Volatility Functions: Empirical Tests. *The Journal of Finance, 6,* 2059-2106. http://dx.doi.org/10.1111/0022-1082.00083

Derman, E., & Iraj, K. (1994). The Volatility Smile and Its Implied Tree. *Quantitative Strategies Research Notes,* Goldman Sachs.

Diebold, F., Hickman, A., Inoue, A., & Schuermann, T. (1998). Scale Models. *Risk, 11,* 104-107.

Feng, X. X., & Xie, D. J. (2011). Application of MCMC Algorithm in Interest Rate Modeling. *Proceeding of the International MutiConference of Engineers and Scientists*, 16-18.

Feng, X. X., & Xie, D. J. (2012). Bayesian Estimation of CIR model. *Journal of Data Science, 10*, 271-280.

Hillebrand, E. (2003). A Mean-Reversion Theory of Stock-Market Crashes. Center for Complex Systems and Visualization, Universitat Bremen, Department of Mathematics, Stanford University.

Hunter, J. (1986). The Exponentially Weighted Moving Average. *Journal of Quality Technology, 18*, 203-209.

Heston, S. L. (1992). Invisible Parameters in Option Prices. *The Journal of Finance, 48,* 933-947. http://dx.doi.org/10.2307/2329021

Heston, S. L. (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies, 6,* 327-343. http://dx.doi.org/10.1093/rfs/6.2.327

Kunitomi, N., & Sato, S. (2009). Realized Volatility, Covariance and Hedging Coefficient of the Nikkei-225 Futures with Micro-Market Noise. *18th World IMACS / MODSIM Congress*, 17-19.

Kladıvko, K. (2011). Maximum Likelihood Estimation of The Cox-Ingersoll-Ross Process: The MATLAB Implementation. Department of Statistics and Probability Calculus, University of Economics, Pragueand Debt Management Department, Ministry of Finance of the Czech Republic.

Larsen, N., Mausser, H., & Ursvsev, S. (2001). Algorithms for Optimization of Value-at-Risk. *Financial Engineering, E-Commerce and Supply Chain,* 129-157.

Olowe, R., & Ayodeji, R. (2009). Stock Return, Volatility And The Global Financial Crisis in An Emerging Market: The Nigerian Case. *International Review of Business Research Papers, 5,* 426-447.

Peñ̂a, I., Serna, G., & Rubio, G. (1999). Why do We Smile? On the Determinants of The Implied Volatility Function? *Journal of Banking and Finance, 23,* 1151-1179.

Romer, C. (2011). The Great Crash and the Onset of the Great Depression. *The Quarterly Journal of Economics, 105,* 597-1179.

Raunig, B., & Scharler, J. (2011). Stock Market Volatility and the Business Cycle. *Monetary Policy* & *Economy, Q2,* 54-63.

Stein, E. M., & Stein, J. C. (1991). Stock Price Distributions with Stochastic Volatility: An Analytic Approach. *Review of Financial Studies, 4,* 727-752.

Xie, D. D., Chen, X. F., & Chadam, J. (2007). Optimal Payment of Mortgages. *European Journal of Applied Mathematics, 18,* 363-388.

Xie, D. D., Edwards, D. A., Schleiniger, G., & Zhu, Q. H. (2011). Characterization of the American Put Option Using Convexity. *Applied Mathematical Finance, 18,* 353-365.

Table 1. Maximum Likelihood estimates of CIR model using historical daily data of S&P 500 Index. The starting dates for the data are January 3rd, 1995, January 2nd, 1997, January 4th, 1999 and January 2nd, 2002, respectively. The closing date is December 29th, 2006

SP 500	k	θ	σ^2	$\bar{v(t)}$
1995−2006	5.751066	0.042041	0.087743	0.041898
1997−2006	6.071555	0.039969	0.082144	0.038925
1999−2006	7.057089	0.034299	0.059359	0.034474
2002−2006	7.611455	0.029819	0.050125	0.029942

Table 2. Maximum Likelihood estimates of CIR model using historical daily data of Nikkei 225 Index. The starting dates for the data are January 4th, 1995, January 6th, 1997, January 4th, 1999 and January 4th, 2002, respectively. The closing date is December 29th, 2006

Nikkei 225	k	θ	σ^2	$\bar{v(t)}$
1995−2006	4.008421	0.051751	0.086224	0.045755
1997−2006	26.79940	0.044146	0.083093	0.044349
1999−2006	627.2639	0.041893	0.064212	0.042393
2002−2006	85.44733	0.041163	0.064792	0.041298

Table 3. Maximum Likelihood estimates of CIR model using historical daily data of S&P 500 Index from January 3rd, 1995 to December 29th, 2006. The table shows 5-day moving average, 10−day moving average and 15-day moving average for Chaikins Volatility

SP 500	k	θ	σ^2
n=5	226.3401	0.057505	0.132672
n=10	5.751065	0.042041	0.087743
n=15	95.10144	0.038993	0.078978

Table 4. Maximum Likelihood estimates of CIR model using historical daily data of Nikkei 225 Index from January 3rd, 1995 to December 29th, 2006. The table shows 5−day moving average, 10−day moving average and 15−day moving average for Chaikins Volatility

Nikkei 225	k	θ	σ^2
n=5	174.3905	0.060898	0.121118
n=10	4.008421	0.051751	0.086224
n=15	79.60699	0.040406	0.060591

Table 5. Maximum Likelihood estimates of CIR model using historical daily, weekly and monthly data of S&P 500 Index from January 3rd, 1995 to December 29th, 2006

SP 500	k	θ	σ^2
Daily	5.751066	0.042041	0.087743
Weekly	17.58786	0.059539	0.080261
Monthly	299.2399	0.085253	0.060909

Table 6: Maximum Likelihood estimates of CIR model using historical daily, weekly and monthly data of Nikkei 225 Index from January 3rd, 1995 to December 29th, 2006

Nikkei 225	k	θ	σ^2
Daily	4.008421	0.051751	0.086224
Weekly	13.63029	0.053029	0.081287
Monthly	164.2547	0.045228	0.053284

Table 7. The estimated results k of S&P 500 index on sample periods before and after the day of stock plunge on October 16, 1987. The data from October 16, 1987 to October 26, 1987 are not included so that the fluttering itself does not affect the sampling process

quarters before Oct.16	k	quarters after Oct.26	k
1	133.61	1	41.823
2	63.106	2	92.779
3	107.88	3	76.940
4	85.051	4	196.08
5	50.758	5	324.93
6	24.601	6	35.162
7	42.492	7	129.60
8	9.5577	8	11.100
9	60.435	9	72.700
10	77.535	10	4.8500

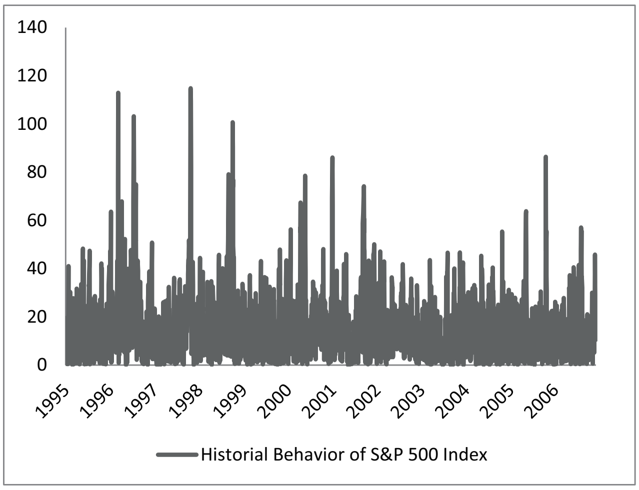


Figure 1. The daily implied volatility of S&P 500 Index over a duration of 12 years

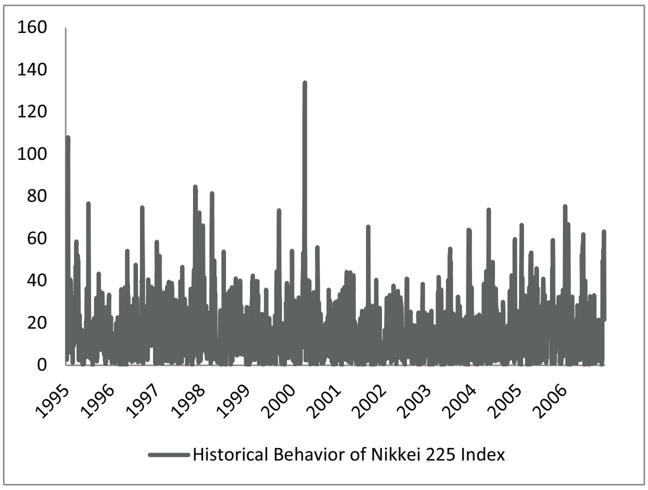


Figure 2. The daily implied volatility of Nikkei 225 Index over a duration of 12 years

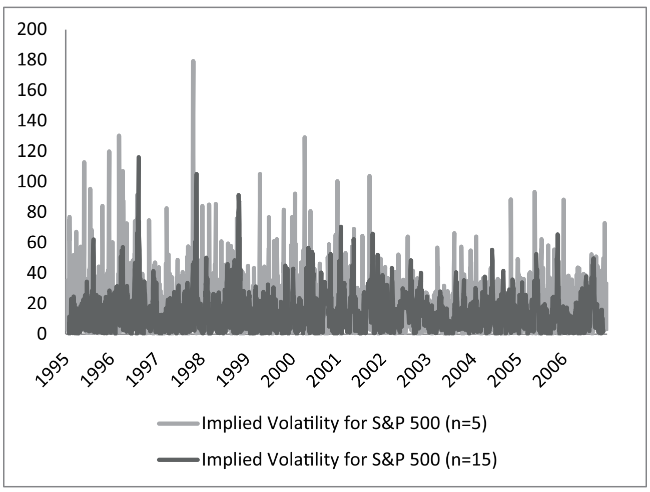


Figure 3. The daily implied volatility of S&P 500 Index over a duration of 12 years, for n=5 and n=15

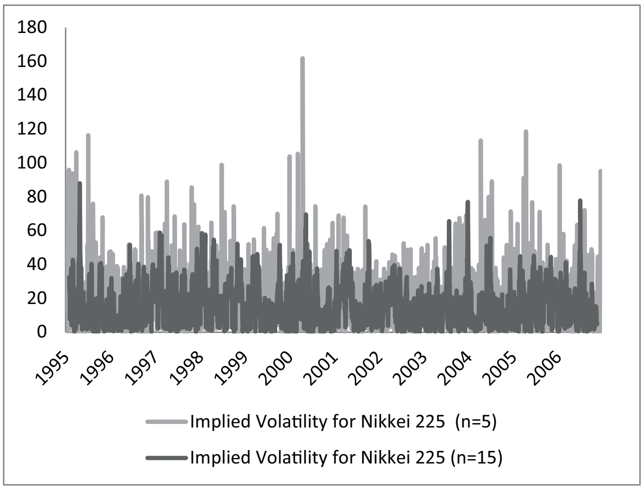


Figure 4. The daily implied volatility of Nikkei 225 Index over a duration of 12 years, for *n*=5 and *n*=15

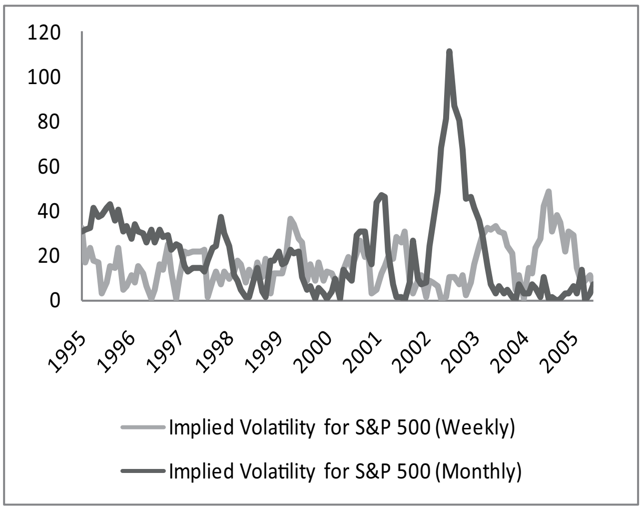


Figure 5. The weekly and monthly implied volatility of S&P 500 Index over a duration of 12 years

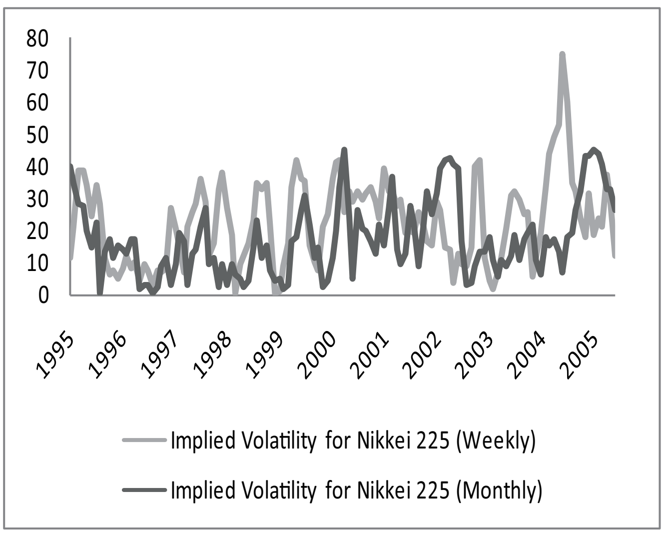


Figure 6. The Weekly and Monthly implied volatility of Nikkei 225 Index over a duration of 12 years

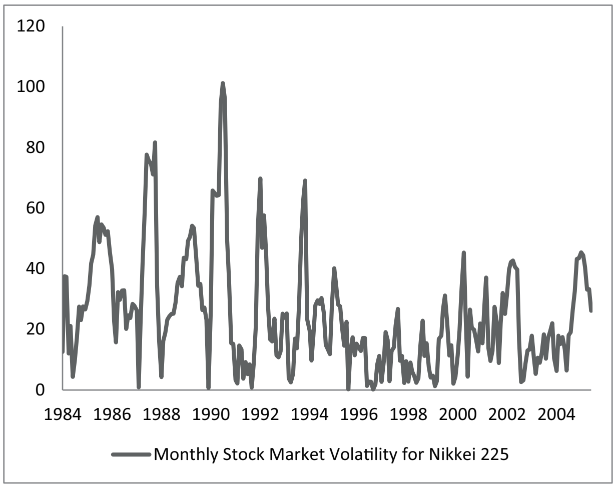


Figure 7. The monthly implied volatility of Nikkei 225 Index over a duration of 2 decades from 1984 to 2006

Figure 8. The estimated value for k of S&P 500 Index over a duration of 20 quarters before and after the stock crash on October 16, 1987. Specifically, the data set is from April 16, 1985 to April 26, 1990. The x-axis refers to nth quarter before or after the stock crisis

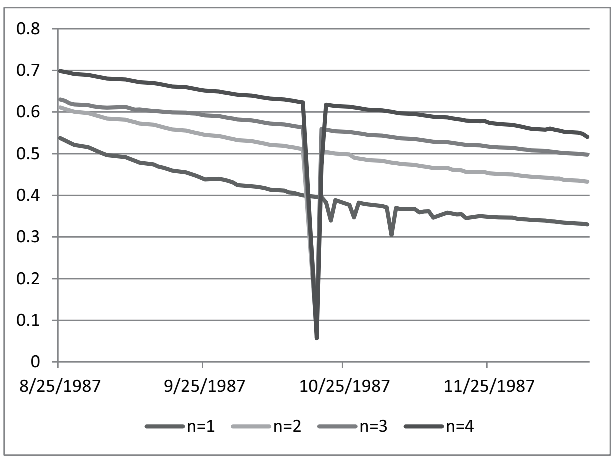


Figure 9. Likelihood-ratio test for Model 1 against Model 2 in Section 4.3. The estimation calculated for $n = 1, 2, 3, 4$, where n represents the quarter before or after the crash. The likelihood ratio jumps dramatically near to zero when the crisis happens, which means Model 2 is better than the null model during the period of the crisis. The graph shows the speed of mean-reversion time is higher after the stock crisis

Structural Credit Modeling Under Stochastic Volatility

Marcos Escobar (Corresponding author)
Department of Mathematics, Ryerson University
Toronto M3A 2K7, Canada
Tel: 1-416-979-5000 E-mail: escobar@ryerson.ca

Tim Friederich
Department of Mathematics, Technische Universität München
Parkring 11 85748 Garching b. Mnchen, Germany
Tel: 49-89-289-17417 E-mail: friederich@tum.de

Mikhail Krayzler
Department of Mathematics, Technische Universität München
Parkring 11 85748 Garching b. Mnchen, Germany
Tel: 49-89-289-17417 E-mail: krayzler@tum.de

Luis Seco
Department of Mathematics, University of Toronto
Toronto M5S 2E4, Canada
Tel: 1-416-978-3984 E-mail: seco@math.utoronto.ca

Rudi Zagst
Department of Mathematics, Technische Universität München
Parkring 11 85748 Garching b. Mnchen, Germany
Tel: 49-89-289-17401 E-mail: zagst@tum.de

Received: February 16, 2012 Accepted: March 5, 2012 Published: May 1, 2012
doi:10.5539/ijsp.v1n1p20 URL: http://dx.doi.org/10.5539/ijsp.v1n1p20

Abstract

This paper presents a structural credit model with underlying stochastic volatility, a CIR process, combining the Black/Cox framework with the Heston Model. We allow to calibrate a Heston Model for a non-observable process as underlying of the Black/Cox Model. A closed-form solution for the price of a down-and-out call option on the assets with the debt as barrier and strike price is derived using the concept of optional sampling. Furthermore, estimators are derived with the Method of Moments for Hidden Markov Chains. As an application in Statistical Finance, the default probabilities of Merrill Lynch during the financial crisis are examined.

Keywords: Credit models, Barrier options, Stochastic volatility, Black/Cox model, Heston model

1. Introduction

In order to describe the performance of a company on a daily basis, we can often refer to quoted stock prices only. These prices reflect the market value of the company's equity. Yet, it's mainly the assets and the liabilities which drive the performance of a stock, and which may develop independently. The higher the leverage ratio (i. e. the ratio between debt and equity capital) the more important it becomes to consider these underlying factors. Yet in general, we cannot observe daily time series for assets or liabilities.

Following the theory of structural models, whose foundations have been laid mainly by Merton (1973) and Black and Cox (1976), the company's value i.e. its equity's value is modeled as an option on the asset value with the value of the debt as strike price. In this paper, the value is modeled as the value of a down-and-out call option (DOC) with the assets as underlying and the value of the debt as both, knock-out barrier and strike price. A simple call option pays the difference between the underlying and the strike price at maturity if this difference is positive. A down-and-out call option only pays this difference if the underlying does not fall below a certain threshold at any time before maturity. This threshold is called knock-out barrier. Knock-out barrier and strike price do not necessarily have to be equal (see Escobar et al., 2012).

Modeling the assets as a Brownian motion as in the Black/Cox Model would have the drawback of assuming normally distributed returns and constant volatility. However, in particular the financial crisis has shown that volatility is not constant over time. Heston Model, which is used here to model the assets allows the company's assets to have stochastic volatility. This model is chosen for its suitability in the pricing of financial products as well as for its closed-form expression for higher conditional moments. It should be noted that Heston Model is a continuous-time stochastic volatility (SV) process as oppose to popular discrete-time SV processes known as ARCH/GARCH (see Bollerslev, 1986). There is a recent literature on continuous-time extensions of GARCH models, see for example COGARCH by Brockwell et al., (2006); but the complexity of these extensions for estimation and pricing purposes makes well-established SV models as Heston model still dominant for academics and practitioners alike. The first part of the paper focuses on the modeling and introduces a closed-form formula for the down-and-out option. In the second part, estimators are derived using the method of moments on a discretization of the process as well as the mixing properties of the Hidden Markov Chain. This generalization is inspired by the work of Genon-Catalot et al., (2000) and allows for the calibration of the parameters of the asset process.

The model not only allows to describe the value of the company and its assets. It also enables to simulate multiple scenarios of possible paths how the stock would evolve over time according to the model. These scenarios can be analyzed under different aspects. As an example, the default probabilities in these scenarios are examined. In this example, the model is applied to a case which can be regarded as one of the key events of the financial crises: the takeover of Merrill Lynch by the Bank of America, which *The Wall Street Journal* headlined "The End of Wall Street".

This paper is organized as follows: Section 2.1 introduces the stochastic volatility model for the assets. Section 2.2 presents the relationship between assets and equity as well as some numerical results. Section 2.3 derives the estimation methodology, while in Section 3 the approach is applied to data from Merrill Lynch. Section 4 gives a conclusion of the study.

2. The Stochastic Volatility Model

Asset returns, and hedge fund returns in particular, are not normally distributed. Thus, from today's point of view, a Black/Scholes Model is barely able to map reality in an appropriate way, as one of the main assumptions is that the returns are normally distributed. Furthermore, the Black/Scholes Model assumes the volatility of the asset process to be constant over time. This characteristic feature is called homoscedasticity. However, even the stock market suggests that this is not the case. For example, it can be observed that those times when the stock markets take a hit, the (implied) volatility is higher than in "peaceful" times of strong markets, see e.g. Engle (1982) or Heston (1993). Therefore, we propose a model with stochastic volatility to describe the asset process which also implies a stochastic volatility model for the equity process of the company.

2.1 Model Setup

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{Q})$ be a filtered probability space on the domain Ω with sigma algebra $\mathcal{F}$, filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}$, and a probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$. The underlying asset process A and the variance v of that process can be expressed through the following SDEs:

$$dA(t) = \mu A(t)dt + \sqrt{v(t)}A(t)dZ(t) \tag{1}$$

$$dv(t) = \kappa_v (v_\infty - v(t))\, dt + \varepsilon_v \sqrt{v(t)} dZ^v(t) \tag{2}$$

where

A	=	underlying asset process,
v	=	variance of asset process, where the volatility $\sigma_A = \sqrt{v}$,
μ	=	drift of the assets,
Z, Z^v	=	two independent Wiener processes in the probability space $(\Omega, \mathcal{F}, \mathbb{Q})$,
v_∞	=	long-term value of volatility,
ε_v	=	volatility of the variance process,
κ_v	=	mean-reversion speed.

These assumptions provide the bare minimum to obtain a setting where closed-form solutions are available not only for the price of the credit products but also for a fast and reliable estimation of the parameters. For example, in a more general formulation, Z and Z^v can be correlated, however the assumption of zero correlation is required to obtain closed-form expressions for the estimators in Section 2.3. The parameters v_∞, κ_v and ε_v have to fulfill the following conditions:

$$\kappa_v \cdot v_\infty \geq \frac{1}{2}\varepsilon_v^2 \tag{3}$$

$$\kappa_v > 0 \tag{4}$$

These conditions guarantee strict stationarity and α-mixing, see Genon-Catalot et al. (2000) which is necessary for the derivation of estimators in Section 2.3. Furthermore, the barrier as the debt is exponentially growing with the risk-free rate r:

$$D(t) = D(0) \cdot e^{\int_0^t r\,ds} = D(T) \cdot e^{-\int_t^T r\,ds} \tag{5}$$

where $D(t)$ denotes the value of the company's debt (or "liabilities") at time t.

The model defined by (1) and (2) is a so-called stochastic volatility model. Heston (1993) developed the model to overcome the shortcomings of the Black/Scholes Model where the latter seems to be too restrictive to map reality. Therefore, the model described by (1–2) is often referred to as the "Heston Model". In this model, the variance (2) is a stochastic process, introduced by Cox, Ingersoll, and Ross (1985), with the following features: First, the variance process has a long-term value or a "limit" v_∞ around which it oscillates. Next, there is the mean-reversion speed κ_v. This parameter describes how fast the variance will adjust to its mean v_∞. The term $(v_\infty - v(t))$ is the deviation of the current variance from v_∞. This term is weighted with the mean-reversion rate κ_v. Therefore, the higher κ_v the faster the variance will tend towards v_∞. Up until now, the variance process would be deterministic. In order to make it stochastic, the diffusion term $\varepsilon_v \sqrt{v(t)}dZ^v(t)$ is added, where ε_v is the volatility of the variance process itself. The characteristics of the stochastic volatility model become "visible" in Figure 1.

2.2 Barrier Options in the Stochastic Volatility Model

The aim is to price a barrier option with strike $D(T)$ on the assets A:

$$C(t, A) = \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_t^T r(s)\,ds} \cdot \max\{A(T) - D(T), 0\} \cdot \mathbf{1}_{\{\tau > T\}} \,\middle|\, \mathcal{F}_t\right] \tag{6}$$

where

$$C(t, A) = \text{option price of the barrier option with underlying } A \text{ at time } t,$$

$$\mathbf{1}_{\{\tau>T\}} = \begin{cases} 1 & \text{if } \tau > T \\ 0 & \text{else} \end{cases}$$

and τ is the time of default of the option, i.e. the first time, the asset process A crosses the barrier D. It is modeled as a stopping time on the interval $(t, T]$:

$$\tau = \inf\{t' \in (t, T] \ : \ A(t') < D(t')\} \tag{7}$$

The symbol $\mathbb{E}_{\mathbb{Q}}$ denotes the expected value under the arbitrage-free measure $\mathbb{Q}$.

Equation (6) describes a down-and-out call option with underlying A, strike and knock-out barrier D. This means that when the price of the underlying hits or falls below the barrier the option expires (Note 1).

2.2.1 Derivation of Option Pricing Formula

For deriving the formula of the option price described by (6), the most straight-forward approach would be to solve the corresponding SDEs. However, in this particular case, the option price can also be derived with the help of optional sampling, which actually is more elegant than working through the SDEs. As it is shown in detail in the appendix, the price of the option is:

$$C(t, A) = (A(t) - D(t)) \cdot \mathbf{1}_{\{\tau > t\}}. \tag{8}$$

The big advantage of (8) is that it provides a straight-forward inverse to calculate the asset value if both debt and equity value are known: Given the value of the equity $C(t)$ as well as the liabilities $D(t)$, and assuming that the company didn't default up to time t, i.e. $1_{\{\tau>t\}}$, the underlying assets $A(t)$ can be calculated as follows:

$$A(t) = C(t) + D(t) \tag{9}$$

Next, the option price – given a set of parameters specifying (1) and (2) – is calculated according to this formula and compared to a numerical simulation of the option price which is gained by evaluating (6). For this purpose, a large number of possible underlying asset paths is simulated with the same set of parameters. Then, the option price is the expected value of the discounted payoffs of all paths.

Figure 1 gives five examples for possible paths following the stochastic volatility model defined in (1) and (2). The parameters were chosen as follows, and are the same for all five paths:

$A(0)$	r	$v(0) = v_\infty$	κ_v	ε_v
100	0.04	0.01	0.5	0.1

Figure 1 displays the variance process $v(t)$, the according volatility process which is the square root of the variance process, and the asset process $A(t)$, where the shading of lines in these three graphs indicates that the three time series belong together.

For validating the formula, the price for several options is calculated analytically according to (8). This price is then compared to a numerical simulation of the option price. For that purpose, 10 000 paths under the given parameter set (the same as above, with $t = 0$ and $T = 5$) for the underlying asset process are simulated. For each of those paths, the payoff is calculated as

$$PO \quad = \quad \max\{A(T) - D(T),\, 0\} \cdot \mathbf{1}_{\{\tau > T\}} \tag{10}$$

in accordance with (6). The "numerical option price" is the discounted mean over those 10 000 payoffs, i.e.

$$C_{num}(t,\, A) = e^{-\int_t^T r(s)\,ds} \cdot \overline{PO}$$

where $\overline{PO} = \frac{1}{n}\sum_{i=1}^{n} PO_i$.

Table 1 presents the outcomes of this test. From left to right, the columns give the values of the initial debt $D(t) = D(0)$, and the number of paths for which the payoff equals zero. Next, the discounted mean over all payoffs is provided (C_{num}) along with the "correct" analytical option price, i.e. the option price according to Formula (8), together with their difference and the standard deviation of the 10 000 values for C_{num}, i.e. the discounted payoff. The last four columns provide some validation for the derived formula in form of the confidence interval. Considering these number, this test numerically validates the analytical derivation of Equation (8).

2.3 Estimation and Fitting of Parameters

Having the formula for the barrier option at hand, the next tool required is the estimators for the method of moments. The aim is to find the parameters of the asset process which is calculated from the equity and debt time series employing (8). In the following, there are always three different time series of interest: (1) the underlying asset path A of the company, (2) its debt D, which is also referred to as its liabilities, and (3) the company's equity C which is also referred to as the stock price itself or the equity. Furthermore, there are different possible interpretations of the initial value of the debt $D(0)$. If $A(0) = 1$, then $D(0)$ can be interpreted as the percentage of debt capital in the financing structure of the company at time $t = 0$. If $C(0) = 1$, then $D(0)$ is the leverage ratio the company operates with, because the leverage ratio is defined as the ratio between debt and equity capital.

The derivation of the estimators follows the lines of Genon-Catalot et al. (2000). As already mentioned, the two Wiener processes Z and Z^v denoting the stochastic component of the asset and volatility process are assumed to be uncorrelated. This leaves the following unknown parameters of the model described in (1) and (2): $\Theta = (\mu,\, v_\infty,\, \kappa_v,\, \varepsilon_v)$ which are to be fitted. The information on the debt D as well as the risk-free rate r can be observed and are assumed to be known. Actually, there is one more parameter which is unknown from the start which is $v(t)$, the level of variance at the time t the option is to be priced for (usually, $t = 0$). However, having $v_0 = v(0)$ as well as v_∞ as free parameters might bear problems for the fitting. Therefore, it is assumed that $v_0 = v_\infty$. Note that this does not imply that v_∞ is equal to the current volatility at the time the option price is calculated. As the volatility v itself is not observable and thus cannot directly influence the fitting, v_∞ and v_0 are rather regarded as two parameters of the model.

First of all, the actual calculation of suitable estimators requires some preparation. Assuming the model given by Equations (1) and (2), an application of Itô's lemma to the process $Y := \log A$ immediately shows:

$$dY(t) = \left(\mu - \frac{v(t)}{2}\right) dt + \sqrt{v(t)}dZ(t) \tag{11}$$

Under the condition that the variance process (2) is known and assuming $Y(0) = 0$ i.e. $A(0) = 1$,

$$(Y(t)|v) := (Y(t)|v(s),\ 0 < s < t) \sim \mathbf{N}\left(\mu \cdot t - \frac{1}{2}\int_0^t v(s)\,ds,\ \int_0^t v(s)\,ds\right) \tag{12}$$

where $X \sim \mathbf{N}(\mu,\ \sigma^2)$ means that the random variable X is normally distributed with mean μ and variance σ^2.

Lemma 2.1 *Assume the partition* $\{t_i\}_{i=1}^n$ *of the time interval* $[0, t]$ *where* $\triangle := t_i - t_{i-1}$. *Set*

$$\begin{aligned} R(i) &= \left(\mu - \frac{\overline{V}(i)}{2}\right) \cdot \sqrt{\triangle} + \frac{1}{\sqrt{\triangle}} \int\limits_{(i-1)\triangle}^{i\triangle} \sqrt{v(s)}dZ(s) \\ \overline{V}(i) &= \frac{1}{\triangle} \int\limits_{(i-1)\triangle}^{i\triangle} v(s)\,ds \end{aligned} \tag{13}$$

then

$$R(i) = \frac{Y(i\triangle) - Y((i-1)\triangle)}{\sqrt{\triangle}} \tag{14}$$

Using (12) and the fact that the sum of two normally distributed random variables is also normally distributed, it follows that:

$$(R(i)|v) \sim \mathbf{N}\left(\sqrt{\triangle}\mu - \frac{1}{2\sqrt{\triangle}}\int_{(i-1)\triangle}^{i\triangle} v(s)\,ds,\ \ \frac{1}{\triangle}\int_{(i-1)\triangle}^{i\triangle} v(s)\,ds\right)$$

and in terms of (13)

$$\left(R(i), |\overline{V}\right) \sim \mathbf{N}\left(\left(\mu - \frac{1}{2}\overline{V}(i)\right) \cdot \sqrt{\triangle},\ \ \overline{V}(i)\right). \tag{15}$$

The concept of α-mixing was introduced by Rosenblatt (1956) under the term "strongly mixing". A formal definition of α-mixing can be found in the appendix. It can be interpreted as a form of asymptotic independence of a process: $X(t)$ and $X(s)$ are more nearly independent the farther apart s and t, as commented by Davidson (1993). Thus, α-mixing is a weaker form of dependence. Under the mixing condition a central limit theorem can be shown to hold for general processes. A deeper discussion of α-mixing and other forms of the mixing-condition can be found in Bradley (2005).

Proposition 2.2 *Assume the process (2) satisfies (3) and (4). Then, for the process (14) there exist the following estimators for the moments of R:*

$$\frac{1}{n}\sum_{i=0}^{n-1} R(i+1) \xrightarrow{a.s.} \sqrt{\triangle}\mu - \frac{\sqrt{\triangle}}{2}\mathbb{E}\left[\overline{V}(1)\right] \tag{16}$$

$$\frac{1}{n-1}\sum_{i=0}^{n-2} R(i+1)R(i+2) \xrightarrow{a.s.} \triangle\mu^2 - \triangle\mu\mathbb{E}\left[\overline{V}(1)\right] + \frac{\triangle}{4}\mathbb{E}\left[\overline{V}(1)\overline{V}(2)\right] \tag{17}$$

$$\frac{1}{n}\sum_{i=0}^{n-1} R^2(i+1) \xrightarrow{a.s.} \triangle\mu^2 - (\triangle\mu - 1)\,\mathbb{E}\left[\overline{V}(1)\right] + \frac{\triangle}{4}\mathbb{E}\left[\overline{V}^2(1)\right] \tag{18}$$

$$\begin{aligned} \frac{1}{n-1}\sum_{i=0}^{n-2} R^2(i+1)R(i+2) \xrightarrow{a.s.} &\left(\triangle\mu^2 - (\triangle\mu - 1)\,\mathbb{E}\left[\overline{V}(1)\right] + \frac{\triangle}{4}\mathbb{E}\left[\overline{V}^2(1)\right]\right) \\ &\cdot \left(\sqrt{\triangle}\mu - \frac{\sqrt{\triangle}}{2}\mathbb{E}\left[\overline{V}(1)\right]\right) \end{aligned} \tag{19}$$

$$\begin{aligned} \frac{1}{n}\sum_{i=0}^{n-1} R^4(i+1) \xrightarrow{a.s.} &\triangle^2\mu^4 + \left(6\,\triangle\,\mu^2 - 2\,\triangle^2\,\mu^3\right)\mathbb{E}\left[\overline{V}(1)\right] \\ &+ \left(\frac{3}{2}\,\triangle^2\,\mu^2 - 6\,\triangle\,\mu + 3\right)\mathbb{E}\left[\overline{V}^2(1)\right] \end{aligned} \tag{20}$$

where $\mathbb{E}\left[\overline{V}(1)\right]$, $\mathbb{E}\left[\overline{V}^2(1)\right]$, *and* $\mathbb{E}\left[\overline{V}(1)\overline{V}(2)\right]$ *can be calculated as (see Proposition 4.1 in Genon-Catalot et al., 2000):*

$$\mathbb{E}\left[\overline{V}(1)\right] = v_\infty \tag{21}$$

$$\mathbb{E}\left[\overline{V}^2(1)\right] = v_\infty^2 + \frac{\varepsilon_v^2 v_\infty}{\kappa_v}\frac{\left(\kappa_v \triangle -1 + e^{-\kappa_v\triangle}\right)}{\kappa_v^2\Delta^2} \tag{22}$$

$$\mathbb{E}\left[\overline{V}(1)\overline{V}(2)\right] = v_\infty^2 + \frac{\varepsilon_v^2 v_\infty}{2\kappa_v}\frac{\left(1 - e^{-\kappa_v\triangle}\right)^2}{\kappa_v^2\Delta^2} \tag{23}$$

In order to find the optimal parameters, the sum of deviations between the empirical and theoretical statistics of equations (16) to (20) is minimized:

$$\min_{\Theta} \sum_{i\in\{(16),(17),(18),(19),(20)\}} (empStat_i(\Theta) - theoStat_i(\Theta))^2 \tag{24}$$

where $empStat_i$ denotes the empirical estimator (i.e. the left hand side) of equation i, and $theoStat_i$ the theoretical estimator (i.e. the right hand side). (24) represents the basic optimization problem of how to fit the parameters. This minimization could be avoided if the system of equations (16–20) led to closed-form solutions, but the nonlinearity of the system avoids such a quick path. Equation (24) could be altered by assigning different weights on the specific estimators or completely neglecting one or more of the five estimators. Theoretically, all these approaches would lead to consistent estimators although not efficient ones unless using a suitable more complicated set of weights as in the Generalized Method of Moments. Furthermore, we could also include moments of order higher than 4. However, the theoretical estimators get more complex and contain the parameters with higher exponents. This would make the numerical estimations less stable.

3. Application: Merrill Lynch and the Financial Crisis

In order to show an application of the model, one of the most spectacular stories of the financial crises is scrutinized: the downfall of former Wall-Street giant Merrill Lynch (Note 2). Merrill Lynch stocks have long been deemed a save investment. But in September 2008, the existence of Merrill Lynch could only be saved by the takeover by the Bank of America. The stock price reached its peak in early 2007 being close to 100 USD. After still announcing record earnings in early 2007, Merrill Lynch was severely struck by the subprime crisis. In the summer of 2008, the events speeded up dramatically. Finally, the investment bank with the bull as figurehead fell prey to the bear. On September 15, Bank of America announced its intention to acquire Merrill Lynch. This way, Merrill Lynch was prevented from pursuing the same tragic fate as Lehman Brothers which were filing for insolvency that very same day.

The first period which is to be examined in this example is the time between July 2001 and June 2007. The performance over that time horizon seemed to be satisfying (see Figure 2), yet what was already striking is the fairly high volatility of the stock price. The daily mean of log-returns is 0.0227% (corresponding to an annual return of 5.71% which is calculated assuming 252 trading days per year), the standard deviation is 1.85% (which can be approximated by an annual standard deviation of 29.35%), the skewness is–0.140, and the kurtosis is 6.06. But at first sight, there seemed to be no obvious sign that Merrill Lynch could default so soon.

3.1 Fitting the Model Parameters

The introduced stochastic volatility model requires several parameters which have to be determined before the model can be applied: the time to maturity T, the risk-free rate r, and the level of debt $D(0)$. The time to maturity is the time until the barrier option introduced in Section 2.2 extinguishes. It is therefore assumed to represent the average time to maturity of the company's liabilities. Unfortunately, this information is not provided by Merrill Lynch. For that reason, the time to maturity is assumed to be 5 years, i.e. $T = 5$ to represent the average time of maturity of Merrill Lynch's liabilities. In order to get an appropriate risk-free rate, the average of the 5-year treasury rate over the according time horizon is used. Therefore, $r = 3.93\%$.

Furthermore, the initial debt $D(0)$ is required. In order to gain that information, the adjusted leverage ratio (Note 3) is extracted from the company's annual reports which are summarized in Table 2. The average adjusted leverage ratio over the considered time horizon is 12.7. However, note that Merrill Lynch defines the leverage as assets (not liabilities) divided by equity. Thus, the adjusted leverage ratio according to the common definition,, is 11.7. In order to define the initial value of the debt $D(0)$ assume that the initial value of the equity is $C(0) = 1$. Therefore, $D(0) = 11.7$.

The problem with fitting the model parameters is that the variance process cannot directly be observed, in particular κ_v and ε_v. Thus, a sufficiently large number of data points is required in order to capture the characteristics and influence of these parameters. For finding the model parameters, the procedure works as follows: A value for ε_v is fixed, and the remaining parameters μ, v_∞, and κ_v are fitted. The optimal set of parameters (including the optimal ε_v) is that set which minimizes the error function (i.e. the sum of squared residuals as described in (24) as a global minimum.

The parameters $\Theta = (\mu,\ v_\infty,\ \kappa_v\ \varepsilon_v)$ for the time series of the Merrill Lynch stock between July 2001 and June 2007 are the following:

μ	r	v_∞	κ_v	ε_v
0.040916	0.000323	0.355731	0.5	0.012545

At first sight, the long-term variance seems to be fairly low. Yet, one has to bear in mind that the model has one important assumption which is that the debt is deterministic, i.e. it has no stochastic component and does not allow for a volatility of the returns of the debt. Thus, only the volatility of the equity can influence the volatility of the assets and vice versa. Furthermore, the leverage ratio is fairly high which dilutes the volatility of the equity in the asset process.

The value of κ_v already indicates the necessity to apply the stochastic volatility model, as the parameter is clearly different from 0 ($\kappa_v = 0$, under Equation 4, would make the stochastic volatility model a common Black/Scholes Model). However, as this statement has no scientific relevance, an ARCH-test is applied. This test has been developed by Engle (1982) in order to test time series for homoscedasticity (constant volatility over time). In this case, the hypothesis of constant volatility can be rejected at a significance of $4.8 \cdot 10^{-7}$ emphasizing the use of the stochastic volatility model.

Figure 3 shows the histogram of daily returns of Merrill Lynch between July 2001 and June 2007. Furthermore, the normal distribution with same mean and standard deviation as the historic distritbuion as well as the distribution according to the stochastic volatility model is displayed. For the latter, more than 1 million daily asset returns (1 000 time series each consisting of 1 506 daily returns) have been simulated with the estimated parameters and the according equity returns have been calculated. The graph already shows that the stochastic volatility model maps reality much better than a common Black Scholes model. In order to validate this observation, a chi-square test is applied (see Chernoff and Lehmann, 1954). First, it is tested whether the historic returns follow a normal distribution. The test statistic with value 961.85 is clearly higher than the 95% critical value of the chi-square distribution 98.48 (accounting for two estimated parameters, mean and standard deviation). Therefore, the hypothesis can clearly be rejected. Next, the stochastic volatility model is tested. The resulting value of the test statistic is 89.82, and thus lower than the critical value of 96.22 (now accounting for four estimated parameters). As we cannot reject the hypothesis, we conclude that the stochastic volatility model maps the distribution of historic returns well.

3.2 The Default Risk of Merrill Lynch

One interesting aspect which is worth to study is the default risk of Merrill Lynch. In July 2007, all the information which has been used so far was known. Therefore, the estimated parameters $\Theta = (\mu,\ v_\infty,\ \kappa_v, \varepsilon_v)$ reflect the information given at that point of time. In order to study the default risk, 10 000 underlying asset time series for Merrill Lynch are simulated over a 100-year time horizon applying the estimated parameters μ, v_∞, κ_v, and ε_v starting on July 1, 2007. As already discussed above, it is arguable whether these simulated asset time series reflect Merrill Lynch's true assets. However, a model is always a simplification of reality, and despite the shortcomings of the simulated asset time series, this does not affect the validity of the equity time series gained by that model.

Every simulated asset time series is compared to the debt which is assumed to follow (5). Merrill Lynch would have to declare insolvency once its liabilities exceeded its assets. Therefore, the time of default is calculated as τ according to (7), as the first point of time where the assets fall below the debt. Figure 4 shows the results of that test: As already mentioned, 10 000 asset time series have been simulated daily over a time horizon of 100 years. Figure 4 indicates how many paths default in the respective year, for example, 11 of them already default within the first year which represents July 2007 until June 2008, 34 default in the second year which represents July 2008 until June 2009. The last bar comprises those paths which have not defaulted within the first 100 years (i.e. until June 2107) which are 5 913 paths. As the total of simulated time series is 10 000, the numbers can easily be interpreted as default probabilities in the respective years: The probability for a default amounts to 0.11% in the first year (starting on July 2007) and 0.34% in the second year.

With that information another interesting question can be answered: What is the probability for Merrill Lynch to survive over a certain span of time? This answer is given in Figure 5. Assuming the probabilities for the event of a default given in Figure 4, this figure provides the probability that Merrill Lynch survives until the respective year, which is 1 minus the sum over the default probabilities until that year. As the time series have been simulated daily, Figure 5 provides an almost continuous function. This graph tells that, for example, the probability to survive 10 years is 88.88%, and the probability to survive 100 years is 59.13%.

Figure 4 already points out that there is a significant default probability for Merrill Lynch in the years from July 2007 on, given the information described above. The probability for a default until June 2009 would be estimated to be 0.45% by this model. Although this number still seems to be fairly low (but is not negligible though), it can already be concluded that even Merrill Lynch bears a certain default risk. However, the takeover by the Bank of America did not mean a default for Merrill Lynch. Merrill Lynch accepted an offer by the Bank of America in order to overcome its own default risk, and the Bank of America finally paid 50 billion USD for that acquisition according to the Wall Street Journal.

3.3 Development of Default Risk

It is even more interesting to compare the findings so far with the condition of Merrill Lynch not before but in the middle of the financial crisis. Therefore, the situation of Merrill Lynch in January 2008 is examined. In order to have a comparable data basis, but which still allows to employ the stochastic volatility model, the time between January 2003 and December 2007 is considered for fitting the parameters. The average 5-year US treasury rate is 3.94%, and from the average adjusted leverage ratio we get $D(0) = 12.6$.

The fitted parameters $\Theta = (\mu, \nu_\infty, \kappa_\nu, \varepsilon_\nu)$ for Merrill Lynch based on the time between January 2003 and December 2007 are the following:

μ	r	ν_∞	κ_ν	ε_ν
0.041350	0.000637	0.595501	0.5	0.024382

Once more, 10 000 asset paths are simulated with those parameters and the event of default is examined. Figure 7 shows the number of defaults in the respective year. In this graph, the first year represents the time from January 2008 until December 2008, the second year is 2009, and so on. Especially when comparing these outcomes to those presented in Figure 4, the highly increased default risk becomes obvious. In the new setting, 0.47% of the simulated time series default within the first year (compared to 0.11%), and 6.27% default within the first three years (compared to 1.40%). In only 41.39% of the time series, the company does not default over the simulated time horizon of 100 years (compared to 59.13%). Figure 8 displays the development of the survival probabilities for the two examined periods (the first year of a possible default starts in July 2007, or January 2008). The survival probability decreases notably. Especially in the first years, the probability to survive is dramatically lower for the situation in January 2008. This means that already on January 1st, 2008 this model only shows a 3:1 chance for Merrill Lynch to survive the next ten years. Therefore, the “unexpected” development of July, August and September 2008 does not really come as such a big surprise anymore. Considering that Merrill Lynch actually has not defaulted in September 2008 when being taken over by the Bank of America, it was almost clear that with these prospects, something “had to happen”.

Furthermore, Figure 8 compares the survival probabilities of this model with the survival probabilities derived from the S&P credit rating of Merrill Lynch. In January 2008, Merrill Lynch was rated “A+”. Assuming that the transition matrix for the rating classes will not change over the next 100 years, yields the survival probabilities indicated by the dotted line. It becomes clear that, in particular, in the near future, the credit rating severely underestimates the probability of a default compared to the credit model under stochastic volatility.

4. Conclusion

This paper presented a model for a company’s asset process if only the equity process is observable. Companies’ equity is regarded as a barrier option, namely a down-and-out call on the company’s assets where barrier and strike price are equal to the debt. The assets are modeled with stochastic volatility providing for more degrees of freedom than a common Black/Scholes Model. Thus, this paper combines the credit model as in Black and Cox (1976) with the stochastic volatility model in Heston (1993). As a first result, a closed-form solution for the price of a down-and-out call option on the assets with the debt as barrier and strike price has been derived. Second, a method of gaining consistent estimators was derived from Genon-Catalot et al. (2000), and five estimators were given explicitly which are utilized during the fitting process for single time series. Given the equity value of the company and information on the liabilities, we can derive an asset process where the assets have stochastic volatility. These contributions provide a computationally and statistically friendly setting for practitioners in economics and finance.

The applications in this study have been limited to one dimension. One potential extension of this research is to look at several companies and estimate their dependence structure within our stochastic volatility setting. Furthermore, it would also be of interest to examine a setting in the stochastic volatility model where the barrier is not equal to the strike price and/or where the debt is assumed to follow a stochastic process to account for uncertainty on the liability side as well. Yet, this would probably make it a lot more difficult to arrive at (closed-form) formulas for option prices or to obtain estimators with sound statistical properties. However, models with a higher complexity than the Black/Scholes Model where assets follow a Geometric Brownian Motion are very important, in particular stochastic volatility models which are able to better capture realistic features of market behavior, and thus better describe reality.

Appendix

Derivation of Option Pricing Formula The following lemma is stated from Zagst (2002, Theorem 2.21, p. 21):

Lemma 5.3 (Optional Sampling) *Let* $(\Omega, \mathcal{F}, \mathbb{Q}, \mathbb{F})$ *be a filtered probability space and* $X = (X_t)_{t\geq 0}$ *a right-continuous martingale which means that all paths of* X_t *are right-continuous. Furthermore, let* τ *be a stopping time and* $0 \leq s \leq t < \infty$. *Then, for all* $t \geq 0$ *it holds* $\mathbb{Q}$*-a.s. that*

$$\mathbb{E}\left[X_{t\wedge\tau} \mid \mathcal{F}_s\right] = X_{s\wedge\tau} \tag{25}$$

where the stopped process $X_{t\wedge\tau}$ *is defined as*

$$X_{t\wedge\tau}(\omega) := \begin{cases} X_t(\omega) & \text{if} \quad t \leq \tau(\omega) \\ X_\tau(\omega) & \text{if} \quad t > \tau(\omega) \end{cases}$$

In the present scenario, the option (6) can be rewritten as

$$\begin{aligned} C(t, A) &= \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_t^T r(s)\,ds} \cdot \max\{A(T) - D(T),\, 0\} \cdot \mathbf{1}_{\{\tau>T\}} \,\middle|\, \mathcal{F}_t\right] \\ &= \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_t^T r(s)\,ds} \cdot (A(T) - D(T)) \cdot \mathbf{1}_{\{\tau>T\}} \,\middle|\, \mathcal{F}_t\right] \end{aligned} \tag{26}$$

because if the asset value is below the value of the debt at maturity, i.e. $A(T) - D(T) < 0$, the indicator function $\mathbf{1}_{\{\tau>T\}}$ with $\tau = \inf\{t' \;:\; A(t') < D(t')\}$ already is zero. Thus, taking the maximum becomes redundant.

Now define

$$X_t = X(t) := \tilde{A}(t) - \tilde{D}(t) \tag{27}$$

with the discounted processes

$$\begin{aligned} \tilde{A}(t) &:= e^{-\int_0^t r(s)\,ds} A(t) \\ \tilde{D}(t) &:= e^{-\int_0^t r(s)\,ds} D(t) \end{aligned}$$

for assets and debt. As $A(t)$ and $D(t)$ are continuous X_t is continuous, and thus, $X_{t\wedge\tau}$ is continuous as well, in particular for $t = \tau$. As a tradable underlying $\tilde{A}(t)$ can be assumed a $\mathbb{Q}$-martingale. Therefore, X_t is a martingale. The stopped process is

$$\begin{aligned} X_{t\wedge\tau}(\omega) &= \begin{cases} \tilde{A}(t) - \tilde{D}(t) & \text{if} \quad t \leq \tau(\omega) \\ 0 & \text{if} \quad t > \tau(\omega) \end{cases} \\ &= \left(\tilde{A}(t) - \tilde{D}(t)\right) \cdot \mathbf{1}_{\{\tau>t\}} \end{aligned}$$

Using the stopped process, the option price (27) is:

$$\begin{aligned} C(t, A) &= \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_t^T r(s)\,ds} \cdot (A(T) - D(T)) \cdot \mathbf{1}_{\{\tau>T\}} \,\middle|\, \mathcal{F}_t\right] \\ &= e^{\int_0^t r(s)\,ds} \cdot \mathbb{E}_{\mathbb{Q}}\left[\left(\tilde{A}(T) - \tilde{D}(T)\right) \cdot \mathbf{1}_{\{\tau>T\}} \,\middle|\, \mathcal{F}_t\right] \\ &= e^{\int_0^t r(s)\,ds} \cdot \mathbb{E}_{\mathbb{Q}}\left[X_{T\wedge\tau} \mid \mathcal{F}_t\right] \\ &\underset{(25)}{=} e^{\int_0^t r(s)\,ds} \cdot X_{t\wedge\tau} \\ &= e^{\int_0^t r(s)\,ds} \cdot \left(\tilde{A}(t) - \tilde{D}(t)\right) \cdot \mathbf{1}_{\{\tau>t\}} \\ &= (A(t) - D(t)) \cdot \mathbf{1}_{\{\tau>t\}} \end{aligned} \tag{28}$$

Given that the company didn't default up to time t, i.e. $\mathbf{1}_{\{\tau>t\}} = 1$,

$$C(t, A) = A(t) - D(t). \tag{29}$$

Definition 5.4 (α-mixing) Let X be a stochastic process (which is a sequence of random variables $\{X_k\}_{k\in\mathbb{Z}}$) on a probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ and define

$$\mathcal{F}_A^B := \sigma\left(X_k,\, k \in [A, B]\right)$$

for $-\infty \leq A \leq B \leq \infty$. $\sigma(E)$ denotes the generated σ-algebra with generator E.

Furthermore,

$$\alpha(n) := \sup_{j\in\mathbb{Z}} \alpha\left(\mathcal{F}_{-\infty}^{j}, \mathcal{F}_{j+n}^{\infty}\right).$$

with

$$\alpha(\mathcal{F}_1, \mathcal{F}_2) := \sup_{E_1\in\mathcal{F}_1, E_2\in\mathcal{F}_2} |\mathbb{Q}(E_1 \cap E_2) - \mathbb{Q}(E_1)\cdot\mathbb{Q}(E_2)|$$

The process X is α-mixing, if

$$\alpha(n) \overset{n\to\infty}{\longrightarrow} 0$$

Proof. **(Lemma 2.1)** The stochastic differential equation (11) can be re-written as:

$$Y(t) = y_0 + \int_0^t \mu - \frac{v(s)}{2}\, ds + \int_0^t \sqrt{v(s)}dZ(s)$$

Therefore,

$$\begin{aligned} Y(i\triangle) - Y((i-1)\triangle) &= \int_{(i-1)\triangle}^{i\triangle} \mu - \frac{v(s)}{2}\, ds + \int_{(i-1)\triangle}^{i\triangle} \sqrt{v(s)}dZ(s) \\ &= \mu\,\triangle - \int_{(i-1)\triangle}^{i\triangle} \frac{v(s)}{2}\, ds + \int_{(i-1)\triangle}^{i\triangle} \sqrt{v(s)}dZ(s) \end{aligned}$$

and

$$\begin{aligned} \frac{Y(i\triangle) - Y((i-1)\triangle)}{\sqrt{\triangle}} &= \mu\sqrt{\triangle} - \frac{1}{\sqrt{\triangle}}\int_{(i-1)\triangle}^{i\triangle} \frac{v(s)}{2}\, ds + \frac{1}{\sqrt{\triangle}}\int_{(i-1)\triangle}^{i\triangle} \sqrt{v(s)}dZ(s) \\ &= \left(\mu - \frac{\overline{V}(i)}{2}\right)\cdot\sqrt{\triangle} + \frac{1}{\sqrt{\triangle}}\int_{(i-1)\triangle}^{i\triangle} \sqrt{v(s)}dZ(s) \end{aligned}$$

Proof. **(Proposition 2.2)** The proof of this proposition works along Sections 3 and 4 of Genon-Catalot (2000). According to them, the volatility process v which is characterized by (2) is α-mixing if the following conditions hold, as shown in detail in the example in their Section 4.2:

$$\kappa_v > 0 \tag{30}$$

$$\kappa_v v_\infty \geq \frac{\varepsilon_v^2}{2} \tag{31}$$

If this is the case, due to Proposition 3.2 in Genon-Catalot (2000) (Note 4), $v(t)$ in (2) is also α-mixing. It follows that:

$$\frac{1}{n}\sum_{i=0}^{n-2} f(R(i+1), R(i+2)) \overset{n\to\infty}{\longrightarrow} \mathbb{E}[\mathbb{E}[f(R(1), R(2))]] \quad a.s. \tag{32}$$

where f is chosen (Note 5) to be

$$f(A, B) = A^p \cdot B^q$$

In particular $(p, q) \in \{(1, 0), (1, 1), (2, 0), (2, 1), (4, 0)\}$. Therefore, $\frac{1}{n}\sum_{i=0}^{n-1} R(i+1) \overset{a.s.}{\longrightarrow} \mathbb{E}[R(1)]$, $\frac{1}{n}\sum_{i=0}^{n-2} R(i+1)R(i+2) \overset{a.s.}{\longrightarrow}$ $\mathbb{E}[R(1)R(2)]$, and so on.

$$\begin{aligned} \mathbb{E}[R(1)] &= \mathbb{E}\left[\mathbb{E}\left[R(1)|\overline{V}\right]\right] \\ &\underset{(15)}{=} \mathbb{E}\left[\sqrt{\triangle}\mu - \frac{1}{2}\sqrt{\triangle}\,\overline{V}(1)\right] \\ &= \sqrt{\triangle}\mu - \frac{1}{2}\sqrt{\triangle}\,\mathbb{E}\left[\overline{V}(1)\right] \end{aligned} \tag{33}$$

$$\begin{aligned}
\mathbb{E}[R(1)R(2)] &= \mathbb{E}\left[\mathbb{E}\left[R(1)R(2)|\overline{V}\right]\right] \\
&= \mathbb{E}\left[\mathbf{Cov}(R(1),\, R(2)|\overline{V}) + \mathbb{E}\left[R(1)|\overline{V}\right]\cdot\mathbb{E}\left[R(2)|\overline{V}\right]\right] \\
&= \mathbb{E}\left[0 + \left(\sqrt{\triangle}\mu - \frac{1}{2}\sqrt{\triangle}\overline{V}(1)\right)\left(\sqrt{\triangle}\mu - \frac{1}{2}\sqrt{\triangle}\overline{V}(2)\right)\right] \\
&= \mathbb{E}\left[\triangle\mu^2 - \frac{1}{2}\triangle\mu\overline{V}(1) - \frac{1}{2}\triangle\mu\overline{V}(2) + \frac{1}{4}\triangle\overline{V}(1)\overline{V}(2)\right] \\
&= \triangle\mu^2 - \frac{1}{2}\triangle\mu\mathbb{E}\left[\overline{V}(1)\right] - \frac{1}{2}\triangle\mu\mathbb{E}\left[\overline{V}(2)\right] + \frac{1}{4}\triangle\mathbb{E}\left[\overline{V}(1)\overline{V}(2)\right] \\
&= \triangle\mu^2 - \triangle\mu\mathbb{E}\left[\overline{V}(1)\right] + \frac{1}{4}\triangle\mathbb{E}\left[\overline{V}(1)\overline{V}(2)\right] \qquad (34)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[R^2(1)\right] &= \mathbb{E}\left[\mathbb{E}\left[R^2(1)|\overline{V}\right]\right] \\
&= \mathbb{E}\left[\mathbf{Var}(R(1)|\overline{V}) + \mathbb{E}\left[R(1)|\overline{V}\right]^2\right] \\
&= \mathbb{E}\left[\overline{V}(1) + \left(\sqrt{\triangle}\mu - \frac{1}{2}\sqrt{\triangle}\overline{V}(1)\right)^2\right] \\
&= \mathbb{E}\left[\overline{V}(1) + \triangle\mu^2 - \triangle\mu\overline{V}(1) + \frac{1}{4}\triangle\overline{V}^2(1)\right] \\
&= \triangle\mu^2 - (\triangle\mu - 1)\mathbb{E}\left[\overline{V}(1)\right] + \frac{1}{4}\triangle\mathbb{E}\left[\overline{V}^2(1)\right] \qquad (35)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[R(1)^2R(2)\right] &= \mathbb{E}\left[\mathbb{E}\left[R(1)^2R(2)|\overline{V}\right]\right] \\
&= \mathbb{E}\left[\mathbf{Cov}(R(1)^2,\, R(2)|\overline{V}) + \mathbb{E}\left[R(1)^2|\overline{V}\right]\cdot\mathbb{E}\left[R(2)|\overline{V}\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[R(1)^2|\overline{V}\right]\right]\cdot\mathbb{E}\left[\mathbb{E}\left[R(2)|\overline{V}\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[R(1)^2|\overline{V}\right]\right]\cdot\mathbb{E}\left[\mathbb{E}\left[R(1)|\overline{V}\right]\right] \\
&\underset{(33),(35)}{=} \left(\triangle\mu^2 - (\triangle\mu - 1)\mathbb{E}\left[\overline{V}(1)\right] + \frac{1}{4}\triangle\mathbb{E}\left[\overline{V}^2(1)\right]\right) \\
&\quad \cdot\left(\sqrt{\triangle}\mu - \frac{1}{2}\sqrt{\triangle}\mathbb{E}\left[\overline{V}(1)\right]\right) \qquad (36)
\end{aligned}$$

For the derivation of $\mathbb{E}\left[R^4(1)\right]$ it is used that if X is normally distributed, then the fourth moment of X is

$$\mathbb{E}[X^4] = \mathbb{E}[X]^4 + 6\cdot\mathbb{E}[X]^2\cdot\mathbf{Var}(X) + 3\cdot\mathbf{Var}(X)^2 \qquad (37)$$

$$\begin{aligned}
\mathbb{E}[R^4(1)] &= \mathbb{E}\left[\mathbb{E}[R^4(1)|\overline{V}]\right] \\
&= \mathbb{E}\left[\underbrace{\mathbb{E}[R(1)|\overline{V}]^4}_{\text{(I)}} + \underbrace{6\cdot\mathbb{E}[R(1)|\overline{V}]^2\cdot\mathbf{Var}(R(1)|\overline{V})}_{\text{(II)}} + \underbrace{3\cdot\mathbf{Var}(R(1)|\overline{V})^2}_{\text{(III)}}\right] \\
&= \mathbb{E}\left[\underbrace{\left(\triangle\mu^2 - \triangle\mu\overline{V}(1) + \frac{1}{4}\triangle\overline{V}^2(1)\right)^2}_{\text{(I)}}\right. \\
&\qquad \left. \underbrace{+\,6\left(\triangle\mu^2 - \triangle\mu\overline{V}(1) + \frac{1}{4}\triangle\overline{V}^2(1)\right)\overline{V}(1)}_{\text{(II)}} + \underbrace{3\overline{V}^2(1)}_{\text{(III)}}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\Bigg[\underbrace{\triangle^2\mu^4 + \triangle^2\mu^2\overline{V}^2(1) + \frac{1}{16}\triangle^2\,\overline{V}^4(1) - 2\triangle^2\mu^3\overline{V}(1) + \frac{1}{2}\triangle^2\mu^2\overline{V}^2(1)}_{\text{(I)}} \\
&\qquad \underbrace{-\frac{1}{2}\triangle^2\mu\overline{V}^3(1)}_{\text{(I)}} + \underbrace{6\triangle\mu^2\overline{V}(1) - 6\triangle\mu\overline{V}^2(1) + \frac{6}{4}\triangle\overline{V}^3(1)}_{\text{(II)}} + \underbrace{3\overline{V}^2(1)}_{\text{(III)}}\Bigg] \\
&\approx \triangle^2\mu^4 + \left(6\triangle\mu^2 - 2\triangle^2\mu^3\right)\mathbb{E}\left[\overline{V}(1)\right] \\
&\quad + \left(\frac{3}{2}\triangle^2\mu^2 - 6\triangle\mu + 3\right)\mathbb{E}\left[\overline{V}^2(1)\right] \qquad (38)
\end{aligned}
$$

Note that the influence of those terms containing $\mathbb{E}\left[\overline{V}^3(1)\right]$ or $\mathbb{E}\left[\overline{V}^4(1)\right]$ is negligible. Summing up, (33), (34), (35), (36) and (38) prove (16), (17), (18), (19) and (20).

(21), (22) and (23) have already been calculated in Genon-Catalot (2000).

References

Black, F., & Cox, J. (1976). Valuing Corporate Securities: Some Effects of Bond Indenture Provisions. *Journal of Finance, 31*(2), 351-367. http://dx.doi.org/10.2307/2326607

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics, 31*, 307-327. http://dx.doi.org/10.1016/0304-4076(86)90063-1

Bradley, R. (2005). Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. *Probability Surveys, 2*, 107-144.

Brockwell, P., Chadraa, E., & Lindner, A. (2006). Continuous time GARCH processes. *The Annals of Applied Probability, 16*(2), 790-826. http://dx.doi.org/10.1214/105051606000000150

Chernoff, H., & Lehmann, E. (1954). The Use of Maximum Likelihood Estimates in Chi-Square Tests for Goodness of Fit. *The Annals of Mathematical Statistics, 25*(3), 579-586. http://dx.doi.org/10.1214/aoms/1177728726

Cox, J., Ingersoll, J., & Ross, S. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica, 2*, 385-408. http://dx.doi.org/10.2307/1911242

Davidson, R., & MacKinnon, J. (1993). *Estimation and Inference in Econometrics*. Oxford University Press.

Embrechts, P., Frey, R., & McNeil, A. (2005). *Quantitative Risk Management*. Princeton University Press.

Engle, R. (1982). Autoregressive Contidional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica, 50*(4), 987-1008. http://dx.doi.org/10.2307/1912773

Escobar, M., Friederich, T., Seco, L., & Zagst, R. (2011). A General Structural Approach for Credit Modeling under Stochastic Volatility. *Journal of Financial Transformation, 32*, 123-132.

Feller, W. (1951). Two Singular Diffusion Problems. *The Annals of Mathematics, 54*(1), 173-182. http://dx.doi.org/10.2307/1969318

Genon-Catalot, V., Jeantheau, T., & Laredo, C. (2000). Stochastic Volatility Models as Hidden Markov Models and Statistical Applications. *Bernoulli, 60*(6), 1051-1079. http://dx.doi.org/10.2307/3318471

Heston, S. (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies, 6*(2), 327-343. http://dx.doi.org/10.1093/rfs/6.2.327

Merton, R. (1973). Theory of Rational Option Pricing. *Bell Journal of Economics and Management Science, 4*, 141-183. http://dx.doi.org/10.2307/3003143

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences, 42*, 43-47. http://dx.doi.org/10.1073/pnas.42.1.43

Zagst, R. (2002). *Interest Rate Models*. Springer Finance.

Table 1. Numerical and analytical pricing of a DOC in the stochastic volatility model

D(t)	zero payoffs	C_{num} mean	C analytical	diff.	sigma	95% confidence lower	95% confidence upper
0.00	0	99.936	100.000	–0.064	22.507	99.494	100.377
10.00	0	89.877	90.000	–0.123	22.749	89.431	90.323
20.00	0	79.845	80.000	–0.155	22.624	79.402	80.288
30.00	0	70.118	70.000	0.118	22.864	69.670	70.566
40.00	0	60.350	60.000	0.350	22.649	59.906	60.794
50.00	14	49.776	50.000	–0.224	22.026	49.345	50.208
60.00	288	40.240	40.000	0.240	22.888	39.792	40.689
70.00	1 291	29.737	30.000	–0.263	22.098	29.304	30.170
80.00	3 479	20.163	20.000	0.163	21.617	19.739	20.587
90.00	6 637	10.003	10.000	0.003	17.764	9.655	10.352

Table 2. Consolidated financial data of Merrill Lynch 2000–2008

*(in million USD)	2000	2001	2002	2003	2004	2005	2006	2007	2008
EBIT*	5 717	1 377	3 757	5 649	5 836	7 231	10 426	–12 831	–41 831
Net Revenues*	26 766	21 880	18 608	20 154	22 023	26 009	34 659	11 250	–12 593
Total Assets*	407 200	419 419	447 928	494 518	648 059	681 015	841 299	1 020 050	667 543
Total Liabilities*	386 182	396 716	422 395	464 197	616 689	645 415	802 261	988 118	647 540
Stockholders' Equity*	18 304	20 008	22 875	27 651	31 370	35 600	39 038	31 932	20 003
Adjusted Leverage	13.2	13.1	11.1	11.7	13.9	11.6	13.1	17.7	13.3

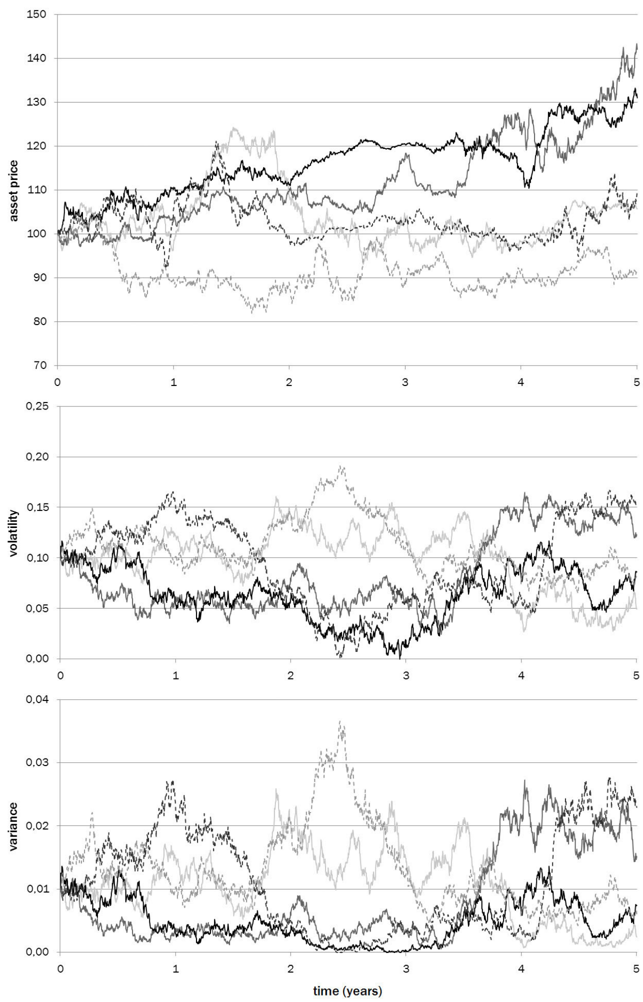


Figure 1. Five simulated paths in the stochastic volatility model

Figure 2. Performance of Merrill Lynch Stocks on NYSE (July 2001 – June 2007)

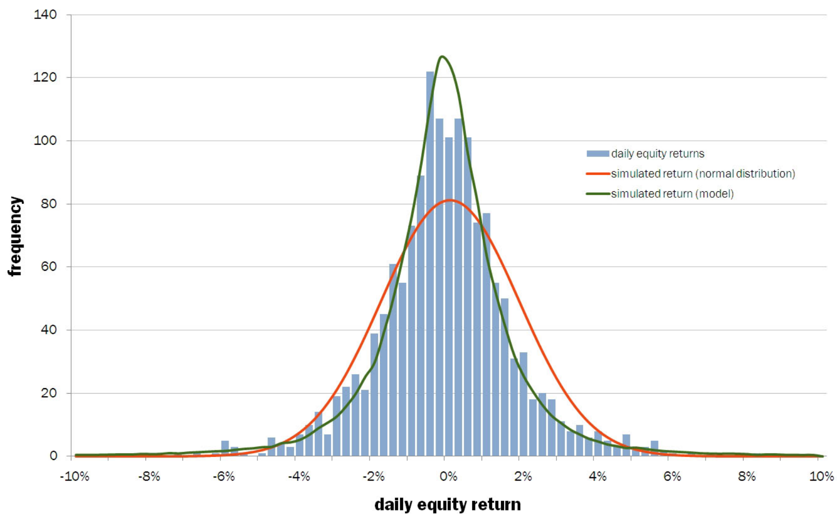


Figure 3. Observed equity returns of Merrill Lynch (July 2001 – June 2007) and simulated returns

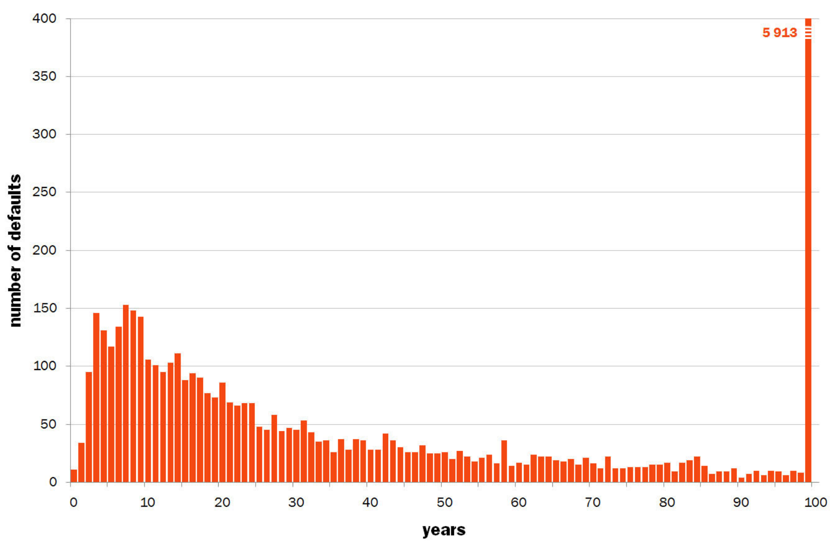


Figure 4. Number of defaults out of 10 000 simulated time series for Merrill Lynch (July 2007)

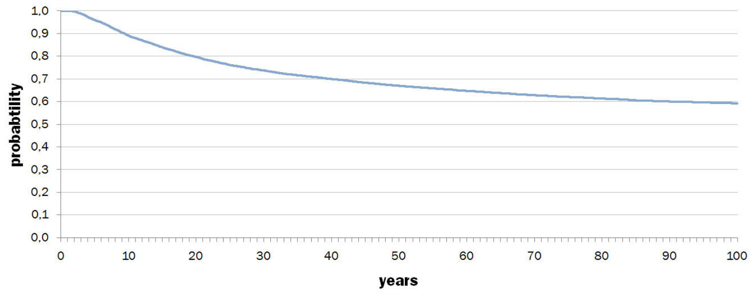


Figure 5. Survival probability for Merrill Lynch (July 2007)

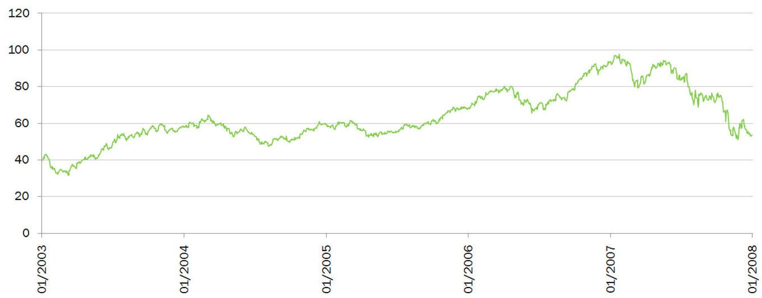


Figure 6. Performance of Merrill Lynch Stocks on NYSE (January 2003 – December 2007)

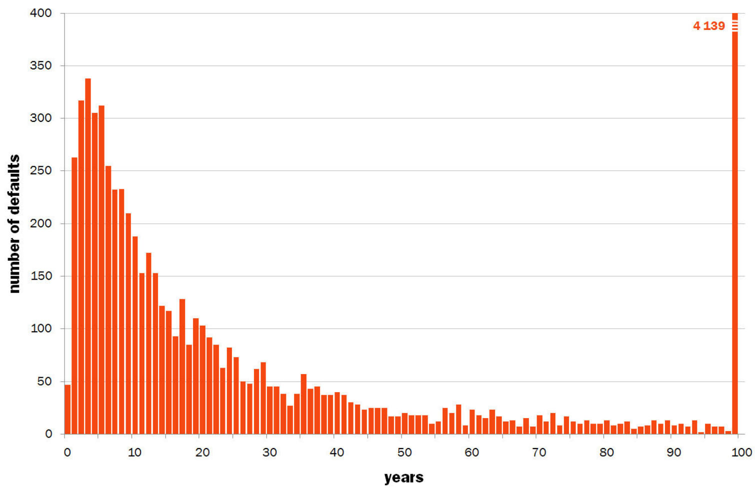


Figure 7. Number of defaults out of 10 000 simulated time series for Merrill Lynch (January 2008)

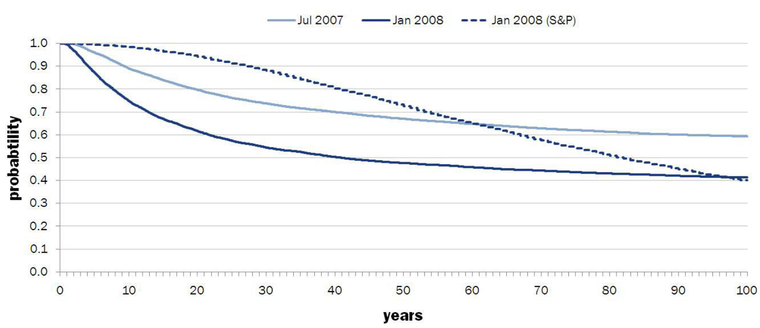


Figure 8. Survival probability for Merrill Lynch (July 2007 and January 2008)

Notes

Note 1. The basic idea of a barrier option is that the option expires if the underlying hits a barrier ("out options") or becomes exercisable only if such a barrier is reached ("in options"). A down-and-out call option therefore describes a call option which expires if the underlying falls below a certain barrier. Thus, barrier options are path-dependent.

Note 2. In order to accurately fit the parameters of the volatility process which is not directly observable, we need sufficiently many data points. Thus, daily data is used in the example.

Note 3. The adjusted leverage ratio is defined in the annual reports as "assets reduced by securities financing transactions and securities received as collateral less trading liabilities net of derivative contracts and segregated cash and securities and separate accounts assets, [...] divided by equity capital."

Note 4. Although it is not stated in the proof of that proposition, it may be noted that the basic for this finding has already been laid by Rosenblatt (1956).

Note 5. The Proposition in Genon-Catalot (2000) deals with a more general class of functions.

Measures on Proportional Reduction in Error by Arithmetic, Geometric and Harmonic Means for Multi-way Contingency Tables

Kouji Yamamoto (Corresponding author)
Center for Clinical Investigation and Research, Osaka University Hospital
2-15, Yamadaoka, Suita 565-0871, Osaka, Japan
E-mail: yamamoto-k@hp-crc.med.osaka-u.ac.jp

Yuri Nozaki
Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science
2641, Yamazaki, Noda City, Chiba 278-8510, Japan
E-mail: yr.nozaki@gmail.com

Sadao Tomizawa
Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science
2641, Yamazaki, Noda City, Chiba 278-8510, Japan
E-mail: tomizawa@is.noda.tus.ac.jp

Received: December 27, 2011 Accepted: January 11, 2012 Published: May 1, 2012
doi:10.5539/ijsp.v1n1p36 URL: http://dx.doi.org/10.5539/ijsp.v1n1p36

Abstract

For multi-way contingency tables with nominal categories, this paper proposes three kinds of proportional reduction in error measures, which describe the relative decrease in the probability of making an error in predicting the value of one variable when the values of the other variables are known, as opposed to when they are not known. The measures have forms of arithmetic, geometric and harmonic means. An example is shown.

Keywords: Arithmetic mean, Geometric mean, Harmonic mean, Proportional reduction in error

1. Introduction

Consider an $R \times C$ contingency table with both nominal categories of the explanatory variable X and the response variable Y. Let p_{ij} denote the probability that an observation will fall in the ith category of X and in the jth category of Y ($i = 1, \ldots, R$; $j = 1, \ldots, C$). Goodman and Kruskal (1954) proposed two kinds of measures, i.e., (1) the measure which describes the proportional reduction in variation (PRV) in predicting the Y category obtained when the X category is known, as opposed to when the X category is not known, and (2) the measure which describes the proportional reduction in error (PRE) in predicting it. Although the details are omitted, some PRV measures are considered by, e.g., Theil (1970), Tomizawa, Seo and Ebi (1997), Tomizawa and Ebi (1998), Tomizawa and Yukawa (2003), and Yamamoto, Miyamoto and Tomizawa (2010).

The present paper considers the PRE measures. Goodman and Kruskal (1954) proposed the PRE measure as

$$\lambda_B = \frac{(1 - p_{\bullet m_0}) - \sum_{i=1}^{R} p_{i\bullet}\left(1 - \left(\frac{p_{im_i}}{p_{i\bullet}}\right)\right)}{1 - p_{\bullet m_0}} = \frac{\sum_{i=1}^{R} p_{im_i} - p_{\bullet m_0}}{1 - p_{\bullet m_0}},$$

where

$$p_{im_i} = \max_j (p_{ij}), \quad p_{\bullet m_0} = \max_j (p_{\bullet j}), \quad p_{i\bullet} = \sum_{t=1}^{C} p_{it}, \quad p_{\bullet j} = \sum_{s=1}^{R} p_{sj};$$

also see Bishop, Fienberg and Holland (1975, p. 388), and Everitt (1992, p. 58). This measure describes the relative decrease in the probability of making an error in predicting the value of Y when the value of X is known, as opposed to when it is not known. The measure λ_B has the properties that (i) $0 \leq \lambda_B \leq 1$, (ii) $\lambda_B = 0$ if and only if the information about the explanatory variable X does not reduce the probability of making an error in predicting the category of the variable Y, and (iii) $\lambda_B = 1$ if and only if no error is made, given knowledge of the explanatory variable X; namely there is complete predictive association.

Next, consider the reverse case which is the explanatory variable Y and the response variable X. The following measure λ_A is suitable for predictions of X from Y, defined by

$$\lambda_A = \frac{\sum_{j=1}^{C} p_{M_j j} - p_{M_0 \bullet}}{1 - p_{M_0 \bullet}},$$

where

$$p_{M_j j} = \max_i(p_{ij}), \quad p_{M_0 \bullet} = \max_i(p_{i\bullet});$$

see Goodman and Kruskal (1954).

The measures λ_B and λ_A are specifically designed for the situation in which the explanatory and response variables are defined. Now consider the situation where the explanatory and response variables are not defined. In this case, the following measure λ is given:

$$\lambda = \frac{\sum_{i=1}^{R} p_{i m_i} + \sum_{j=1}^{C} p_{M_j j} - p_{\bullet m_0} - p_{M_0 \bullet}}{2 - p_{\bullet m_0} - p_{M_0 \bullet}};$$

see Goodman and Kruskal (1954). This indicates the PRE in predicting the category of either variable as between knowing and not knowing the category of the other variable. Also, the measure λ is the weighted sum of the measures λ_B and λ_A.

For a two-way contingency table with both nominal categories, Yamamoto and Tomizawa (2010) proposed new PRE measures, say Λ, expressed as the arithmetic, geometric and harmonic means of λ_B and λ_A. For a two-way contingency table with nominal-ordinal categories, Yamamoto, Nozaki and Tomizawa (2011) proposed a PRE measure although the detail is omitted.

The purpose of the present paper is to extend the Yamamoto and Tomizawa's (2010) measures into T-way contingency tables ($T \geq 3$) with all nominal categories. Section 2 proposes measures for three-way tables ($T = 3$), and Section 3 extends them for multi-way ($T \geq 4$) and expresses as more generalized form including such three kinds of means. Section 4 analyzes data as an example.

2. New PRE Measures for Three-way Contingency Tables

2.1 Measures

Consider an $R \times C \times L$ contingency table with variables X, Y and Z which have all nominal categories. Let p_{ijk} denote the probability of that an observation will fall in the (i, j, k)th cell of the table ($i = 1, \ldots, R; j = 1, \ldots, C; k = 1, \ldots, L$). When the explanatory and response variables are not defined, namely, we cannot specifically define which of the variables is a response, we consider three kinds of prediction, predicting X, predicting Y and predicting Z.

First, consider the table with a response variable X and two explanatory variables Y and Z. In this case, a PRE measure, which describes the relative decrease in the probability of making error in predicting the value of X when the values of the other variables, Y and Z, are known, as opposed to when they are not known, is defined by

$$\lambda_A^{(3)} = \frac{\sum_{j=1}^{C} \sum_{k=1}^{L} p_{m_{jk} jk} - p_{m_1 \bullet\bullet}}{1 - p_{m_1 \bullet\bullet}},$$

where

$$p_{m_{jk} jk} = \max_i(p_{ijk}), \quad p_{m_1 \bullet\bullet} = \max_i(p_{i\bullet\bullet}), \quad p_{i\bullet\bullet} = \sum_{t=1}^{C} \sum_{u=1}^{L} p_{itu}.$$

Similarly, each PRE measure for the table as having a response variable Y and two explanatory variables X and Z and as having a response variable Z and two explanatory variables X and Y is defined by

$$\lambda_B^{(3)} = \frac{\sum_{i=1}^{R}\sum_{k=1}^{L} p_{im_{ik}k} - p_{\bullet m_2 \bullet}}{1 - p_{\bullet m_2 \bullet}},$$

and

$$\lambda_C^{(3)} = \frac{\sum_{i=1}^{R}\sum_{j=1}^{C} p_{ijm_{ij}} - p_{\bullet\bullet m_3}}{1 - p_{\bullet\bullet m_3}},$$

where

$$p_{im_{ik}k} = \max_j(p_{ijk}), \quad p_{\bullet m_2 \bullet} = \max_j(p_{\bullet j \bullet}), \quad p_{\bullet j \bullet} = \sum_{s=1}^{R}\sum_{u=1}^{L} p_{sju},$$

$$p_{ijm_{ij}} = \max_k(p_{ijk}), \quad p_{\bullet\bullet m_3} = \max_k(p_{\bullet\bullet k}), \quad p_{\bullet\bullet k} = \sum_{s=1}^{R}\sum_{t=1}^{C} p_{stk}.$$

Then, we shall propose three kinds of new PRE measures as follows:

$$\lambda_a^{(3)} = \frac{\lambda_A^{(3)} + \lambda_B^{(3)} + \lambda_C^{(3)}}{3},$$

$$\lambda_g^{(3)} = \sqrt[3]{\lambda_A^{(3)}\lambda_B^{(3)}\lambda_C^{(3)}},$$

and

$$\lambda_h^{(3)} = \frac{3}{\dfrac{1}{\lambda_A^{(3)}} + \dfrac{1}{\lambda_B^{(3)}} + \dfrac{1}{\lambda_C^{(3)}}}.$$

The measures $\lambda_a^{(3)}$, $\lambda_g^{(3)}$ and $\lambda_h^{(3)}$ are the arithmetic mean, geometric mean and harmonic mean of the $\lambda_A^{(3)}$, $\lambda_B^{(3)}$ and $\lambda_C^{(3)}$, respectively.

Let λ^* denote each of measures $\lambda_a^{(3)}$, $\lambda_g^{(3)}$ and $\lambda_h^{(3)}$. Each measure has the properties that (i) λ^* must lie between 0 and 1, (ii) $\lambda^* = 0$ if and only if the information about two variables does not reduce the probability of making an error in predicting the category of the other variable, and (iii) $\lambda^* = 1$ if and only if no error is made, given knowledge of two variables; namely there is complete predictive association. We point out that if the variables are independent, then the measure λ^* takes 0, but the converse need not hold. Note that when the values of $\lambda_A^{(3)}$, $\lambda_B^{(3)}$ and $\lambda_C^{(3)}$ are 0 such as the variables are independent, the measure $\lambda_h^{(3)}$ cannot measure the PRE. So in such a case, the measures $\lambda_a^{(3)}$ and $\lambda_g^{(3)}$ should be used as a PRE measure.

We see that

$$\min\left(\lambda_A^{(3)}, \lambda_B^{(3)}, \lambda_C^{(3)}\right) \le \lambda_h^{(3)} \le \lambda_g^{(3)} \le \lambda_a^{(3)} \le \max\left(\lambda_A^{(3)}, \lambda_B^{(3)}, \lambda_C^{(3)}\right),$$

where the equality holds if and only if $\lambda_A^{(3)} = \lambda_B^{(3)} = \lambda_C^{(3)}$.

2.2 Approximate Confidence Interval for Measures

Let n_{ijk} denote the observed frequency in the (i, j, k)th cell of the table ($i = 1, \ldots, R; j = 1, \ldots, C; k = 1, \ldots, L$). Assuming that $\{n_{ijk}\}$ result from full multinomial sampling, we consider an approximate standard error and large-sample confidence interval for λ^*, using the delta method, descriptions of which are given by Bishop et al. (1975, Sec. 14.6). The sample version of λ^*, i.e., $\hat{\lambda}^*$, is given by λ^* with $\{p_{ijk}\}$ replaced by $\{\hat{p}_{ijk}\}$, where $\hat{p}_{ijk} = n_{ijk}/n$ and $n = \sum\sum n_{ijk}$. Using the delta method, $\sqrt{n}(\hat{\lambda}^* - \lambda^*)$ has asymptotically (as $n \to \infty$) a normal distribution with mean 0 and variance $\sigma^2[\lambda^*]$, where

$$\sigma^2[\lambda^*] = \sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{L}\left(\frac{\partial \lambda^*}{\partial p_{ijk}}\right)^2 p_{ijk} - \left(\sum_{s=1}^{R}\sum_{t=1}^{C}\sum_{u=1}^{L}\left(\frac{\partial \lambda^*}{\partial p_{stu}}\right) p_{stu}\right)^2.$$

For measures $\lambda_a^{(3)}$, $\lambda_g^{(3)}$ and $\lambda_h^{(3)}$, the variances are

$$\text{(a)}\ \sigma^2\left[\lambda_a^{(3)}\right] = \frac{1}{9}\left[\sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{L}(U_{ijk})^2 p_{ijk} - \left(\sum_{s=1}^{R}\sum_{t=1}^{C}\sum_{u=1}^{L} U_{stu}p_{stu}\right)^2\right],$$

$$\text{(b)}\ \sigma^2\left[\lambda_g^{(3)}\right] = \frac{1}{9}\left(\lambda_g^{(3)}\right)^{-4}\left[\sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{L}(V_{ijk})^2 p_{ijk} - \left(\sum_{s=1}^{R}\sum_{t=1}^{C}\sum_{u=1}^{L} V_{stu}p_{stu}\right)^2\right],$$

$$\text{(c)}\ \sigma^2\left[\lambda_h^{(3)}\right] = \frac{1}{9}\left(\lambda_h^{(3)}\right)^{4}\left[\sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{L}(W_{ijk})^2 p_{ijk} - \left(\sum_{s=1}^{R}\sum_{t=1}^{C}\sum_{u=1}^{L} W_{stu}p_{stu}\right)^2\right],$$

where

$$\begin{aligned}
U_{ijk} &= \Delta_{ijk(1)} + \Delta_{ijk(2)} + \Delta_{ijk(3)},\\
V_{ijk} &= \Delta_{ijk(1)}\lambda_B^{(3)}\lambda_C^{(3)} + \Delta_{ijk(2)}\lambda_A^{(3)}\lambda_C^{(3)} + \Delta_{ijk(3)}\lambda_A^{(3)}\lambda_B^{(3)},\\
W_{ijk} &= \frac{\Delta_{ijk(1)}}{\left(\lambda_A^{(3)}\right)^2} + \frac{\Delta_{ijk(2)}}{\left(\lambda_B^{(3)}\right)^2} + \frac{\Delta_{ijk(3)}}{\left(\lambda_C^{(3)}\right)^2},
\end{aligned}$$

with

$$\Delta_{ijk(1)} = \frac{I(i=m_{jk})(1-p_{m_1\bullet\bullet}) - I(i=m_1)\left(1-\sum_{j=1}^{C}\sum_{k=1}^{L}p_{m_{jk}\bullet\bullet}\right)}{(1-p_{m_1\bullet\bullet})^2},$$

$$\Delta_{ijk(2)} = \frac{I(j=m_{ik})(1-p_{\bullet m_2\bullet}) - I(j=m_2)\left(1-\sum_{i=1}^{R}\sum_{k=1}^{L}p_{\bullet m_{ik}\bullet}\right)}{(1-p_{\bullet m_2\bullet})^2},$$

$$\Delta_{ijk(3)} = \frac{I(k=m_{ij})(1-p_{\bullet\bullet m_3}) - I(k=m_3)\left(1-\sum_{i=1}^{R}\sum_{j=1}^{C}p_{\bullet\bullet m_{ij}}\right)}{(1-p_{\bullet\bullet m_3})^2},$$

and $I(\cdot)$ is the indicator function.

Let $\hat{\sigma}^2[\lambda^*]$ denote $\sigma^2[\lambda^*]$ with $\{p_{ijk}\}$ replaced by $\{\hat{p}_{ijk}\}$. Then, $\hat{\sigma}[\lambda^*]/\sqrt{n}$ is an estimated standard error for $\hat{\lambda}^*$, and $\hat{\lambda}^* \pm z_{1-\alpha/2}\hat{\sigma}[\lambda^*]/\sqrt{n}$ is an approximate $100(1-\alpha)\%$ confidence interval for λ^*, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$th quantile of the standard normal distribution.

3. Extension to Multi-way Contingency Tables

3.1 Measures

Consider an $R_1 \times R_2 \times \cdots \times R_T$ contingency table with nominal categories in which the $(T-1)$ explanatory variables and one response variable are not defined. Let $p_{i_1 i_2 \cdots i_T}$ denote the probability that an observation will fall in the $(i_1, i_2, \cdots, i_T)$th cell of the table ($i_k = 1,\ldots,R_k;\ k = 1,\ldots,T$) and X_k ($k = 1,\cdots,T$) denote the kth variable. For $k = 1,\ldots,T$, a PRE measure in predicting the value of X_k is defined by

$$\lambda_k^{(T)} = \frac{\sum_{i_1=1}^{R_1}\cdots\sum_{i_{k-1}=1}^{R_{k-1}}\sum_{i_{k+1}=1}^{R_{k+1}}\cdots\sum_{i_T=1}^{R_T} p^{(k)}_{m_{i_1\cdots i_{k-1}i_{k+1}\cdots i_T}} - p^{(k)}_{m_k}}{1-p^{(k)}_{m_k}},$$

where

$$p^{(k)}_{m_{i_1\cdots i_{k-1}i_{k+1}\cdots i_T}} = \max_{i_k}(p_{i_1\cdots i_k\cdots i_T}),\quad p^{(k)}_{m_k} = \max_{i_k}\left(p^{(k)}_{i_k}\right),$$

and $p^{(k)}_{i_k} = \mathrm{P}(X_k = i_k)$. Then, we shall extend the measures as follows:

$$\lambda_a^{(T)} = \frac{1}{T}\sum_{k=1}^{T}\lambda_k^{(T)},$$

$$\lambda_g^{(T)} = \sqrt[T]{\prod_{k=1}^{T} \lambda_k^{(T)}},$$

and

$$\lambda_h^{(T)} = \frac{T}{\displaystyle\sum_{k=1}^{T} \frac{1}{\lambda_k^{(T)}}}.$$

The measures $\lambda_a^{(T)}$, $\lambda_g^{(T)}$ and $\lambda_h^{(T)}$ are the arithmetic mean, geometric mean and harmonic mean of the $\lambda_1^{(T)}$ through $\lambda_T^{(T)}$, respectively.

Let $\Lambda^{(T)}$ denote each of measures $\lambda_a^{(T)}$, $\lambda_g^{(T)}$ and $\lambda_h^{(T)}$. Each measure has the properties that (i) $\Lambda^{(T)}$ must lie between 0 and 1, (ii) $\Lambda^{(T)} = 0$ if and only if the information about $(T-1)$ variables does not reduce the probability of making an error in predicting the category of the other variable, and (iii) $\Lambda^{(T)} = 1$ if and only if no error is made, given knowledge of $(T-1)$ variables; namely there is complete predictive association. We point out that if all variables are independent, then the measure $\Lambda^{(T)}$ takes 0, but the converse need not hold. Note that when $\lambda_k^{(T)} = 0$ $(k = 1, \cdots T)$ such as all variables are independent, the measure $\lambda_h^{(T)}$ cannot measure the PRE. So in such a case, the measures $\lambda_a^{(T)}$ and $\lambda_g^{(T)}$ should be used as a PRE measure.

We see that

$$\min\left(\lambda_1^{(T)}, \cdots, \lambda_T^{(T)}\right) \le \lambda_h^{(T)} \le \lambda_g^{(T)} \le \lambda_a^{(T)} \le \max\left(\lambda_1^{(T)}, \cdots, \lambda_T^{(T)}\right),$$

where the equality holds if and only if $\lambda_1^{(T)}$ through $\lambda_T^{(T)}$ are all equal.

We note that $\Lambda^{(T)}$ when $T = 2$ is equivalent to the measure Λ proposed in Yamamoto et al. (2010).

3.2 Generalization of the Measures

Considering the monotonic function g, we shall propose a generalized measure, which includes the measures $\lambda_a^{(T)}$, $\lambda_g^{(T)}$ and $\lambda_h^{(T)}$, as follows:

$$\Lambda^{(T)} = g^{-1}\left(\frac{\sum_{k=1}^{T} g\left(\lambda_k^{(T)}\right)}{T}\right).$$

The functions g and g^{-1} are differentiable functions. Especially, (i) when $g(x) = x$, the measure $\Lambda^{(T)}$ is identical to $\lambda_a^{(T)}$, (ii) when $g(x) = \log x$, the measure $\Lambda^{(T)}$ is identical to $\lambda_g^{(T)}$, and (iii) when $g(x) = 1/x$, the measure $\Lambda^{(T)}$ is identical to $\lambda_h^{(T)}$.

$\Lambda^{(T)}$ has the same properties as $\lambda_a^{(T)}$, $\lambda_g^{(T)}$ and $\lambda_h^{(T)}$ (see Section 3.1).

3.3 Approximate Confidence Interval for Measures

Let $n_{i_1 i_2 \cdots i_T}$ denote the observed frequency in the $(i_1, i_2, \ldots, i_T)$th cell of the table $(i_k = 1, \ldots, R_k; k = 1, \ldots, T)$. Assume that a multinomial distribution applies to the $R_1 \times R_2 \times \cdots \times R_T$ table. In a similar way to the case of $T = 3$, $\sqrt{n}(\hat{\Lambda}^{(T)} - \Lambda^{(T)})$ (n is sample size and $\hat{\Lambda}^{(T)}$ is the estimated measure) has asymptotically a normal distribution with mean 0 and variance

$$\sigma^2\left[\Lambda^{(T)}\right] = \sum_{j_1=1}^{R_1} \cdots \sum_{j_T=1}^{R_T} \left(\frac{\partial \Lambda^{(T)}}{\partial p_{j_1 \cdots j_T}}\right)^2 p_{j_1 \cdots j_T} - \left(\sum_{s_1=1}^{R_1} \cdots \sum_{s_T=1}^{R_T} \left(\frac{\partial \Lambda^{(T)}}{\partial p_{s_1 \cdots s_T}}\right) p_{s_1 \cdots s_T}\right)^2.$$

For measures $\lambda_a^{(T)}$, $\lambda_g^{(T)}$ and $\lambda_h^{(T)}$, the variances are

$$\text{(a)}\ \ \sigma^2\left[\lambda_a^{(T)}\right] = \frac{1}{T^2}\left[\sum_{j_1=1}^{R_1} \cdots \sum_{j_T=1}^{R_T} (U_{j_1 \cdots j_T})^2 p_{j_1 \cdots j_T} - \left(\sum_{s_1=1}^{R_1} \cdots \sum_{s_T=1}^{R_T} U_{s_1 \cdots s_T} p_{s_1 \cdots s_T}\right)^2\right],$$

$$\text{(b)}\ \ \sigma^2\left[\lambda_g^{(T)}\right] = \frac{1}{T^2}\left(\lambda_g^{(T)}\right)^{2(1-T)}\left[\sum_{j_1=1}^{R_1} \cdots \sum_{j_T=1}^{R_T} (V_{j_1 \cdots j_T})^2 p_{j_1 \cdots j_T} - \left(\sum_{s_1=1}^{R_1} \cdots \sum_{s_T=1}^{R_T} V_{s_1 \cdots s_T} p_{s_1 \cdots s_T}\right)^2\right],$$

$$\text{(c)}\ \ \sigma^2\left[\lambda_h^{(T)}\right] = \frac{1}{T^2}\left(\lambda_h^{(T)}\right)^4\left[\sum_{j_1=1}^{R_1} \cdots \sum_{j_T=1}^{R_T} (W_{j_1 \cdots j_T})^2 p_{j_1 \cdots j_T} - \left(\sum_{s_1=1}^{R_1} \cdots \sum_{s_T=1}^{R_T} W_{s_1 \cdots s_T} p_{s_1 \cdots s_T}\right)^2\right],$$

where

$$U_{j_1\cdots j_T} = \sum_{k=1}^{T} \Delta_{j_1\cdots j_T(k)},$$

$$V_{j_1\cdots j_T} = \sum_{k=1}^{T} \Delta_{j_1\cdots j_T(k)} \lambda_1^{(T)} \cdots \lambda_{k-1}^{(T)} \lambda_{k+1}^{(T)} \cdots \lambda_T^{(T)},$$

$$W_{j_1\cdots j_T} = \sum_{k=1}^{T} \frac{\Delta_{j_1\cdots j_T(k)}}{\left(\lambda_k^{(T)}\right)^2},$$

with

$$\Delta_{j_1\cdots j_T(k)} = \frac{1}{\left(1 - p_{m_k}^{(k)}\right)^2} \left[I(j_k = m_{j_1\cdots j_{k-1} j_{k+1}\cdots j_T}) \left(1 - p_{m_k}^{(k)}\right) - I(j_k = m_k) \left(1 - \sum_{i_1=1}^{R_1} \cdots \sum_{i_{k-1}=1}^{R_{k-1}} \sum_{i_{k+1}=1}^{R_{k+1}} \cdots \sum_{i_T=1}^{R_T} p_{m_{i_1\cdots i_{k-1} i_{k+1}\cdots i_T}}^{(k)} \right) \right],$$

and $I(\cdot)$ is the indicator function.

Then, we can construct an asymptotic confidence interval using estimated variance although the detail is omitted.

4. An Example

Consider the data in Table 1, taken from Goodman (1975), which shows the McHugh test data on creative ability in machine design. This table cross-classifies 137 engineers with respect to their dichotomized scores (above the subtest mean (1) or below the subtest mean (2)) obtained on each of four different subtests that were supposed to measure creative ability in machine design. There are sixteen response patterns because the table has four variables (items A, B, C and D) and each has two categories.

Now, we are interested in what degree the relative decrease in the probability of making an error in predicting the value of one variable when we know the values of the other three variables as opposed to when we do not know them is. We shall analyze these data by using the proposed measure because the explanatory and response variables are not defined. When we use the measure $\lambda_a^{(4)}$, for example, the estimated value of $\lambda_a^{(4)}$ is 0.470 (Table 2). We see that in prediction of one of the variables from the others, the information reduces the probability of making an error by 47.0%. Similarly, the estimated values of $\lambda_g^{(4)}$ and $\lambda_h^{(4)}$ are 0.469 and 0.467, respectively. So we can also obtain a similar interpretation for the data.

We are also interested in the values of test statistic for the hypotheses of independence of (1) item A and items (B, C, D), (2) B and (A, C, D), (3) C and (A, B, D), and (4) D and (A, B, C). The values of Pearson's chi-squared statistic are 35.93 for (1), 37.67 for (2), 48.17 for (3), and 42.06 for (4) with seven degrees of freedom. Therefore, we can see the strong association between one of the variables and the other three variables. So, it would be meaningful to see the values of proposed measures.

5. Concluding Remarks

For analyzing multi-way (T-way) contingency tables with nominal categories, we have proposed three kinds of PRE measures which describes the relative decrease in the probability of making error in predicting value of one variable when the values of the other variables are known, as opposed to when they are not known. The proposed measures include arithmetic mean ($\lambda_a^{(T)}$), geometric mean ($\lambda_g^{(T)}$) and harmonic mean ($\lambda_h^{(T)}$). These measures are useful for analyzing the table which explanatory and response variables are not defined. A point to notice is that the measure $\lambda_h^{(T)}$ cannot measure the PRE when the variables are independent and/or any $\lambda_k^{(T)}$ ($k = 1, \ldots, T$) is 0. In such a case, the measures $\lambda_a^{(T)}$ and $\lambda_g^{(T)}$ should be used. It is difficult to discuss how to choose between three propositions: arithmetic, geometric or harmonic mean. We recommend the use of $\lambda_a^{(T)}$ for the simple interpretation.

In addition, the measure $\Lambda^{(T)}$, including $\lambda_a^{(T)}$, $\lambda_g^{(T)}$ and $\lambda_h^{(T)}$, is invariant under arbitrary permutations of the categories. Therefore the measure is suitable for analyzing the data on a nominal scale, but it is possible for analyzing the data on an ordinal scale because it only requires a categorical scale.

Acknowledgment

The authors would like to thank an editor and a referee for the helpful comments and suggestions.

References

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Massachusetts: The MIT Press.

Everitt, B. S. (1992). *The Analysis of Contingency Tables*. (2nd ed.). London: Chapman and Hall.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association, 49*, 732-764.

Goodman, L. A. (1975). A new model for scaling response patterns: An application of the quasi-independence concept. *Journal of the American Statistical Association, 70*, 755-768.

Theil, H. (1970). On the estimation of relationships involving qualitative variables. *American Journal of Sociology, 76*, 103-154. http://dx.doi.org/10.1086/224909

Tomizawa, S., & Ebi, M. (1998). Generalized proportional reduction in variation measure for multi-way contingency tables. *Journal of Statistical Research, 32*, 75-84.

Tomizawa, S., Seo, T., & Ebi, M. (1997). Generalized proportional reduction in variation measure for two-way contingency tables. *Behaviormetrika, 24*, 193-201. http://dx.doi.org/10.2333/bhmk.24.193

Tomizawa, S., & Yukawa, T. (2003). Proportional reduction in variation measures of departure from cumulative dichotomous independence for square contingency tables with same ordinal classifications. *Far East Journal of Theoretical Statistics, 11*, 133-165.

Yamamoto, K., Miyamoto, N., & Tomizawa, S. (2010). Harmonic, geometric and arithmetic means type uncertainty measures for two-way contingency tables with nominal categories. *Advances and Applications in Statistics, 17*, 143-159.

Yamamoto, K., Nozaki, Y., & Tomizawa, S. (2011). On measure of proportional reduction in error for nominal-ordinal contingency tables. *Journal of Statistics and Management Systems, 14*, 767-773.

Yamamoto, K., & Tomizawa, S. (2010). Measures of proportional reduction in error for two-way contingency tables with nominal categories. *Biostatistics, Bioinformatics and Biomathematics, 2*, 43-52.

Table 1. Frequency of occurrence of response patterns for the four machine design subtests (Goodman, 1975)

Response pattern				
item				Observed
A	B	C	D	frequencies
1	1	1	1	23
1	1	1	2	5
1	1	2	1	5
1	1	2	2	14
1	2	1	1	8
1	2	1	2	2
1	2	2	1	3
1	2	2	2	8
2	1	1	1	6
2	1	1	2	3
2	1	2	1	2
2	1	2	2	4
2	2	1	1	9
2	2	1	2	3
2	2	2	1	8
2	2	2	2	34
Total				137

Table 2. Estimates of the measures, approximate standard errors for them and approximate 95% confidence intervals for the measures, applied to Table 1

Measures	Estimated measure	Standard error	Confidence interval
$\lambda_a^{(4)}$	0.470	0.065	(0.343, 0.597)
$\lambda_g^{(4)}$	0.469	0.066	(0.340, 0.597)
$\lambda_h^{(4)}$	0.467	0.066	(0.337, 0.598)

Association between Spatial Patterns of Acute Malnutrition and Household Income in Iraq-2004

Faisal G. Khamis (Corresponding author)

AL-Zaytoonah University of Jordan, Faculty of Economics and Administrative Sciences, Jordan

E-mail: faisal_alshamari@yahoo.com

Ghaleb A. El-Refae

Al Ain University of Science and Technology, U.A.E

Received: February 6, 2012 Accepted: February 20, 2012 Published: May 1, 2012

doi:10.5539/ijsp.v1n1p43 URL: http://dx.doi.org/10.5539/ijsp.v1n1p43

Abstract

In this research acute malnutrition (AM) and household income (HI) are investigated. Historically, governorates of Iraq suffered inequality in AM and HI for several reasons, such as government's focus on heart of city in specific governorates; like Baghdad, Basra, and Nineveh. Question is raised whether the spatial patterns of AM and HI are existed in Iraq? If so, can the pattern of HI explain the pattern of AM?

The present paper investigates the spatial structure of AM across different governorates in Iraq and its spatial correlation to HI. This investigation will provide implications for policy makers, finding local clusters and showing visual picture for each of AM and HI.

The study utilizes a cross-sectional survey data collected in 2004 for 18 governorates. Mapping is used as the first step to conduct visual inspection for AM and HI using quartiles. Several spatial econometric techniques are available in the literature, which deal with the spatial autocorrelation in geographically referenced data. Two statistics of spatial autocorrelation, based on sharing boundary neighbours, known as global and local Moran's I, are carried out. Wartenberg's measure is used to detect bivariate spatial correlation.

In conclusion, based on visual inspection of mapping, global clustering in high level of AM and low level of HI were in general concentrated in western-southern governorates. This global clustering for AM was confirmed by significant global Moran's I statistic, but was not confirmed for HI. Out of 18 governorates, three and one governorates were found as local clusters in AM and HI respectively based on local Moran's I_i. Bivariate spatial correlation between AM and HI was not found significant.

Keywords: Spatial autocorrelation, Acute malnutrition, Household income, Mapping, Moran measure, Iraq

1. Introduction

Areas, independent of their geographic level of aggregation, are known to be interrelated partly due to their relative locations. Similar economic performance among areas can be attributed to proximity. Consequently, a proper understanding and accounting of spatial liaisons are needed in order to effectively forecast regional economic variables. Acute malnutrition (AM) and household income (HI) were studied in many developed and some developing countries using different statistical measures. We have developed a framework to better understand how AM and HI are spatially associated across governorates of Iraq. Martin et al. (2011) examined how poverty and education influence adolescent overweight. They found that family poverty and parental education do not predict adolescent overweight. Janjua, Iqbal, and Mahmood (2001) assessed the relationship between socioeconomic position (SEP) and under- and over-nutrition in Pakistan using multinomial regression based on cross-sectional data in 2006. They didn't find significant association between SEP and underweight.

The relationship between malnutrition and socioeconomic indicators was studied by several authors in many countries. Based on demographic and health survey 2004 to 2006, Urke, Bull, and Mittelmark (2011) investigated the association of parents socioeconomic status (SES) using parental education, occupation, and household wealth index, with malnutrition using the indicator child stunting in Peruvian Andes and in Peru nationally. In both samples, SES was significantly related to stunting. Subramanyam et al. (2011) assessed the association between changes in state per capita income and the risk of under-nutrition among children in India. They applied logistic models that accounted for the clustering of data and they didn't find consistent evidence that economic growth leads to reduction in childhood under-nutrition. Magalhães et al. (2011) estimated geographical risk profile of anemia accounting for malnutrition, malaria, and helminth infections in Burkina Faso, Ghana, and Mali in 2003-2006. They found malnutrition and parasites make to anemia. In Kenya, The high prevalence of chronic malnutrition suggested that stunting is a sustained problem within this urban informal settlement, not specifically resulting from the relatively brief political crisis (Burke et al., 2011). As stated by Filho, Kawachi, and Gotlieb (2012), income inequality is associated with worse population health in Sõ Paulo, Brazil. One of malnutrition reasons is the deprivation from basic requirements of life. There is greater variation in death rates and socio-demographic characteristics among the most deprived constituencies in Britain. Socio-demographic factors that are most strongly correlated with death rates among the most deprived places differ from areas of all deprivation levels and include population density, ethnicity and migration (Tunstall et al., 2011).

Twenty-five years ago, Iraq had the best living standard comparing to its neighbours. In the last years, Iraq was fallen due to some indicators. In some cases, was fallen far behind. The growing income inequality was contributed significantly to the deprivation increasing in Iraq. By the beginning of 1980s, other factors contributed to the rise in disparities among some governorates. These factors include geographic location of governorate in terms of its proximity to battlefields, and social composition of governorate's population especially in north and south. In other governorates, disparities stemmed according to political and tribal origins. The ratios of AM at main regions, South, Baghdad, Center, and North, in 2004 were 2.5, 1.8, 1.4, and 1.2 respectively. Median total of HI (in Iraqi Dinar) at these main regions in 2004 were 105 500, 100 000, 100 000, and 129 081 respectively. In recent years, a growing interest has been seen in examining the existence of spatial autocorrelation of AM and its spatial relationship to several indicators such as HI, education, etc. Low wage flexibility and limited labour mobility involve persistent unemployment differentials across governorates of Iraq. Governorates are tightly linked by migration, commuting, and inter-governorate trade. These types of spatial interaction are exposed to the frictional effects of distance, possibly causing the spatial dependence of governorate labour market conditions.

To understand linkages between socioeconomic health indicators, investigation should focus on features of areas rather than on compositional characteristics of residents of area, which cannot fully describe the social environment in which people live (Macintyre, Maciver, & Sooman, 1993). So, the aim of the research is to investigate geographical mapping and spatial autocorrelation of AM and its spatial relationship to HI. Spatial autocorrelation is the term used for the interdependence of the values of lattice data over space. However, it was argued that lattice data are spatially correlated. Mapping plays an important role in monitoring health status of people. Maps can reveal spatial patterns that is neither recognized previously nor suspected from the examination of statistics table. It reveals high risk communities or problem areas (Lawson & Williams, 2001). The purpose of spatial analysis is to identify pattern in geographical data and then attempts to explain this pattern. Findings are expected to enhance health monitoring and policy interventions across governorates of Iraq.

The importance of this research objective emanates the studies conducted by Amaral et al., (2011) in Brazilian community and Burke et al., (2011) in sub-Saharan Africa. They stated that malnutrition represents the strongest risk factor for morbidity and mortality. Also, to author's knowledge, no previous studies used spatial analysis techniques and geographical mapping in studying spatial inequality in AM and HI in Iraq. The importance of mapping was stated by Koch (2005): why make the map if detailed statistical tables carry the same results? Perhaps the most important reason for studying spatial statistics is not only interested in answering the "how much" question, but the "how much is where" question (Schabenberger & Gotway, 2005). Therefore, the usefulness of the paper is to suggest where to intervene geographically. In light of these: (1) the existence of spatial global clustering, (2) spatial local clusters for each of AM and HI were investigated, (3) mapping was displayed for each of AM and HI and for their local Moran's values, and (4) bivariate spatial correlation between AM and HI was examined based on Wartenberg's (1985) statistic. Study design was a cross-sectional analysis in a survey conducted in Iraq in 2004.

The paper is structured as follows: Section one reviews the literature relating to AM and HI disparities generally in several countries and particularly in Iraq. Materials and methods including data and analysis are presented in the second Section. Third section explains the results with many details. Discussion is explained in fourth Section. Last section is closed with several conclusions.

2. Materials and Methods

2.1 Data

Data were collected from the ministry of planning and development cooperation (2005), based on a survey conducted at Iraq in 2004. For each of (N=18) governorates, transformed AM and HI data were applied. AM is a devastating public health problem of epidemic proportions. The prevalence of AM is now one of the most widely used indicators of the severity of humanitarian crises throughout the world. This is endorsed by a wide array of UN organizations, donors, national governments and international agencies. AM for male and female in Iraq in 2004 were approximately same as accounted 1.9 and 1.8 respectively. AM is measured as low weight for height. The individual's weight is compared to the 'normal' weight for that height. Normal weights for children are determined by studies that have weighed thousands of healthy children. Based on this information, World Health Organization (WHO) developed charts known as international standards for expected growth. If the weight is less than international standards, the individual is considered acutely malnourished. WHO created cut-off points to indicate the severity of malnutrition? If the individual's weight-for-height is less than -3 z-scores (or standard deviations) of normal, s/he suffers from severe acute malnutrition. The formula for calculating this index is given by:

z = (measured value - median of reference population) / standard deviation of the reference population

HI (in Iraqi Dinars) is measured as the median total of household income. Overall, households receive 45.3% of their income from wages and salaries; 25.0% from self-employment and employer income; 19.8% from property income; 5.2% from social payments; and 4.7% from "transfers". These percentages vary geographically. For example, wages and salaries account for 31.4% of household income in Al-Najaf but 56.7% in Basrah; self-employment and employer income account for 8.8% in Diala but 43.1% in Al-Najaf; and property income accounts for 14.2% in Al-Muthanna but 27.3% in Erbil.

2.2 Analysis

Data analysis involved six steps. In step 1, AM and HI variables were tested for normal distribution. Both variables aren't found to follow normal distribution. Therefore, both variables are transformed to follow normal distribution using LISREL software. LISREL scales normal scores so that transformed variable has the same sample mean and standard deviation as the original variable. Thus, normal score is a monotonic transformation of original score with same mean and standard deviation (this characteristic can be considered as an advantage in this transformation) but with values of skewness and kurtosis much reduced. In step 2, visual inspection based on quantified gradients for each of AM and HI was conducted using quartiles. Step 3 included the calculation of global Moran's -statistic for each of AM and HI to detect global clustering. Also, the significance of -statistic was examined using permutation test. Step 4 involved the calculation of local Moran's for governorate and it's -value using Monte Carlo simulation to detect local clusters for each of AM and HI. In step 5, visual inspection for local Moran values was inspected using choropleth mapping. In Step 6, bivariate spatial correlation between AM and HI was examined using Wartenberg's (1985) statistic.

Variables were categorized by four intervals. These intervals were used for all maps using darker shades of gray to indicate increasing values of studied variables. Such approach enables qualitative evaluation of spatial pattern. In neighbourhood researches, neighbours may be defined as areas border each other or within a certain distance of each other. In this research neighbouring structure was defined as governorates share a boundary. The *second order* method (queen pattern) that include both the first-order neighbours (rook pattern) and those diagonally linked (bishop pattern) was used. A neighbourhood structure of Iraq's governorates is explained in Figure 1, where the ID neighbours for each governorate are shown.

Although maps allow visual assessment for spatial pattern, they have two important limitations. First, their interpretation varies from person to person. Second, there is possibility that a perceived pattern is actually the result of randomness and thus not meaningful. For these reasons, it makes sense to compute a numerical measure for spatial pattern, which can accomplish using spatial autocorrelation measures.

2.2.1 Identification of Global Spatial Clustering

Moran's I coefficient was used to measure the strength of spatial autocorrelation. In this exploratory spatial analysis, the spatial autocorrelation was tested using standard normal deviate (z-value) of Moran's I under normal assumption. The null hypothesis of no spatial autocorrelation or spatially independent versus the alternative of positive spatial autocorrelation is as follows:

H_0 : No clustering exists (no spatial autocorrelation)
H_1 : Clustering exists (positive spatial autocorrelation)

Moran's I is calculated as follows (Cliff & Ord, 1981):

$$I = \frac{N\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{S_0\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

and

$$S_0 = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij},\ i \neq j$$

Where, $n = 18$ is the number of governorates, $w_{ij} = 1$ is a weight denoting the strength of the connection if the two governorates i and j shares a common boundary, otherwise, $w_{ij} = zero$, and x_i and x_j represent AM or HI in i*th* and j*th* governorate respectively.

A significant positive value of Moran's I indicates positive spatial autocorrelation, showing the overall pattern for governorates having a high/low level of AM or HI similar to their neighbouring governorates. A significant negative value indicates negative spatial autocorrelation, showing the governorates having a high/low level of AM or HI unlike their neighbouring governorates. To test the significance of global Moran's I, z-statistic which follows a standard normal distribution was applied. It is calculated as follows (Weeks, 1992):

$$z = \frac{I - E(I)}{\sqrt{var(I)}}$$

Permutation test was applied. A permutation test tells us that a certain pattern in data is or is not likely to have arisen by chance. The observations of AM or HI were randomly reallocated 1 000 times with 1 000 of spatial autocorrelations were calculated in each time to test the null hypothesis of randomness. The hypothesis under investigation suggests that there will be a tendency for a certain type of spatial pattern to appear in data. Whereas the null hypothesis says that if this pattern is present, then this is a pure chance effect of observations in a random order. Null hypothesis is: no pattern in the data.

The analysis suggests an evidence of clustering if the result of the global test is found significant; though it doesn't identify the location of any particular cluster. Besides clustering that represents global characteristic of AM or HI, the existence and location of localized spatial clusters are of interest in geographic sociology. Accordingly, local spatial statistic was advocated for identifying and assessing potential clusters.

2.2.2 Identification of Local Spatial Clusters

A global index can suggest *clustering* but cannot identify individual *clusters* (Waller & Gotway, 2004). Anselin (1995) proposed local Moran's I_i statistic to test local autocorrelation. Local spatial clusters, sometimes referred to as hot spots, may be identified as those locations or sets of contiguous locations for which the local Moran's I_i is significant. Local Moran statistic was used to test the null hypothesis of no clusters. Moran's I_i for i*th* governorate may be defined by Waller and Gotway (2004) as:

$$I_i = \frac{(x_i - \bar{x})}{S} \sum_{j=1}^{N} \left(w_{ij} \Big/ \sum_{k=1}^{N_i} w_{ik} \right) \frac{(x_j - \bar{x})}{S}, \; i = 1, 2, ..., 18$$

Where, analogous to the global Moran's I, x_i and x_j represent the AM or HI in i*th* and j*th* governorate respectively, N = number of neighbours for i*th* governorate, and S is the standard deviation. It is noteworthy to mention that the number of neighbours for i*th* governorate was taken into account by the amount: $(w_{ij} / \sum_{k=1}^{N_i} w_{ik})$, where w_{ij} was measured in the same manner as in Moran's I statistic.

Cluster could be due to either aggregation of high values, aggregation of low values, or aggregation of moderate values. Thereby, high value of I_i suggests a cluster of similar (but not necessarily large) values across several governorates. Low value of I_i suggests an outlying cluster in a single governorate i (being different from most or all of its neighbours). A positive local Moran value indicates local stability, such as a governorate that has high/low value surrounded by governorate that has high/low value. A negative local Moran value indicates local instability, such as a governorate having low value surrounded by governorates having high value or vice versa. Each governorate's I_i value was mapped to provide insight into the location of governorates with comparatively high or low local association with their neighbouring values.

2.2.3 Bivariate Spatial Association

So far, only univariate spatial correlation is presented. It quantifies the spatial structure of one variable at a time. Rates of malnutrition on their own do not mean very much, unless the underlying related indicators of malnutrition are understood: i.e. whether a socioeconomic indicator such as HI is correlated with malnutrition? Spatial dependence or spatial clustering causes losing in the information that each observation carries. When N observations are made on a variable that is spatially dependent and that dependence is positive so that nearby values tend to be similar, the amount of information carried by the sample is less than the amount of information that would be carried if the N observations are independent. Due to a certain amount of information carried by each observation is duplicated by other observations in the cluster. A general consequence of this is that the sampling variance of statistics is underestimated. As the level of spatial dependence increases, the underestimation increases. The problem prevails when spatial autocorrelation is present. The variance of sampling distribution of e.g. Pearson correlation coefficient, which is a function of number of pairs of observations, is underestimated. Spatial autocorrelation coefficient can be modified to estimate the bivariate spatial correlation between two variables (Wartenberg, 1985):

$$I_{xy} = \frac{1}{S_0} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(y_j - \bar{y})}{[\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2/N}][\sqrt{\sum_{j=1}^{N}(y_j - \bar{y})^2/N}]}$$

Where x and y are AM and HI variables respectively. Although the mathematics is quite straightforward, very few software packages offer the option of computing, I_{xy}. Thus, programming was used to find I_{xy}. To test the significance of I_{xy}, z-statistic was applied: $z = I_{xy}\sqrt{N-1}$, which follows approximately standard normal distribution. In statistical analysis, all programs were performed using $S + 8$ Software.

3. Results

Figure 1 shows the study area explaining all governorates with their identification numbers (ID). Figure 2a and b show visual insight for AM and HI respectively. Visual inspection shows that high level in AM was concentrated in western-southern governorates, particularly in governorates 9, 10, 14, and 15. While low level of HI was concentrated in northern and western-southern governorates, particularly in governorates 12, 13, 15, and 16. Several governorates are not observed visually as hot spots, such as 1, 2, and 4 in AM, and 7 and 18 in HI. But, after considering the information of their neighbours, i.e. calculating their local Moran values, the pattern of their hot spots can be obviously seen. Figure 3a and b show visual insight for local Moran values of AM and HI respectively. Darkest shade corresponds to the highest quartile. These maps display geographical inequalities across governorates of Iraq. Based on visual inspection for AM taken from Figure 2a, an overall worsening pattern (high scores) was found in eastern-southern parts. Based on visual inspection for HI taken from Figure 2b, an overall worsening pattern (low scores) was found in northern and eastern-southern parts.

The suggestion of spatial clustering in HI that results from a visual inspection of mapping was not confirmed by global Moran's I of .01 with an associated -statistic of .53 and $p = .593$. The suggestion of spatial clustering in AM that results from a visual inspection of mapping was confirmed by a positive slightly significant global Moran's I of .16 with an associated -statistic of 1.68 and $p = .092$. To investigate global clustering, permutation test was done. For AM, permutation $p = .051$ was found slightly significant, and for HI, permutation $p = .294$ was found not significant. Thus, the null hypothesis of no spatial autocorrelation was rejected for AM while for HI was not rejected. Three and one significant local clusters were found for AM (their ID are 12, 14, and 15) and HI (its ID is 15) respectively. Clusters of AM and HI are located, as shown in Figures 3a and b respectively, in northern and southern parts. Transformed AM and HI variables, their local Moran values, and their significant p-values in boldface are shown in Table 1.

Simulated data are useful for validating the results of spatial analysis. When the word simulation is used, it is referred to an analytical method meant to imitate a real-life system, especially when other analyses are too mathematically complex or too difficult to reproduce. However, using Monte Carlo simulation, 9 999 random samples, eighteen values for each sample, were simulated. The process of simulation was conducted under standard normal distribution to calculate p-values for local Moran values of AM and HI. While results are specific to these data the case study helps to identify general concepts for future studies.

The AM is influenced by several variables in various ways. Spatial relationship between AM and HI was investigated. Pearson correlation coefficient between AM and HI, that is not spatial measure, was found -.18. Pearson correlation was not found significant with $p = .243$. Bivariate spatial correlation between AM and HI was found ($I_{xy} = -.24$), that is not significant with $z = -.99$ and $p = 0.990$.

4. Discussion

This study aims to clarify how the risk of AM is spatially clustered and how is associated with HI across governorates of Iraq. The above framework revealed some noteworthy findings. Such findings allow policy makers to better identify what types of resources are needed and precisely where they should be employed across governorates. After rejecting the null hypothesis for AM, it becomes possible to state that there was a form of global clustering and it was, of course, of interest to know the exact nature of this clustering. Are there hot-spot clusters? If so, how many hot-spots are there? Where are they located?

Exploratory tools such as descriptive Table and somewhat small area choropleth maps were used. Maps provide powerful means to communicate data to others. Unlike information displayed in graphs, tables and charts; maps also provide bookmarks for memories. In this way, maps are not passive mechanism for presenting information. Usually, in the spatial analysis and geographical mapping, small spatial areas should be used such as districts, counties. . . etc. But, in this research governorates were used which are considered somewhat larger than for example districts due to data were not available for smaller areas. Most often the word 'neighbourhood' suggests a relatively small area surrounding individuals' homes. But researchers commonly make use of larger spatial area such as census tracts (Coulton et al., 2001). Often, choices about neighbourhood spatial definitions were made with respect to convenience and availability of contextual data rather than study purpose (Schaefer-McDaniel et al., 2009). Schaefer-McDaniel et al. stated that, researchers might utilize census data and thus rely on census-imposed boundaries to define neighbourhoods even though these spatial areas may not be the best geographic units for the study topic.

Although, this work was conducted as part of a wider study, its immediate implications are more for policy makers and practitioners than for researchers. This study adds to the global body of knowledge on the utilization of spatial analysis to strengthen the research–policy interface in the developing countries such as Iraq. The spatial pattern of high AM in Iraq was concentrated in old governorates such as AL-Muthana, Salahuddin, and AlQadisiya. This pattern was characterized by underprivileged living conditions. This was consistent with what found by Yuan, Fulong, and Xueqiang (2011) in the city of Guangzhou, China. Our contribution shows that the patterns of AM and HI are quite complicated. Also, Investigators should draw from different research strands to understand where and why socioeconomic indicators particularly HI could matter for AM. The findings of this study suggest that tackling AM is a high priority. There should be fostered efforts to ensure that malnutrition-prevention strategies include the poor governorates.

As noted by Waller and Jacques (1995), the test for spatial pattern employs alternative hypotheses of two types; the omnibus not the null hypothesis or more specific alternatives. Tests with specific alternatives include focused tests that are sensitive to monotonically decreasing risk as distance from a putative exposure source (the focus) increases. Acceptance of either types (the omnibus or a more specific alternatives) only demonstrates that some spatial pattern exist, and does not implicate a cause (Jacques, 2004). Hence, the existence of spatial pattern alone cannot demonstrate nor prove a causal mechanism.

It is very well known that employment is considered a major source for household income. Iraq faced specific challenges with regard to jobs in its state-owned enterprises. With over 500 000 workers on the payroll, state-owned enterprises were a major source of employment. Analysis of other episodes of conflict in Iraq indicates a very strong reciprocal relationship between the lack of security and high unemployment (World Bank, 2006). While reconstruction and associated public-sector jobs are important in the initial phase of Iraq's recovery, they will not create a sufficient number of jobs to meet the population's needs in the long term; even if recovery is on a massive scale. Researchers recommend designing global development strategies that focus on job creation and income generation. These strategies incorporate elements of basic social protection and social dialogue at the global and local levels as an attempt to reduce inequalities in AM and HI.

The usual correlation coefficients, such as Pearson coefficient, only test whether there is an association between two attributes by comparing values at the same location. While measuring spatial correlation involves more than pair wise comparison between data recorded at the same locations as spatial units are arbitrary subdivisions of study region, people could move around from one area to another. People could be affected by the variation in HI and other socioeconomic indicators in areas other than the area they live in. i.e., AM inequality in ith governorate is thought to be influenced and explained by the inequality of HI and probably other socioeconomic variables not just in ith governorate but also in neighbouring governorates. However, the population size in these governorates is not equal; i.e. the rate of AM does not express the absolute size of the problem.

Anselin (1995) stated that indication of local pattern of spatial association may be in line with a global indication, although this is not necessarily be the case. It is quite possible that the local pattern is an aberration that the global indicator would not pick up, or it may be that a few local patterns run in the opposite direction of the global spatial trend. This case is found in this research. Local values that are very different from the mean (or median) would indicate locations that contribute more than their expected share to the global statistic. These may be outliers or high leverage points and thus would invite closer scouting. Although global clustering was not found significant for HI and was found slightly significant for AM, several local clusters were found significant in AM and HI.

The application of statistical techniques to spatial data faces an important challenge, as expressed in the first law of geography: "everything is related to everything else, but closer things are more related than distant things" (Tobler, 1979). The quantitative expression of this principal is the effect of *spatial dependence*, i.e. when the observed values are spatially clustered, the samples are not independent. Increasing in AM level in governorate generates increasing in AM level in governorate . This mechanism of transmission causes a spatial autocorrelation. The obvious question after finding significant clusters in AM is-why? Could this pattern associated by the spatial pattern of socioeconomic indicators such as HI or by the limitation of economic resources? However, further research is required regarding this bivariate spatial association between AM and other socioeconomic indicators. This research will be our interest in Iraq and other developing countries in the near future.

It should be emphasized that AM problem cannot overcome in the short-run, but long-term efforts are needed to tackle inequality across governorates. In turn, enabling the economy to create more job opportunities and establish new projects, especially in the governorates that found as hot spot clusters. It means that the place of the problem is now clearly shown. Health inequality is ubiquitous around the world. Thus, fresh perspective to tackle inequality is always welcomed by research community invested in reducing and eventually eliminating this inequality. Finally, this kind of studies should be conducted periodically in light of the changing of socioeconomic and political conditions.

5. Conclusions

Conclusions are comprehensive in at least five aspects. First, visual inspection shows that high level in AM was concentrated in western-southern governorates. While low level of HI was concentrated in northern and western-southern governorates. Second, several governorates are not observed visually as hot spots in am and HI. But, after considering the information of their neighbours, the pattern of their hot spots can be obviously seen. Third, global clustering in AM was found slightly significant but was not found significant in HI. Based on local Moran measure, three governorates were found to be significant local clusters in AM located in central and western-southern governorates. Only one significant local cluster in HI was found. The opposite being the case for those with low AM were in general seen in central and eastern governorates and those with high HI were seen in central and some southern governorates. Forth, from negative local Moran values, looking at the local variation, some governorates were represented as areas of dissimilarity in AM and in HI. Means that these governorates with low AM and/or HI surrounded by governorates with high AM and/or HI or vice versa. Fifth, spatial correlation between AM and HI was not found significant. This is consistent with what found by several studies. Martin et al. (2011) concluded that poverty status is not associated with overweight. Janjua, Iqbal, and Mahmood (2011) stated that no significant association found between SEP and underweight. Subramanyam et al. (2011) didn't find consistent evidence that economic growth leads to reduction in childhood under-nutrition.

Maps display geographical inequality in AM and HI across governorates of Iraq. The analytical approach used here delineates governorates of relatively high AM. This permits policy makers to develop strategies to minimize this inequality. Policy which pays attention to area characteristics will reduce the variation in AM and HI. Consequently this will improve prosperity which in turn improves population health. In summary, the study supports the hypothesis of a spatial clustering in AM at governorate level that probably reflects the inequality distribution in several socioeconomic indicators across governorates of Iraq. Researchers recommend that direct investments in appropriate health interventions may be necessary to reduce both AM level in each governorate and the variation across all governorates.

References

Amaral, T. F., Marucci, M. F. N., et al. (2011). Under-nutrition as a major risk factor for death among older Brazilian adults in the community-dwelling setting: SABE survey. *Nutrition, 27*(10), 1017-1022. http://dx.doi.org/10.1016/j.nut.2010.11.008

Anselin, L. (1995). Local Indicators of Spatial Association-LISA. *Geographical Analysis, 27*, 2, 93-115. http://dx.doi.org/10.1111/j.1538-4632.1995.tb00338.x

Cliff, A. D., & Ord, J. K. (1981). *Spatial Processes: Models & Applications*. London: Page Bros.

Coulton, C. J., Korbin, J., Chan, T., & Su, M. (2001). Mapping residents' perceptions of neighbourhood boundaries: A methodological note. *American Journal of Community Psychology, 29*(2), 371-383. http://dx.doi.org/10.1023/A:1010303419034

Filho, A. C., Kawachi I., & Gotlieb, S. D. (2012). Propensity score matching approach to test the association of income inequality and mortality in São Paulo, Brazil. *Journal of Epidemiology & Community Health, 66*, 14-17. http://dx.doi.org/10.1136/jech.2010.108852

Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge university press. http://dx.doi.org/10.1017/CBO9780511754944

Jacques, G. M. (2004). Current practices in the spatial analysis of cancer: flies in the ointment. *International Journal of Health Geographics, 3*(22), 1-10. [Online] Available: http://www.ij-healthgeographics.com/content/3/1/22

Janjua, N. Z., Iqbal, R., & Mahmood, B. (2011). Association of socioeconomic position with under- and over-nutrition in Pakistan. *Annals of Epidemiology, 21*(12), 884-891. http://dx.doi.org/10.1016/j.annepidem.2011.08.006

Koch, T. (2005). *Cartographies of Disease: Maps, Mapping, and Medicine, 1st*. Ed. California: Guilford Press.

Lawson, A. B., & Williams, F. R. (2001). *An Introductory Guide to Disease Mapping*. New York: Wiley & Sons. http://dx.doi.org/10.1002/0470842571

Macintyre, S., Maciver, S., & Sooman, A. (1993). Area, class, and health: Should we be focusing on places or people? *Journal of Social Policy, 22*, 213-234. http://dx.doi.org/10.1017/S0047279400019310

Magalhães, S., Ricardo J., & Archie, C. A. (2011). Mapping the Risk of Anemia in Preschool-Age Children: The Contribution of Malnutrition, Malaria, and Helminth Infections in West Africa. *PLoS Medicine, 8*(6), 1-16.

Martin, M. A., Frisco, M. L., et al. (2011). Social stratification and adolescent overweight in the UN: How income and educational resources matter across families and schools. *Social Science* & *Medicine*. (Article in press). http://dx.doi.org/10.1016/j.socscimed.2011.11.006

Ministry of Planning and Development Cooperation / The Republic of Iraq. (2006). Unsatisfied basic needs mapping and living standards in Iraq. United Nations Development Program, 1-25.

Olack, B., Burke, H., Cosmas, L., et al. (2011). Nutritional Status of Under-five Children Living in an Informal Urban Settlement in Nairobi, Kenya. *Journal of Health, Population*, & *Nutrition, 29*(4), 357-363. http://dx.doi.org/10.3329/jhpn.v29i4.8451

Schabenberger, O., & Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. Boca Raton: Chapman & Hall.

Schaefer-McDaniel, N., Caughy, M. O'B., O'Campo, P., & Gearey, W. (2009). Examining methodological details of neighbourhood observations and the relationship to health: A literature review. *Social Science* & *Medicine*, 1-16, (article in press).

Subramanyam, M., Kawachi, I., et al. (2011). Is economic growth associated with Reduction in child under-nutrition in India? *PLoS Medicine, 8*(3), 1-15. http://dx.doi.org/10.1371/journal.pmed.1000424

Tobler, W. R. (1979). *Cellular geography, In: Philosophy in Geography*. Holland: DReidel Publishing Company. http://dx.doi.org/10.1007/978-94-009-9394-5_18

Tunstall, H., Mitchell, R., Dorling, D., et al. (2011). Socio-demographic diversity and unexplained variation in death rates between the most deprived areas in Britain. *Journal of Epidemiology and Community Health, 65*. (Article in press). http://dx.doi.org/10.1136/jech.2011.143586.50

Urke, H. B., Bull, T., & Mittelmark, M. B. (2011). Socioeconomic status and chronic child malnutrition: wealth and maternal education matter more in the Peruvian Andes than nationally. *Nutrition Research, 31*(10), 741-747. http://dx.doi.org/10.1016/j.nutres.2011.09.007

Waller, L. A., & Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data*. New Jersey: Wiley & Sons. http://dx.doi.org/10.1002/0471662682

Waller, L. A., & Jacques, G. M. (1995). Disease models implicit in statistical tests of disease clustering. *Epidemiology, 6*, 584-590. http://dx.doi.org/10.1097/00001648-199511000-00004

Wartenberg, D. (1985). Multivariate Spatial Correlation: A Method for Exploratory Geographical Analysis. *Geographical Analysis, 17*(4), 263-282. http://dx.doi.org/10.1111/j.1538-4632.1985.tb00849.x

Webb, P., & Bhatia, R. (2005). A Manual: Measuring and Interpreting Malnutrition and Mortality.

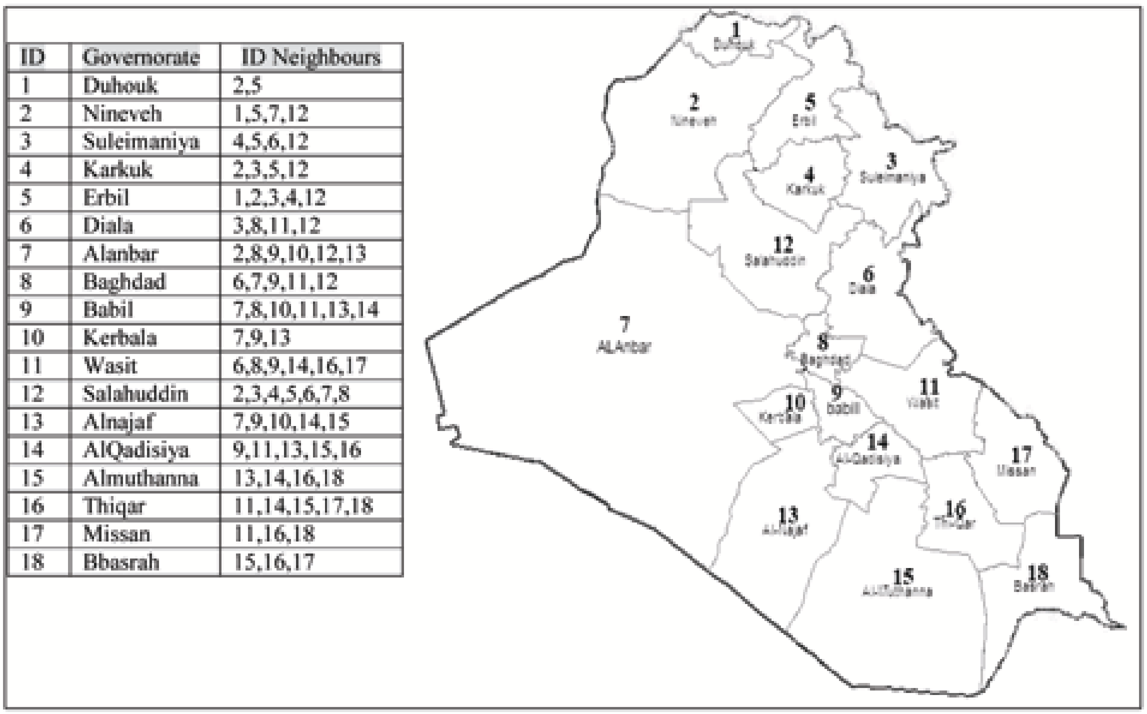
Weeks, J. R. (1992). *Population: An Introduction to concepts and Issues*, *5th*. Ed. Wadsworth, Inc.

World Bank. (2006). Rebuilding Iraq: Economic reform and transition. [Online] Available: http://siteresources.worldbank.org/IRFFI/Resources/IraqCEM-finalComplete.pdf

Yuan, Y., Fulong, W., & Xueqiang, X. (2011). Multiple Deprivations in Transitional Chinese Cities: A Case Study of Guangzhou. *Urban Affairs Review, 47*(5), 721-736. http://dx.doi.org/10.1177/1078087411400370

Table 1. Shows both transformed HI (Iraqi Dinars) and AM (%), local Moran's I_i values for transformed HI and AM, and their corresponding p-values (significant values in boldface at. 1 level)

ID	HI	I_i for HI	p-value	AM	I_i for AM	p-value
1	130688.32	.90	.056	.69	.44	.153
2	94859.43	-.47	.879	.69	.44	.082
3	118281.79	.03	.369	-.13	.37	.106
4	105361.23	-.03	.542	1.09	.61	.070
5	141569.82	-.03	.535	2.53	-.59	.929
6	100400.44	.07	.308	2.29	-.24	.770
7	121537.60	-.39	.884	1.92	.01	.421
8	102918.52	.02	.386	1.54	.00	.435
9	112740.78	-.04	.560	2.90	.29	.085
10	110222.71	.00	.465	3.14	.08	.306
11	107779.99	-.01	.470	1.54	-.26	.816
12	82452.91	-.43	.915	1.28	.36	**.044**
13	87700.21	.23	.144	1.92	-.07	.610
14	97743.74	.35	.083	4.12	.79	**.011**
15	71571.41	.95	**.011**	3.46	.73	**.025**
16	91603.63	.14	.212	2.29	.20	.168
17	125441.02	-.10	.627	2.71	-.05	.568
18	115397.48	-.29	.775	1.92	-.05	.571



ID	Governorate	ID Neighbours
1	Duhouk	2,5
2	Nineveh	1,5,7,12
3	Suleimaniya	4,5,6,12
4	Karkuk	2,3,5,12
5	Erbil	1,2,3,4,12
6	Diala	3,8,11,12
7	Alanbar	2,8,9,10,12,13
8	Baghdad	6,7,9,11,12
9	Babil	7,8,10,11,13,14
10	Kerbala	7,9,13
11	Wasit	6,8,9,14,16,17
12	Salahuddin	2,3,4,5,6,7,8
13	Alnajaf	7,9,10,14,15
14	AlQadisiya	9,11,13,15,16
15	Almuthanna	13,14,16,18
16	Thiqar	11,14,15,17,18
17	Missan	11,16,18
18	Bbasrah	15,16,17

Figure 1. Study area shows all governorates with their ID and the neighbours of each governorate

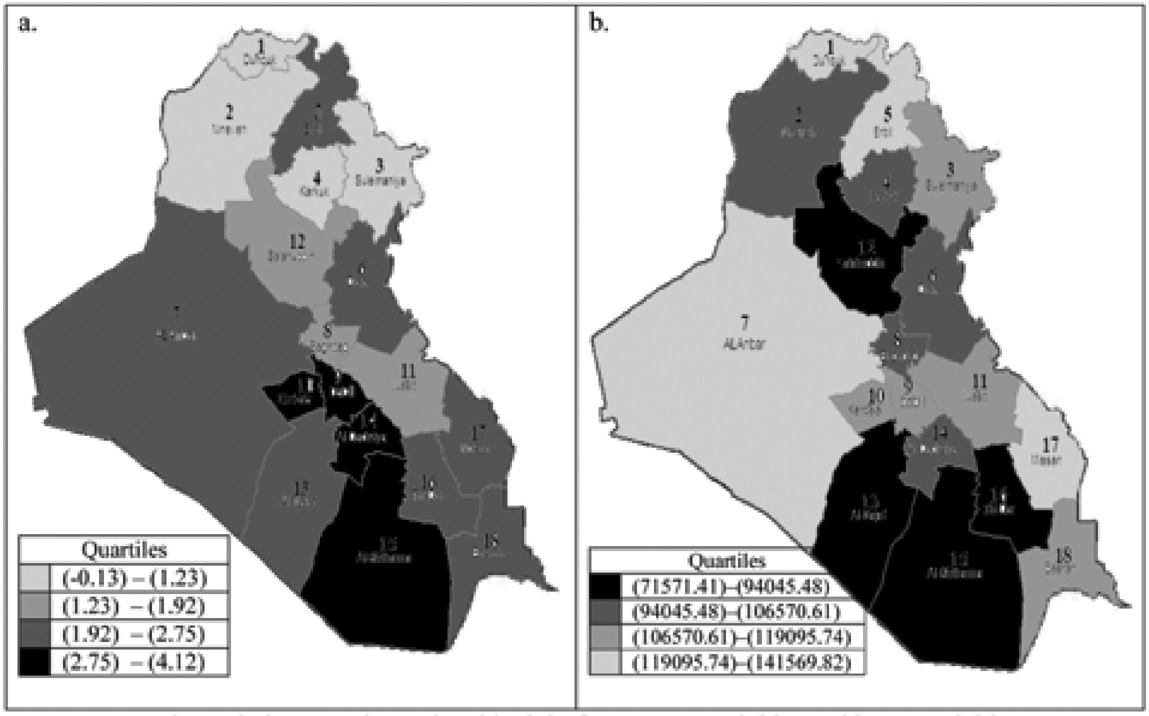


Figure 2. Choropleth maps show visual insight for: a. AM variable, and b. HI variable

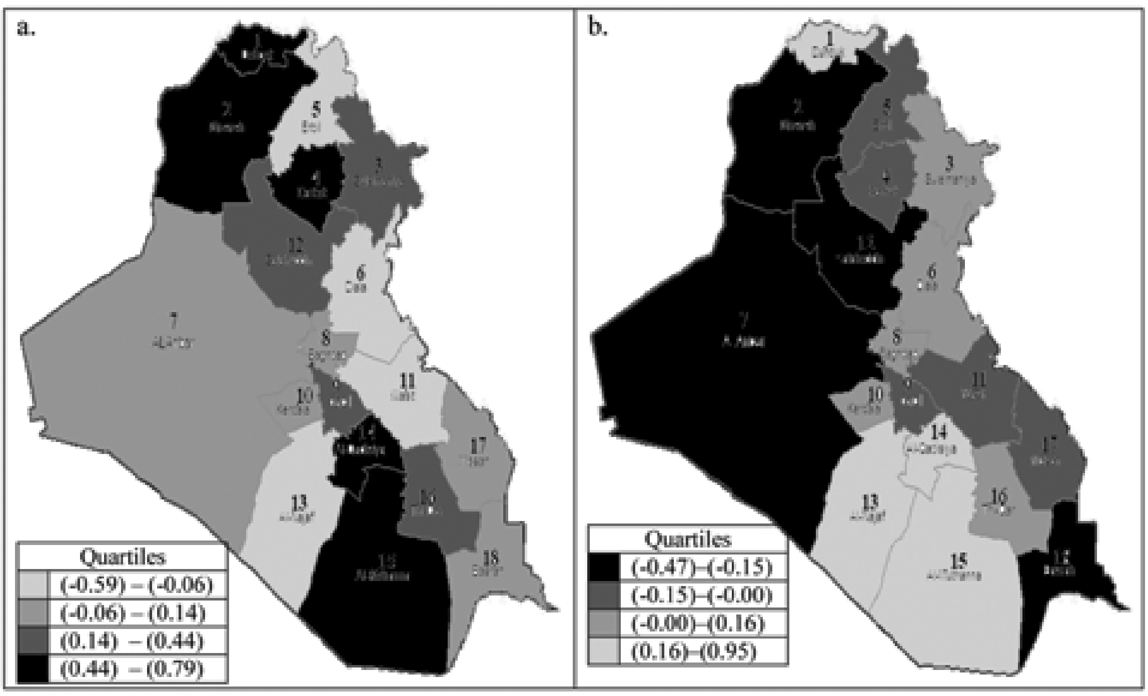


Figure 3. Choropleth maps show visual insight for local Moran values of: a. AM and b. HI variables

The Mc-Γ Distribution and Its Statistical Properties: An Application to Reliability Data

Francisco W. P. Marciano (Corresponding author)
Departamento de Estatística, Universidade Federal de Pernambuco
Cidade Universitária, 50740-540, Recife, PE, Brazil
Tel: 55-81-2126-7434 E-mail: bill.marciano@gmail.com

A. D. C. Nascimento
Departamento de Estatística, Universidade Federal de Pernambuco
Cidade Universitária, 50740-540, Recife, PE, Brazil
Tel: 55-81-2126-7434 E-mail: abraao.susej@gmail.com

M. Santos-Neto (Note 1)
Departamento de Matemática e Estatística, Universidade Federal de Campina Grande
Av. Aprígio Veloso, 882, Bloco CX, 58.109-970, Bodocongó, Campina Grande, PB, Brazil
Tel: 55-83-2101-1637 E-mail: santosneto@dme.ufcg.edu.br

G. M. Cordeiro
Departamento de Estatística, Universidade Federal de Pernambuco
Cidade Universitária, 50740-540, Recife, PE, Brazil
Tel: 55-81-2126-7434 E-mail: gauss@de.ufpe.br

Received: February 15, 2012 Accepted: March 5, 2012 Published: May 1, 2012
doi:10.5539/ijsp.v1n1p53 URL: http://dx.doi.org/10.5539/ijsp.v1n1p53

Abstract

The gamma distribution has been widely used in many research areas such as engineering, hydrology and survival analysis. We propose a new distribution, called the McDonald gamma distribution, which presents greater flexibility to model scenarios involving non-negative data. The new density function is represented as a double linear combination of gamma densities. We also propose analytical expressions for some mathematical quantities: moments, moment generating function, log-moment, mean deviations, Lorentz and Bonferroni curves, order statistics, entropy and quantile function. The score function and the observed information matrix of this new distribution are derived. A real data set is used to illustrate the importance of the proposed model.

Keywords: Beta-Generated class, Entropy, Generalized distribution, Maximum likelihood estimation, Moment

1. Introduction

The gamma distribution is a very general distribution that belongs to the Pearson type III family of distributions. It includes, among other well-known distributions, the exponential and chi-square distributions. Some of its structural properties can be found in Jambunathan (1954). It has a variety of applications and can be used to model the queuing systems, the flow of items through manufacturing and distribution processes, the risk management and some distributions in hydrology. More detailed information on hydrology can be found in (Yevjevich, 1972; Bobee & Ashkar, 1991).

Several generalized distributions have been studied in recent years. The generalization of continuous distributions began with Amoroso (1925). He introduced the generalized gamma (GG) distribution for income rate data. It is also discussed by Stacy (1962). Esteban (1981) showed that the GG distribution has several distributions as special cases or limiting forms, for example, the log-normal, Weibull, gamma, exponential, normal and Pareto distributions. The GG distribution has several applications in areas such as engineering, hydrology and survival analysis, and it is very useful in discriminating between alternative probabilistic models. Nadarajah & Gupta (2007) applied it to drought data. Nadarajah (2008) presented a study on its use in electrical and electronic engineering. Cox (2008) studied the F-generalized family by comparing it with the GG model. Cordeiro et al. (2011) proposed the exponentiated GG distribution with application to lifetime data and Pascoa et al. (2011) defined the called Kumaraswamy GG distribution with application in survival analysis, since it is capable of modeling a bathtub-shaped hazard rate function.

In recent years, several authors published new distributions. Eugene et al. (2002) introduced a class of generalized distributions based on the logit of the beta random variable. They proposed the beta normal distribution. After their work, new distributions have been developed in this class, such as the beta Fréchet (Nadarajah & Gupta, 2004), beta Gumbel (Nadarajah & Kotz, 2004), beta exponential (Nadarajah & kotz, 2005), beta Weibull (Lee et al., 2007), beta Pareto (Akinsete et al., 2008), beta half-normal (Pescim et al., 2006), beta generalized exponential (Barreto-Souza et al., 2010) and beta power (Cordeiro & Brito, 2012) distributions.

In this sense, our purpose is to present a new distribution, called the McDonald gamma distribution (Mc-Γ for short), which extends the gamma model and has several other models as special cases. Some of its mathematical properties are obtained and the method of maximum likelihood estimation is discussed. The new model provides greater flexibility than other distributions, since it has more shape parameters, yielding a large variety of forms. It can also be useful for testing the goodness of fit of its sub-models.

The rest of the article is organized as follows. In Section 2, we discuss the modeling of the Mc-Γ distribution, in which the class Mc is contextualized. We also present its density function, cumulative distribution and hazard function. Some expansions of its mathematical quantities are derived in Section 3. Additionally, its moment generating function (mgf) and the limiting density and cumulative distribution functions are derived. Moreover, we propose analytical expressions for the following statistical measures: Shannon and Rényi entropies, mean deviations, Bonferroni and Lorentz curves, skewness, kurtosis, order statistics and quantile function. In Section 4, we discuss maximum likelihood estimation. In Section 5, the Mc-Γ distribution is applied to a real data set. Finally, in Section 6, we provide some concluding remarks.

2. The Mc-Γ Distribution

The Mc-Γ distribution originated from the work of Eugene et al. (2002) who defined a general class of distributions as follows: if G denotes the cumulative distribution function (cdf) of a random variable, then a generalized class of distributions can be defined as

$$F(x) = I_{G(x)}(a,b) = \frac{1}{B(a,b)} \int_0^{G(x)} \omega^{a-1}(1-\omega)^{b-1}\mathrm{d}\omega, \tag{1}$$

for $a > 0$, $b > 0$, where $I_y(a,b) = B_y(a,b)/B(a,b)$ denotes the incomplete beta function ratio and $B_y(a,b) = \int_0^y \omega^{a-1}(1-\omega)^{b-1}\mathrm{d}\omega$ is the incomplete beta function. The probability density function (pdf) corresponding to (1) can be expressed as

$$f(x) = \frac{1}{B(a,b)}\, G(x)^{a-1}\,[1-G(x)]^{b-1}\, g(x),$$

where $g(x) = \partial G(x)/\partial x$ is the baseline density function.

We start with the generalized beta distribution of the first kind (or beta type I) introduced by McDonald (1984). Its pdf is given by

$$f(x) = \frac{c}{B\left(\frac{a}{c}, b\right)}\, x^{a/c-1}\,(1-x^c)^{b-1},$$

where $a > 0$, $b > 0$ and $c > 0$ are shape parameters.

Alexander et al. (2011) introduce the generalized beta-generated (GBG) distribution which has as sub-models the classical beta-generated, Kumaraswamy-generated and exponentiated distributions. Consider starting from an arbitrary baseline continuous distribution function $G(x)$, the cdf $F(x)$ of the GBG distribution can be expressed as

$$F(x) = I_{G(x)^c}\left(\frac{a}{c}, b\right) = \frac{1}{B\left(\frac{a}{c}, b\right)} \int_0^{G(x)^c} \omega^{a/c-1}(1-\omega)^{b-1}\mathrm{d}\omega. \tag{2}$$

We follow the works of Eugene et al. (2002), Jones (2004), Cordeiro & de Castro (2011) and Alexander et al. (2011) to define the Mc-Γ distribution. If we substitute the gamma cdf in (2), we obtain the Mc-Γ cumulative distribution.

Here and henceforth, consider the pdf of the $\Gamma(\alpha,\beta)$ distribution given by

$$h(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, \exp(-\beta x), \quad x > 0,$$

where $\alpha > 0$ and $\beta > 0$ are shape and scale parameters, respectively.

Thus, the Mc-Γ cumulative distribution is

$$F_{\text{Mc-}\Gamma}(x;\alpha,\beta,a,b,c) = \frac{1}{B\left(\frac{a}{c},b\right)} \int_0^{\gamma_1(\alpha,\beta x)^c} w^{\frac{a}{c}-1}(1-w)^{b-1}\mathrm{d}w, \tag{3}$$

where $\gamma_1(\alpha,\beta x) = \frac{1}{\Gamma(\alpha)}\int_0^{\beta x} t^{\alpha-1}\mathrm{e}^{-t}\mathrm{d}t$ is the incomplete gamma function ratio, $\beta > 0$ is a scale parameter and $\alpha, a, b, c > 0$ are shape parameters.

The Mc-Γ density function is obtained by differentiating (3). We have

$$f_{\text{Mc-}\Gamma}(x;\alpha,\beta,a,b,c) = \frac{c\,\beta^\alpha\, x^{\alpha-1}\,\mathrm{e}^{-\beta x}}{\Gamma(\alpha)\,B(\frac{a}{c},b)}\,\gamma_1(\alpha,\beta x)^{a-1}\,[1-\gamma_1(\alpha,\beta x)^c]^{b-1}. \tag{4}$$

If X is a random variable with density (4), we write $X \sim$ Mc-$\Gamma(\alpha,\beta,a,b,c)$. Two important special sub-models are the beta gamma ($\mathcal{B}$-Γ) distribution (when $c = 1$), proposed by Jones (2004), and the Kumaraswamy gamma (Kw-Γ) distribution (when $a = 1$), studied by Cordeiro & de Castro (2011). The Mc-Γ density function for selected parameter values is plotted in Figs. 1(a)-1(b).

The survival and hazard rate function of the Mc-Γ distribution are

$$S_{\text{Mc-}\Gamma}(x;\alpha,\beta,a,b,c) = 1 - \frac{1}{B\left(\frac{a}{c},b\right)} \int_0^{\gamma_1(\alpha,\beta x)^c} w^{\frac{a}{c}-1}(1-w)^{b-1}\mathrm{d}w$$

and

$$h_{\text{Mc-}\Gamma}(x;\alpha,\beta,a,b,c) = \frac{c\,\beta^\alpha\, x^{\alpha-1}\mathrm{e}^{-\beta x}\,\gamma_1(\alpha,\beta x)^{a-1}\,[1-\gamma_1(\alpha,\beta x)^c]^{b-1}}{\Gamma(\alpha)\,B(\frac{a}{c},b)\,S_{\text{Mc-}\Gamma}(x;\alpha,\beta,a,b,c)},$$

respectively.

In Figs. 2(a)-2(b), we plot the hazard function for selected parameter values. This function is quite flexible and may take different forms: constant, increasing, decreasing and bathtub.

To generate random numbers from the Mc-$\Gamma(\alpha,\beta,a,b,c)$ distribution, we have to solve the nonlinear equation

$$\gamma_1(\alpha,\beta X) - U^{1/c} = 0, \tag{5}$$

where $U \sim \mathcal{B}\left(\frac{a}{c},b\right)$. We use the R programming language (R Development Core Team, 2008) for solving this equation. Figs. 3(a)-3(b). present the theoretical and approximate densities for different sample sizes, $n \in \{500, 2000, 10000\}$. In Fig. 3(b), we provide the histogram of the simulated data for $n = 500$. Clearly, the data are well-fitted by the Mc-Γ theoretical density.

2.1 Special Sub-models

The Mc-Γ distribution contains as special cases several well-known distributions shown in Fig. 4. In addition to these distributions, it contains other special cases such as: Mc-χ^2 ($\alpha = k/2, \beta = 1/2$), Mc-Exp ($\alpha = 1$), Kw-$\chi^2$ ($\alpha = k/2, \beta = 1/2, a = 1$), Kw-Exp ($\alpha = 1, a = 1$), $\mathcal{B}$-χ^2 ($\alpha = k/2, \beta = 1/2, c = 1$), $\mathcal{B}$-Exp ($\alpha = 1, c = 1$), $\mathcal{L}_1$-χ^2 ($\alpha = k/2, \beta = 1/2, b = 1$), $\mathcal{L}_1$-Exp ($\alpha = 1, b = 1$), $\mathcal{L}_2$-χ^2 ($\alpha = k/2, \beta = 1/2, a = 1, c = 1$), $\mathcal{L}_2$-Exp ($\alpha = 1, a = 1, c = 1$). Here, Kw-G denotes the family of Kumaraswamy G distributions, $\mathcal{B}$-G the family of beta G distributions, $\mathcal{L}_1$-G the family of Lehmann type I G distributions and $\mathcal{L}_2$-G the family of Lehmann type II G distributions.

3. Theoretical Properties

3.1 Expansions for Important Mathematical Quantities

Theorem 1. *Here and henceforth, let $X \sim$ Mc-$\Gamma(\alpha,\beta,a,b,c)$. If s is a non-negative real number, we obtain the double linear combination*

$$f^s_{Mc\text{-}\Gamma}(x;\alpha,\beta,a,b,c) = \sum_{v,m=0}^{\infty} w^{(s)}_{v,m}\; h(x;\alpha v + m + 1 + s(\alpha-1),\beta),$$

where $h(x;\alpha v + m + 1 + s(\alpha-1),\beta)$ is the gamma density with shape parameter $\alpha v + m + 1 + s(\alpha-1)$ and scale parameter β,

$$w^{(s)}_{v,m} = \frac{t_{m,v}\,\Gamma(\alpha v + m + 1 + s(\alpha-1))}{\beta^{m+1-s}\Gamma(\alpha)^v\sum_{i,k=0}^{\infty}\binom{ic+s(a-1)}{k}\binom{s(b-1)}{i}\binom{k}{v}(-1)^{i+k+v}},$$

and the quantity $t_{m,v}$ is defined in Appendix A.

The proof of this theorem is given in Appendix A.

Corollary 1. *We obtain the double linear combination for the density function of X*

$$f_{Mc\text{-}\Gamma}(x;\alpha,\beta,a,b,c) = \sum_{v,m=0}^{\infty} w_{v,m}^{(1)}\ h(x;\alpha v + m + \alpha,\beta),$$

where $w_{v,m}^{(1)}$ is given in Theorem 1.

Corollary 2. *If $Y \sim \Gamma(\alpha v + m + \alpha,\beta)$, explicit expressions for the cdf, nth moment and mgf of X are given by*

$$F_{Mc\text{-}\Gamma}(x;\alpha,\beta,a,b,c) = \sum_{v,m=0}^{\infty} w_{v,m}^{(1)}\ \gamma_1(\alpha v + m + \alpha,\beta),$$

$$\mathrm{E}(X^n) = \sum_{v,m=0}^{\infty} w_{v,m}^{(1)}\ \mathrm{E}(Y^n) = \sum_{v,m=0}^{\infty} w_{v,m}^{(1)}\ \frac{\Gamma(\alpha v + m + \alpha + n)}{\Gamma(\alpha v + m + \alpha)\beta^n},$$

$$\mathrm{M}_X(t) = \sum_{v,m=0}^{\infty} w_{v,m}^{(1)}\ \mathrm{M}_Y(t) = \sum_{v,m=0}^{\infty} w_{v,m}^{(1)}\ (1-\beta t)^{-(\alpha v + m + \alpha)},$$

for $t < \beta$, respectively, where $w_{v,m}^{(1)}$ is given in Theorem 1.

3.2 Asymptotic Density and Cumulate Distribution Functions

Consider the representation in power series in Appendix B and the asymptotic results for the exponentiated gamma density function given by Nadarajah and Kotz (2006). The following approximations for the asymptotic density and cumulate distribution functions hold:

$$f_{\mathrm{Mc}\text{-}\Gamma}(x;\alpha,\beta,a,b,c) \sim \begin{cases} \sum_{k=1}^{\infty} w'_k \ \left\{\frac{(ck+1)\alpha^{-ck}x^{\alpha ck}}{\Gamma(\alpha)^{ck+1}}\right\}, & x \to 0, \\ \sum_{k=1}^{\infty} w'_k \ \left\{\frac{(ck+1)x^{\alpha-1}\exp(-x)}{\Gamma(ck+1)}\right\}, & x \to \infty \end{cases}$$

and

$$F_{\mathrm{Mc}\text{-}\Gamma}(x;\alpha,\beta,a,b,c) \sim \begin{cases} \sum_{k=1}^{\infty} w'_k \ \left\{\frac{(ck+1)\alpha^{-ck}x^{\alpha ck+1}}{(\alpha ck+1)\Gamma(\alpha)^{ck+1}}\right\}, & x \to 0, \\ \sum_{k=1}^{\infty} w'_k \ \left\{\frac{(ck+1)\Gamma(\alpha)}{\Gamma(ck+1)}\right\}, & x \to \infty \end{cases}$$

where

$$w'_k = \frac{c\beta(-1)^k\binom{b-1}{k}}{(ck+1)B(\frac{a}{c},b)}.$$

3.3 Alternative Forms for Ordinary Moments

The following discussion proposes an alternative expression for the *nth* ordinary moment of X. We have

$$\mathrm{E}(X^n) = \frac{c\beta}{\Gamma(\alpha)B(\frac{a}{c},b)}\int_0^{\infty} x^{n+\alpha-1}\mathrm{e}^{-\beta x}[\gamma_1(\alpha,\beta x)]^{a-1}[1-\gamma_1(\alpha,\beta x)^c]^{b-1}\mathrm{d}x. \tag{6}$$

Since $\gamma_1(\alpha,\beta x) < 1$ and $c > 0$, the following expansion holds

$$[1-\gamma_1(\alpha,\beta x)^c]^{b-1} = \sum_{i=0}^{\infty}(-1)^i\binom{b-1}{i}\gamma_1(\alpha,\beta x)^{ci}.$$

Additionally, we work with an expansion for the incomplete gamma function

$$\gamma_1(\alpha,\beta x) = \frac{(\beta x)^{\alpha}}{\Gamma(\alpha)}\sum_{m=0}^{\infty}\frac{(-\beta x)^m}{(\alpha+m)m!}. \tag{7}$$

Thus, equation (6) can be reduced to

$$\begin{aligned}\mathrm{E}(X^n) &= \frac{c\beta^{\alpha}}{\Gamma(\alpha)B(\frac{a}{c},b)}\sum_{i=0}^{\infty}(-1)^i\binom{b-1}{i}\int_0^{\infty}x^{n+\alpha-1}\mathrm{e}^{-\beta x}\,\gamma_1(\alpha,\beta x)^{ic+a-1}\mathrm{d}x\\ &= \frac{c\beta^{\alpha}}{\Gamma(\alpha)B(\frac{a}{c},b)}\sum_{i=0}^{\infty}\frac{(-1)^i}{\Gamma(\alpha)^{ic+a-1}}\binom{b-1}{i}\int_0^{\infty}x^{n+\alpha-1}\mathrm{e}^{-\beta x}\left\{(\beta x)^{\alpha}\sum_{m=0}^{\infty}\frac{(-\beta x)^m}{(\alpha+m)m!}\right\}^{ic+a-1}\mathrm{d}x.\end{aligned}$$

Setting $u=\beta x$, we have

$$\begin{aligned}\mathrm{E}(X^n) &= \frac{c\beta^{-n}}{\Gamma(\alpha)B(\frac{a}{c},b)}\sum_{i=0}^{\infty}\frac{(-1)^i}{\Gamma(\alpha)^{ic+a-1}}\binom{b-1}{i}\underbrace{\int_0^{\infty}u^{n+\alpha-1}\exp(-u)\left\{u^{\alpha}\sum_{m=0}^{\infty}\frac{(-u)^m}{(\alpha+m)m!}\right\}^{ic+a-1}\mathrm{d}u}_{\text{Lauricella function of type A (Cordeiro \& Nadarajah, 2011)}}.\\ &= \frac{c\beta^{-n}}{\Gamma(\alpha)B(\frac{a}{c},b)}\sum_{i=0}^{\infty}\frac{(-1)^i\alpha^{-(ic+a-1)}}{\Gamma(\alpha)^{ic+a-1}}\binom{b-1}{i}\Gamma(n+\alpha(ic+a))F_A^{(ic+a-1)}(n+\alpha(ic+a);\alpha,\ldots,\alpha;\alpha+1,\ldots,\alpha+1;-1,\ldots,-1).\end{aligned}$$

3.4 Expression for the Rényi Entropy

Let Y be a random variable with density $f(y;\theta)$ and support $y\in\mathcal{D}\subset\mathbb{R}$. The Rényi entropy is defined by

$$H_{\mathrm{R}}^{s}(Y)=\frac{1}{1-s}\log\left\{\mathrm{E}[f(y;\theta)^{s-1}]\right\}=\frac{1}{1-s}\log\left(\int_{\mathcal{D}}f(y;\theta)^{s}\mathrm{d}y\right),$$

where $s\geq 0$ and $s\neq 1$.

From the expansion given in Theorem 1, we can write

$$\begin{aligned}H_{\mathrm{R}}^{s}(X) &= \frac{1}{1-s}\log\left(\sum_{v,m=0}^{\infty}w_{v,m}^{(s)}\int_0^{\infty}h(x;\alpha v+m+1+s(\alpha-1),\beta)\mathrm{d}x\right)\\ &= \frac{1}{1-s}\log\left(\sum_{v,m=0}^{\infty}w_{v,m}^{(s)}\right),\end{aligned}$$

where $w_{v,m}^{(s)}$ is given in Theorem 1.

3.5 Expressions for the Log-moment and Shannon Entropy

Let Y be defined as in Section 3.4. The log-moment is given by

$$\mathrm{E}\{[\log(Y)]^n\}=\int_{\mathcal{D}}\log(y)^n f(y;\theta)\mathrm{d}y.$$

Now, we consider $Y\sim\Gamma(\alpha,\beta)$ with density function $h(x;\alpha,\beta)$. Minor manipulations yield

$$\mathrm{E}\{[\log(Y)]^n\}=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\left\{\frac{\partial^{(n)}}{\partial\alpha^n}\left(\frac{\Gamma(\alpha)}{\beta^{\alpha}}\right)\right\}\xrightarrow{n=1}\mathrm{E}[\log(Y)]=\psi(\alpha)-\log(\beta),$$

where $\psi(\cdot)$ is the digamma function.

Combining this result with Corollary 1, if $Z\sim\Gamma(\alpha v+m+\alpha,\beta)$, we have

$$\begin{aligned}\mathrm{E}\{[\log(X)]^n\} &= \sum_{v,m=0}^{\infty}w_{v,m}^{(1)}\ \mathrm{E}\{[\log(Z)]^n\}\\ &= \sum_{v,m=0}^{\infty}w_{v,m}^{(1)}\frac{\beta^{\alpha v+m+\alpha}}{\Gamma(\alpha v+m+\alpha)}\left\{\frac{\partial^{(n)}}{\partial j^n}\left(\frac{\Gamma(j)}{\beta^{j}}\right)\right\}_{j=\alpha v+m+\alpha}\\ &\xrightarrow{n=1}\\ \mathrm{E}[\log(X)] &= \sum_{v,m=0}^{\infty}w_{v,m}^{(1)}\ \{\psi(\alpha v+m+\alpha)-\log(\beta)\}.\end{aligned}$$

In the following discussion, we derive the Shannon entropy defined by

$$H_S(X) = \mathrm{E}\{-\log[f_{\text{Mc-}\Gamma}(X)]\} = \int_0^\infty -\log[f_{\text{Mc-}\Gamma}(x)] f_{\text{Mc-}\Gamma}(x)\mathrm{d}x.$$

The log-likelihood function relative to one observation follows from (4) as

$$\ell(\alpha,\beta,a,b,c;x) = \log\left[\frac{c\beta^\alpha}{\Gamma(\alpha)B\left(\frac{a}{c},b\right)}\right] + (\alpha-1)\log(x) - \beta x + (a-1)\log[\gamma_1(\alpha,\beta x)] + (b-1)\log[1-\gamma_1(\alpha,\beta x)^c].$$

It is known that the expected value of the score function vanishes, and then from $\mathrm{E}\{\partial\ell(\theta)/\partial a\} = 0$ and $\mathrm{E}\{\partial\ell(\theta)/\partial b\} = 0$, we obtain

$$\mathrm{E}\{\log[\gamma_1(\alpha,\beta X)]\} = \frac{\psi(\frac{a}{c}) - \psi(\frac{a}{c}+b)}{c}$$

and

$$\mathrm{E}\{\log[1-\gamma_1(\alpha,\beta X)^c]\} = \psi(b) - \psi\left(\frac{a}{c}+b\right),$$

respectively.

Hence, the Shannon entropy of the Mc-Γ distribution reduces to

$$H_S(X) = \log\left(\frac{\Gamma(\alpha)B\left(\frac{a}{c},b\right)}{c\beta^\alpha}\right) - (\alpha-1)\sum_{v,m=0}^{\infty} w_{v,m}^{(1)}\{\psi(\alpha v+m+\alpha) - \log(\beta)\}$$
$$+ \sum_{v,m=0}^{\infty} w_{v,m}^{(1)}\,(\alpha v+m+\alpha) + \frac{(a-1)}{c}\left\{\psi\left(\frac{a}{c}\right) - \psi\left(\frac{a}{c}+b\right)\right\} + (b-1)\left\{\psi(b) - \psi\left(\frac{a}{c}+b\right)\right\}.$$

3.6 Means Deviations

The mean deviations of a random variable X with respect to the mean and the median are

$$\delta_1(X) = \int_0^\infty |x-\mu|\, f(x)\,\mathrm{d}x \quad \text{and} \quad \delta_2(X) = \int_0^\infty |x-M|\, f(x)\,\mathrm{d}x,$$

respectively, where $\mu = \mathrm{E}(X)$ and $M = \mathrm{Median}(X)$ denotes the median. These quantities can be expressed as

$$\delta_1(X) = 2\mu F(\mu) - 2\mu + 2T(\mu) \quad \text{and} \quad \delta_2(X) = 2T(M) - \mu,$$

where $F(\mu)$ is the cdf of X and $T(q) = \int_q^\infty x f(x)\mathrm{d}x$.

Based on the Corollary 1, the quantity $T(q)$ for the Mc-Γ distribution becomes

$$\begin{aligned} T(q) &= \sum_{v,m=0}^{\infty} w_{v,m}^{(1)} \int_q^\infty x\ h(x;\alpha v+m+\alpha,\beta)\mathrm{d}x \\ &= \sum_{v,m=0}^{\infty} w_{v,m}^{(1)}\left[1 - \int_0^q x\ h(x;\alpha v+m+\alpha,\beta)\mathrm{d}x\right] \\ &= \sum_{v,m=0}^{\infty} w_{v,m}^{(1)}\left[1 - \frac{(\alpha v+m+v)}{\beta}\ \gamma_1(\alpha v+m+v+1,\beta q)\right] \end{aligned}$$

and $F(x)$ comes from equation (3).

3.7 Bonferroni and Lorentz Curves

Bonferroni ($B(\cdot)$) and Lorentz ($L(\cdot)$) curves have been applied in many fields such as economics, reliability, demography, insurance and medicine. Expressions of these measures for several important probability distributions were proposed by Giorgi & Nadarajah (2010). For the Mc-Γ distribution, these quantities are defined by

$$B(p) = \frac{1}{p\mu}\int_0^q x\, f_{\text{Mc-}\Gamma}(x)\,\mathrm{d}x \quad \text{and} \quad L(p) = \frac{1}{\mu}\int_0^q x\, f_{\text{Mc-}\Gamma}(x)\,\mathrm{d}x = p\,B(p),$$

where $q = F^{-1}(p) = Q_{\text{Mc-}\Gamma}(p)$ is the Mc-Γ quantile function for a given probability p and $\mu = E(X)$. The Mc-Γ quantile function can be calculated by inverting (5) and it will be studied in Section 3.8 (see also equation (8)). We obtain

$$\begin{aligned} B(p) &= \frac{1}{p\mu\beta}\sum_{v,m=0}^{\infty} w_{v,m}^{(1)}\ (\alpha v + m + \alpha)\int_0^q x\,h(x;\alpha v + m + \alpha,\beta)\mathrm{d}x \\ &= \frac{1}{p}\left\{\frac{\sum_{v,m=0}^{\infty} w_{v,m}^{(1)}\ (\alpha v + m + \alpha)\ \gamma_1(\alpha v + m + \alpha,\beta q)}{\sum_{v,m=0}^{\infty} w_{v,m}^{(1)}\ (\alpha v + m + \alpha)}\right\}. \end{aligned}$$

3.8 Skewness and Kurtosis

We can express the Mc-Γ quantile function in terms of the quantile functions of the $\Gamma(\alpha,\beta)$ and $\mathcal{B}\left(\frac{a}{c},b\right)$ distributions, denoted by $Q_\Gamma(p)$ and $Q_{\mathcal{B}}(p)$, respectively. Using (3), we have $F_{\text{Mc-}\Gamma}(x) = I_{\gamma_1(\alpha,\beta x)^c}(\frac{a}{c},b)$. By inverting $I_{\gamma_1(\alpha,\beta x)^c}(\frac{a}{c},b) = p$, we obtain $\gamma_1(\alpha,\beta x)^c = Q_{\mathcal{B}}(p)$, and the Mc-Γ quantile function becomes

$$Q_{\text{Mc-}\Gamma}(p) = Q_\Gamma(Q_{\mathcal{B}}(p)^{1/c}). \tag{8}$$

There are several robust measures in the literature for location and dispersion. The median, for example, can be used for location and the interquartile range. Both the median and the interquartile range are based on quantiles. From this fact, Bowley (1920) proposed a coefficient of skewness based on quantiles given by

$$SK = \frac{Q(3/4) + Q(1/4) - 2Q(1/2)}{Q(3/4) - Q(1/4)},$$

where $Q(\cdot)$ is the quantile function of a given distribution F. It can be shown that Bowley's coefficient of skewness takes the value zero for symmetric distributions. Additionally, its largest value is one and the lowest is -1.

Moors (1986) demonstrated that the conventional measure of kurtosis may be interpreted as a dispersion around the values $\mu + \sigma$ and $\mu - \sigma$. Thus, the probability mass focuses around μ or on the tails of the distribution. Therefore, based on this interpretation, Moors (1988) proposed, as an alternative to the conventional coefficient of kurtosis, a robust measure based on octiles given by

$$KR = \frac{(Q(7/8) - Q(5/8)) + (Q(3/8) - Q(1/8))}{Q(6/8) - Q(2/8)}.$$

Figs. 5(a)-5(b) provide the plots of the Bowley's skewness and Moors's kurtosis, respectively, for the proposed distribution.

3.9 Order Statistics

In the following discussion, we derive the order statistics and their vth moments. The pdf of the ith order statistic $X_{i:n}$, for $i = 1, 2, \ldots, n$, is given by

$$f_{i:n}(x) = \frac{f(x)}{B(i, n-i+1)}\sum_{k=0}^{n-i}(-1)^k\binom{n-i}{k}F^{i+k-1}(x).$$

From equation (18) given in Appendix C, we have

$$f_{i:n}x) = \frac{1}{B(i,n-i+1)}\sum_{k=0}^{n-i}(-1)^k\binom{n-i}{k}\left\{\sum_{h_1,h_2,r,m,s_1,s_2=0}^{\infty} W_{h_1,h_2,r,m,s_1,s_2}^{(i+k-1)}\right\}h(x;\alpha s_1 + s_2 + \alpha + \alpha h_2 + m,\beta). \tag{9}$$

Additionally, the vth ordinary moment of $X_{i:n}$ is

$$\mathrm{E}(X_{i:n}^v) = \int_0^\infty x^v f_{i:n}(x)\mathrm{d}x = \frac{1}{B(i,n-i+1)}\sum_{k=0}^{n-i}(-1)^k\binom{n-i}{k}\underbrace{\int_0^\infty x^v f(x)F^{i+k-1}(x)\mathrm{d}x}_{\mu_{v,i+k-1}\triangleq \mathrm{E}\{X^v F^{i+k-1}(X)\}},$$

where the quantity $\mu_{v,i+k-1}$ is the probability weighted moment (PWM) of the Mc-Γ distribution. From equation (9), we obtain

$$\mathrm{E}(X_{i:n}^v) = \frac{1}{B(i,n-i+1)}\sum_{k=0}^{n-i}(-1)^k\binom{n-i}{k}\mu_{v,i+k-1},$$

where

$$\mu_{v,i+k-1} = \sum_{h_1,h_2,r,m,s_1,s_2=0}^{\infty} W_{h_1,h_2,r,m,s_1,s_2}^{(i+k-1)} \left[\frac{\Gamma(\alpha s_1 + s_2 + \alpha + \alpha h_2 + m + v)}{\Gamma(\alpha s_1 + s_2 + \alpha + \alpha h_2 + m)\beta^v} \right].$$

3.10 Expansion for the Quantile Function

Quantile functions are in widespread use in general statistics and often find representations in terms of lookup tables for key percentiles. An extensive discussion of the use of quantile functions in mainstream statistics is given in the book by Gilchrist (2000).

The quantile function $Q(u)$ usually does not have closed-form expressions for several important distributions, such as the normal, Student t, gamma and beta distributions. As a potential solution, this function can be expressed in terms of a power series of a transformed variable v, which takes the form $v = p\,(q\,u - t)^{\rho}$, for p, q, t and ρ known constants:

$$Q(u) = \sum_{i=0}^{\infty} \epsilon_i\, v^i, \tag{10}$$

where the coefficients ϵ_i are suitably chosen real numbers. Steinbrecher (2002) explored the solution of this equation by standard power series methods.

According to Steinbrecher & Shaw (2008), the following two results hold:

(r1): For the gamma distribution, equation (10) is defined by $v = \beta[\Gamma(\alpha + 1)u]^{1/\alpha}$ and

$$\epsilon_i = m_i = \begin{cases} 0, & \text{if } i = 0 \\ 1, & \text{if } i = 1 \\ a_{i+1}, & \text{if } i \geq 1, \end{cases}$$

where

$$a_{i+1} = \frac{1}{i(\alpha + i)} \Bigg\{ \sum_{r=1}^{i} \sum_{s=1}^{i-s+1} a_r\, a_s\, a_{i-r-s+2}\, s\,(i - r - s + 2) - \Delta(i) \sum_{r=2}^{i} a_r\, a_{i-r+2}\, r\,[r - \alpha - (1 - \alpha)(i + 2 - r)] \Bigg\},$$

$\Delta(i) = 0$ if $i < 2$ and $\Delta(i) = 1$ if $i \geq 2$. Here, the first coefficients are $a_2 = 1/(\alpha+1)$, $a_3 = (3\alpha+5)/[2(\alpha+1)^2(\alpha+2)], \ldots$ Hence, the power series for the gamma quantile is given by

$$Q_{\Gamma}(u) = \sum_{i=0}^{\infty} m_i\, [\beta\, \Gamma(\alpha + 1)^{1/\alpha}]^i\, u^{i/\alpha}. \tag{11}$$

(r2): For the beta quantile, the power series reduces to

$$Q_{\mathcal{B}}(u) = \sum_{i=0}^{\infty} d_i'\, u^{ic/a}, \tag{12}$$

where the transformed variable is $v = [ac^{-1}B(ac^{-1}, b)u]^{ca^{-1}}$, $d_i' = d_i\,[ac^{-1}B(ac^{-1}, b)]^{ic/a}$ and d_i is given by

$$d_i = \begin{cases} 0, & \text{if } i = 0 \\ 1, & \text{if } i = 1 \\ \lambda_i, & \text{if } i \geq 2, \end{cases}$$

$$\lambda_i = \frac{1}{[i^2 + (a/c - 2)i + (1 - a/c)]} \Bigg\{ (1 - \delta_{i,2}) \sum_{r=2}^{i-1} \lambda_r\, \lambda_{i+1-r}\, [r(1 - a/c)(i - r)$$
$$- r(r - 1)] + \sum_{r=1}^{i-1} \sum_{s=1}^{i-r} \lambda_r\, \lambda_s\, \lambda_{i+1-r-s}\, [r(r - a/c) + s(a/c + b - 2)(i + 1 - r - s)] \Bigg\},$$

$\delta_{i,2} = 1$ if $i = 2$ and $\delta_{i,2} = 0$ if $i \neq 2$. The first quantities are:

$$\begin{aligned}
\lambda_2 &= \frac{b-1}{a/c+1},\\
\lambda_3 &= \frac{(b-1)\,(a/c^2 + 3ba/c - a/c + 5b - 4)}{2(a/c+1)^2(a/c+2)},\\
\lambda_4 &= (b-1)[a/c^4 + (6b-1)a/c^3 + (b+2)(8b-5)a/c^2 + (33b^2 - 30b + 4)a/c\\
&\qquad + b(31b - 47) + 18]/[3(a/c+1)^3(a/c+2)(a/c+3)], \ldots
\end{aligned}$$

Inserting equation (11) into (8), we obtain

$$Q_{\text{Mc-}\Gamma}(p) = \sum_{i=0}^{\infty} m_i\,[\beta\,\Gamma(\alpha+1)^{1/\alpha}]^i\,[Q_{\mathcal{B}}(p)]^{i/c\alpha}. \tag{13}$$

Since $0 < Q_{\mathcal{B}}(p) < 1$ and $i/c\alpha > 0$, we have

$$[Q_{\mathcal{B}}(p)]^{i/c\alpha} = \sum_{k,v=0}^{\infty} (-1)^{k+v} \binom{i/c\alpha}{k} \binom{k}{v} [Q_{\mathcal{B}}(p)]^{v}.$$

Additionally, the beta quantile function (12) can be expressed as $p^{c/a} \sum_{j=0}^{\infty} w_j p^{jc/a}$, where $w_j = d'_{j+1} = d_{j+1}[ac^{-1}B(ac^{-1}, b)]^{(j+1)c/a}$ for $j = 0, 1, \ldots$ In this case, the first two quantities are

$$w_0 = [ac^{-1}B(ac^{-1}, b)]^{c/a} \quad \text{and} \quad w_1 = [(b-1)c/(a+c)][ac^{-1}B(ac^{-1}, b)]^{2c/a}.$$

Applying the previous expressions in (13) and using the power series for the beta quantile function in (12), we obtain

$$Q_{\text{Mc-}\Gamma}(p) = \sum_{i,k,v=0}^{\infty} \underbrace{m_i \beta^i\,\Gamma(\alpha+1)^{i/\alpha} (-1)^{k+v} \binom{i/c\alpha}{k} \binom{k}{v}}_{E_{i,k,v}} \left(p^{c/a} \sum_{j=0}^{\infty} w_j p^{jc/a} \right)^{v}.$$

Thus, using a power series raised to a positive integer v (Gradshteyn & Ryzhik, 1980, p. 17), it follows that

$$Q_{\text{Mc-}\Gamma}(p) = \sum_{i,k,v,j=0}^{\infty} E_{i,k,v}\, s_{v,j}\, p^{c(v+j)a^{-1}}.$$

where $s_{v,0} = w_0^v$ and $s_{v,j} = (jw_0)^{-1} \sum_{m=1}^{j} [m(v+1) - j]\, w_m\, s_{v,j-m}$. Notice that the coefficient $s_{v,j}$ can be recursively obtained from $\{s_{v,0}, \ldots, s_{v,j-1}\}$ and $\{w_0, \ldots, w_j\}$.

Finally, setting $\ell = v + j$, the quantile function can be rewritten as

$$Q_{\text{Mc-}\Gamma}(p) = \sum_{\ell=0}^{\infty} N_\ell\, \pi^\ell,$$

where $N_\ell = \sum_{i,k=0}^{\infty} \sum_{v=0}^{\ell} E_{i,k,v}\, s_{v,\ell-v}$ and $\pi = p^{ca^{-1}}$. The last expansion can be used as an alternative way for calculating some mathematical quantities of the new distribution. For example, the result

$$\mathrm{E}(X^n) = \int_0^{\infty} x^n f(x)\mathrm{d}x = \int_0^1 Q(p)^n \mathrm{d}p$$

combined with a power series raised to a positive integer, leads to an alternative form for the ordinary moments

$$\mathrm{E}(X^n) = ac^{-1} \int_0^1 \left(\sum_{\ell=0}^{\infty} N_\ell\, \pi^\ell \right)^n \pi^{ac^{-1}-1} \mathrm{d}\pi = ac^{-1} \sum_{\ell=0}^{\infty} N'_{n,\ell} \int_0^1 \pi^{\ell + ac^{-1} - 1} \mathrm{d}\pi = ac^{-1} \sum_{\ell=0}^{\infty} \frac{N'_{n,\ell}}{\ell + ac^{-1}},$$

where $N'_{n,0} = N_0^n$ and $N'_{n,\ell} = (\ell N_0)^{-1} \sum_{m=1}^{\ell} [m(n+1) - \ell]\, N_m\, N'_{m,\ell-m}$.

4. Estimation

The parameters of the Mc-Γ distribution can be estimated by the method of maximum likelihood. Let $x_1, \cdots, x_n$ be a random sample of size n from the Mc-Γ(α, β, a, b, c) distribution given by (4). The log-likelihood function for the vector of parameters $\theta = (\alpha, \beta, a, b, c)^\top$ can be expressed as

$$\begin{aligned} l(\theta) =& n\log c + n\alpha\log\beta - n\log B\left(\frac{a}{c}, b\right) - n\log\Gamma(\alpha) + (\alpha-1)\sum_{i=1}^{n}\log x_i - \beta\sum_{i=1}^{n} x_i + (a-1)\sum_{i=1}^{n}\log\gamma_1(\alpha,\beta x_i) \\ &+ (b-1)\sum_{i=1}^{n}\log[1-\gamma_1(\alpha,\beta x_i)^c]. \end{aligned}$$

The components of the score vector $\mathbf{U}(\theta)$ are

$$U_\alpha(\theta) = \frac{\partial l(\theta)}{\partial\alpha} = n\log\beta - n\psi(\alpha) + \sum_{i=1}^{n}\log(x_i) + (a-1)\sum_{i=1}^{n}\left\{\frac{\partial\gamma_1(\alpha,\beta x_i)/\partial\alpha}{\gamma_1(\alpha,\beta x_i)}\right\} - c(b-1)\sum_{i=1}^{n}\left\{\frac{\gamma_1(\alpha,\beta x_i)^{c-1}\partial\gamma_1(\alpha,\beta x_i)/\partial\alpha}{1-\gamma_1(\alpha,\beta x_i)^c}\right\},$$

$$U_\beta(\theta) = \frac{\partial l(\theta)}{\partial\beta} = \frac{n\alpha}{\beta} - \sum_{i=1}^{n} x_i + (a-1)\sum_{i=1}^{n} x_i\left\{\frac{\partial\gamma_1(\alpha,\beta x_i)/\partial\beta}{\gamma_1(\alpha,\beta x_i)}\right\} - c(b-1)\sum_{i=1}^{n} x_i\left\{\frac{\gamma_1(\alpha,\beta x_i)^{c-1}\partial\gamma_1(\alpha,\beta x_i)/\partial\beta}{1-\gamma_1(\alpha,\beta x_i)^c}\right\},$$

$$\begin{aligned} U_a(\theta) &= \frac{\partial l(\theta)}{\partial a} = \frac{n}{c}\left[\psi\left(\frac{a}{c}+b\right) - \psi\left(\frac{a}{c}\right)\right] + \sum_{i=1}^{n}\log[\gamma_1(\alpha,\beta x_i)], \\ U_b(\theta) &= \frac{\partial l(\theta)}{\partial b} = n\left[\psi\left(\frac{a}{c}+b\right) - \psi(b)\right] + \sum_{i=1}^{n}\log[1-\gamma_1(\alpha,\beta x_i)^c], \\ U_c(\theta) &= \frac{\partial l(\theta)}{\partial c} = \frac{n}{c} - \frac{na}{c^2}\left[\psi\left(\frac{a}{c}+b\right) - \psi\left(\frac{a}{c}\right)\right] - (b-1)\sum_{i=1}^{n}\left\{\frac{\gamma_1(\alpha,\beta x_i)^c\log\gamma_1(\alpha,\beta x_i)}{1-\gamma_1(\alpha,\beta x_i)^c}\right\}. \end{aligned}$$

These expressions depend on the quantities $\partial\gamma_1(\cdot)/\partial\alpha$ and $\partial\gamma_1(\cdot)/\partial\beta$. Now, we provide formulas for these quantities. Using MATHEMATICA, we obtain

$$\frac{\partial\gamma(\alpha,\beta x)}{\partial\alpha} = \Gamma(\alpha)\psi(\alpha) - \log(\beta x)\gamma(\alpha,\beta x) - G_{2\,3}^{3\,0}\left(\beta x\middle|_{0,\,0,\,\alpha}^{1,\,1}\right),$$

where $G_{2\,3}^{3\,0}(\cdot)$ is a particular case of the Meijer G-function (note 2) given by

$$G_{2\,3}^{3\,0}\left(\beta x\middle|_{0,\,0,\,\alpha}^{1,\,1}\right) = \frac{(\beta x)^\alpha {}_2F_2(\{\alpha,\alpha\};\{\alpha+1,\alpha+1\};-\beta x)}{\alpha^2} - \Gamma(\alpha)\log(\beta x) + \Gamma(\alpha)\psi(\alpha),$$

and ${}_2F_2(\cdot;\cdot;\cdot)$ denotes the hypergeometric function defined by

$${}_2F_2(\{a_1,a_2\};\{b_1,b_2\};z) = \sum_{j=0}^{\infty}\frac{(a_1)_j(a_2)_j}{(b_1)_j(b_2)_j}\frac{z^j}{j!},$$

where, for some parameter μ, the *Pochhammer symbol* $(\mu)_j$ is defined by

$$(\mu)_0 = 1, \quad (\mu)_j = \mu(\mu+1)\cdots(\mu+j-1), \quad j = 1, 2, \ldots$$

We obtain

$$\frac{\partial\gamma(\alpha,\beta x)}{\partial\alpha} = \log(\beta x)\{\Gamma(\alpha) - \gamma(\alpha,\beta x)\} - \frac{(\beta x)^\alpha {}_2F_2(\{\alpha,\alpha\};\{\alpha+1,\alpha+1\};-\beta x)}{\alpha^2}.$$

Finally, we have

$$\frac{\partial\gamma_1(\alpha,\beta x)}{\partial\alpha} = \log(\beta x)\{1-\gamma_1(\alpha,\beta x)\} - \frac{(\beta x)^\alpha {}_2F_2(\{\alpha,\alpha\};\{\alpha+1,\alpha+1\};-\beta x)}{\Gamma(\alpha)\alpha^2} - \psi(\alpha)\gamma_1(\alpha,\beta x).$$

Additionally,

$$\frac{\partial\gamma_1(\alpha,\beta x)}{\partial\beta} = \frac{x^\alpha\beta^{\alpha-1}\exp\{-\beta x\}}{\Gamma(\alpha)}.$$

The maximum likelihood estimate (MLE) $\hat{\theta}$ of θ is the solution of the system of nonlinear equations $\mathbf{U}(\theta) = \mathbf{0}$. For interval estimation and tests of hypotheses on the parameters in θ, we require the 5×5 unit observed information matrix $J = J(\theta)$, whose elements are given in Appendix D. Under certain regularity conditions for the likelihood function, confidence intervals and hypothesis tests can be constructed using the fact that the asymptotic distribution of the MLE $\hat{\theta}$ is $\sqrt{n}(\hat{\theta} - \theta) \sim N_p(0, I(\theta)^{-1})$, where $I(\theta)$ is the Fisher information matrix and p is the number of model parameters (Sen & Singer, 1993). We can substitute $I(\theta)$ by $J(\hat{\theta})$, i.e., the observed information matrix evaluated at $\hat{\theta}$.

The multivariate normal $N_5(0, J(\hat{\theta})^{-1})$ approximated distribution can be used to obtain confidence intervals for the individual parameters. We can compute the maximum values of the unrestricted and restricted log-likelihoods to define likelihood ratio (LR) statistics for testing some sub-models of the Mc-Γ distribution. We may be interested to check if the fit using the Mc-Γ distribution is statistically "superior" to a fit using the $\mathcal{B}$-Γ, Kw-Γ and Γ distributions for a given data set.

5. Application

In this section, we present and compare the performance of the Mc-Γ distribution and its Kw-Γ, $\mathcal{B}$-Γ and Γ sub-models to describe a real data set from USS Halfbeak diesel engine. The data were previously studied by (Ascher, 1984, p. 75) and (Meeker, 1998, p. 415). They represent times of unscheduled maintenance actions for the USS Halfbeak number 4 main propulsion diesel engine over 25.518 operating hours. Table 1 gives some statistical measures for these data. These values indicate that the empirical distribution is *skewed* to the left and *platycurtic*.

In order to compare the fits of the Mc-Γ, $\mathcal{B}$-Γ, Kw-Γ and Γ distributions, the maximum likelihood method was adopted using the subroutine 'NLMixed' in the SAS software. Table 2 lists the MLEs, their standard errors, and three goodness-of-fit statistics: AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) and CAIC (Consistent Akaike Information Criterion). The lowest values of these statistics correspond to the Mc-Γ distribution. Fig. 6 shows that the Mc-Γ distribution provides a closer fit to the histogram of the data than the other three sub-models.

In order to verify the importance of the Mc-Γ distribution in relation to its sub-models by means of hypothesis tests, we consider the LR statistic given by

$$\Lambda = 2\{l(\widehat{\theta}) - l(\widetilde{\theta})\} = 2\{l(\widehat{\alpha}, \widehat{\beta}, \widehat{a}, \widehat{b}, \widehat{c}) - l(\widetilde{\alpha}, \widetilde{\beta}, \widetilde{a}, \widetilde{b}, \widetilde{c})\},$$

where $\widehat{\theta}$ and $\widetilde{\theta}$ are the MLEs of the parameter θ under the alternative and null hypotheses, respectively. The LR statistic can be used to verify if the fit of the Mc-Γ distribution outperforms statistically those fits of their sub-models. Table 3 lists the values of the LR statistics in order to quantify the adequacy of the new distribution. The results provide evidence that the additional parameters of the new distribution are statistically significant in these comparisons, justifying its use for modelling positive real data sets.

6. Conclusions

We present a new five-parameter distribution, called the McDonald gamma (Mc-Γ) distribution, which includes as special cases several commonly used distributions in the literature. Further, the new distribution has proved to be versatile and analytically tractable. We provide a mathematical treatment of this distribution including analytical expressions for the moments, moment generating function, log-moment, mean deviations, Lorentz and Bonferroni curves, order statistics, entropy and quantile function. Additionally, maximum likelihood estimation of the model parameters was discussed and the observed information matrix was derived.

An application to real data was performed in order to quantifying the adequacy of the Mc-Γ distribution and some of its sub-models. The results indicate that the proposed distribution outperforms its main sub-models.

Acknowledgment

The authors are very grateful to a referee and an associate editor for helpful comments that considerably improved the paper. We gratefully acknowledge financial support from CNPq and CAPES.

Appendix

A. Proof of Theorem 1 and Corollary 1: representation in power series for $f_{Mc\text{-}\Gamma}(x)^s$, where s is a positive real number.

First, consider the derivation of the following expression:

$$\delta(\alpha, \beta, a, b, c, s) = \left\{\gamma_1(\alpha, \beta x)^{a-1}[1 - \gamma_1(\alpha, \beta x)^c]^{b-1}\right\}^s.$$

Since $\gamma_1(\alpha,\beta x) < 1$ and $s(a-1)$, $c > 0$, we have

$$[1-\gamma_1(\alpha,\beta x)^c]^{s(b-1)} = \sum_{i=0}^{\infty}(-1)^i[\gamma_1(\alpha,\beta x)]^{ci}\binom{s(b-1)}{i}. \tag{14}$$

Thus,

$$\delta(\alpha,\beta,a,b,c,s) = \sum_{i=0}^{\infty}(-1)^i[\gamma_1(\alpha,\beta x)]^{ci+s(a-1)}\binom{s(b-1)}{i}.$$

The kernel of the above function can be represented as

$$\begin{aligned}[\gamma_1(\alpha,\beta x)]^{ci+s(a-1)} &= \{1-[1-\gamma_1(\alpha,\beta x)]\}^{ci+s(a-1)}\\ &= \sum_{k=0}^{\infty}(-1)^k[1-\gamma_1(\alpha,\beta x)]^k\binom{ci+s(a-1)}{k}\\ &= \sum_{k=0}^{\infty}(-1)^k\binom{ci+s(a-1)}{k}\sum_{v=0}^{\infty}(-1)^v\gamma_1(\alpha,\beta x)^v\binom{k}{v}.\end{aligned}$$

Combining this result with the power series (7), we obtain

$$\delta(\alpha,\beta,a,b,c,s) = \sum_{i,k,v=0}^{\infty}(-1)^{i+k+v}\binom{k}{v}\binom{s(b-1)}{i}\binom{ci+s(a-1)}{k}\left\{\frac{(\beta x)^\alpha}{\Gamma(\alpha)}\right\}^v\left\{\sum_{m=0}^{\infty}\frac{(-\beta)^m x^m}{(\alpha+m)m!}\right\}^v.$$

We consider the result concerning a power series raised to a positive integer v (Gradshteyn & Ryzhik, 1980, p. 17) given by

$$\left\{\sum_{m=0}^{\infty}\underbrace{\frac{(-\beta)^m}{(\alpha+m)m!}}_{a_m}x^m\right\}^v = \sum_{m=0}^{\infty}t_{m,v}\ x^m,$$

where $t_{0,v} = \alpha^{-v}$ and $t_{m,v} = m^{-1}\alpha\sum_{h=1}^{m}[h(v+1)-m]\ a_h\ t_{m-h,v}$ for $m \geq 1$. Hence,

$$\delta(\alpha,\beta,a,b,c,s) = \sum_{i,k,v,m=0}^{\infty}(-1)^{i+k+v}\binom{k}{v}\binom{s(b-1)}{i}\binom{ci+s(a-1)}{k}\left\{\frac{\beta^{v\alpha}x^{\alpha v+m}t_{m,v}}{\Gamma(\alpha)^v}\right\}. \tag{15}$$

Now, we provide a power series for $f_{\text{Mc-}\Gamma}(x)^s$. From equation (15), this quantity can be rewritten as

$$\begin{aligned}&f_{\text{Mc-}\Gamma}(x;\alpha,\beta,a,b,c,)^s = \left\{\frac{c\beta^\alpha x^{\alpha-1}\exp(-\beta x)}{\Gamma(\alpha)B(\frac{a}{c},b)}\right\}^s\delta(\alpha,\beta,a,b,c,s)\\ &= \underbrace{\sum_{v,m=0}^{\infty}\left\{\frac{\Gamma(\alpha v+m+1+s(\alpha-1))}{\beta^{m+1-s}\Gamma(\alpha)^v}t_{m,v}\right\}\sum_{i,k=0}^{\infty}(-1)^{i+k+v}\binom{k}{v}\binom{s(b-1)}{i}\binom{ci+s(a-1)}{k}\Bigg\}}_{w_{v,m}^{(s)}}\times\underbrace{\frac{\beta^{\alpha v+m+1+s(\alpha-1)}x^{\alpha v+m+1+s(\alpha-1)-1}\exp(-\beta x)}{\Gamma(\alpha v+m+1+s(\alpha-1))}}_{h(x;\alpha v+m+1+s(\alpha-1),\beta)}\\ &= \sum_{v,m=0}^{\infty}w_{v,m}^{(s)}\ h(x;\alpha v+m+1+s(\alpha-1),\beta),\end{aligned}$$

where $h(x;\cdot,\cdot)$ denotes the gamma density function. From this result, we have

$$f_{\text{Mc-}\Gamma}(x;\alpha,\beta,a,b,c,) = \sum_{v,m=0}^{\infty}w_{v,m}^{(1)}\ h(x;\alpha v+m+\alpha,\beta), \tag{16}$$

i.e., the density function of the Mc-Γ distribution is a double linear combination of gamma density functions.

B. Representation for $f_{Mc\text{-}\Gamma}(x)$ and $F_{Mc\text{-}\Gamma}(x)$ as functions of the exponentiated gamma distribution.

Applying the result (14) for $s = 1$ in (4), this density can be rewritten as

$$f_{\text{Mc-}\Gamma}(x;\alpha,\beta,a,b,c) = \sum_{k=0}^{\infty}\left[(-1)^k\binom{b-1}{k}\right]\left[\frac{c\beta}{\beta(\frac{a}{c},b)}\right]\frac{(\beta x)^{\alpha-1}\exp(-\beta x)}{\Gamma(\alpha)}[\gamma_1(\alpha,\beta x)]^{ck+a-1}.$$

Setting $u = \beta x$, we have

$$
\begin{aligned}
& f_{\text{Mc-}\Gamma}(u;\alpha,\beta,a,b,c) = \\
&= \sum_{k=1}^{\infty}\left[(-1)^k\binom{b-1}{k}\right]\left[\frac{c\beta}{\beta(\frac{a}{c},b)}\right]\frac{u^{\alpha-1}\exp(-u)}{\Gamma(\alpha)}[\gamma_1(\alpha,u)]^{ck+a-1} \\
&= \sum_{k=0}^{\infty}\left[(-1)^k\binom{b-1}{k}\right]\left[\frac{c\beta}{(ck+a)\beta(\frac{a}{c},b)}\right]\underbrace{(ck+a)\frac{u^{\alpha-1}\exp(-u)}{\Gamma(\alpha)}[\gamma_1(\alpha,u)]^{ck+a-1}}_{h_{\text{EG}}(u;\alpha,ck+a)} \\
&= \sum_{k=0}^{\infty}\underbrace{\left[(-1)^k\binom{b-1}{k}\right]\left[\frac{c\beta}{(ck+a)B(\frac{a}{c},b)}\right]}_{w'_k}h_{\text{EG}}(u;\alpha,ck+a),
\end{aligned}
$$

where $h_{\text{EG}}(x;k_1,k_2)$ is given by

$$h_{\text{EG}}(x;k_1,k_2) = \frac{k_2 x^{k_1-1}\exp(-x)}{\Gamma(k_1)}\left\{\frac{\gamma(k_1,x)}{\Gamma(k_1)}\right\}^{k_2-1},$$

which is the exponentiated standard gamma density with parameters $k_1,k_2 > 0$ (termed here by $\text{EG}(k_1,k_2)$) (Nadarajah & Kotz, 2006). In this case, its cdf can be expressed as

$$F_{\text{Mc-}\Gamma}(u;\alpha,\beta,a,b,c) = \sum_{k=0}^{\infty} w'_k\, H_{\text{EG}}(u;\alpha,ck+a) = \sum_{k=0}^{\infty} w'_k\,[\gamma_1(\alpha,\beta u)]^{ck+a}. \tag{17}$$

C. A linear combination for the quantity $f_{Mc\text{-}\Gamma}(x)F_{Mc\text{-}\Gamma}(x)^{v1}$, where v_1 is a positive integer number.

From equations (16) and (17), we can write

$$f_{Mc\text{-}\Gamma}(x)F_{Mc\text{-}\Gamma}(x)^{v_1} = \left[\gamma_1^a(\alpha,\beta u)\right]^{v_1}\left[\sum_{s_1,s_2=0}^{\infty} w^{(1)}_{s_1,s_2}\; h(x;\alpha s_1+s_2+\alpha,\beta)\right]\left\{\sum_{k=0}^{\infty} w'_k\;[\gamma_1(\alpha,\beta u)]^{ck}\right\}^{v_1}.$$

Since v_1 is a positive integer number, we obtain

$$f_{\text{Mc-}\Gamma}(x)F_{\text{Mc-}\Gamma}(x)^{v_1} = \left[\sum_{s_1,s_2=0}^{\infty} w^{(1)}_{s_1,s_2}\; h(x;\alpha s_1+s_2+\alpha,\beta)\right]\left\{\sum_{k=0}^{\infty} s_{k,v_1}\;[\gamma_1(\alpha,\beta u)]^{ck+av_1}\right\},$$

where $s_{0,v_1} = {w'_0}^{v_1}$ and $s_{k,v_1} = ({w'_0}^{v_1}k)^{-1}\sum_{i=1}^{k}(iv_1-k-i)w'_i s_{k-i,v_1}$ for $k \geq 1$.

Now, following similar arguments of the result (15), we have

$$[\gamma_1(\alpha,\beta x)]^{ck+av_1} = \sum_{h_1,h_2,m=0}^{\infty}(-1)^{h_1+h_2}\binom{h_1}{h_2}\binom{ck+av_1}{h_1}\left\{\frac{t_{m,h_2}x^{\alpha h_2+m}\beta^{\alpha h_2}}{\Gamma(\alpha)^{h_2}}\right\}.$$

Thus,

$$
\begin{aligned}
& f_{\text{Mc-}\Gamma}(x)F_{\text{Mc-}\Gamma}(x)^{v_1} = \\
&= \sum_{h_1,h_2,k,m,s_1,s_2=0}^{\infty} s_{k,v_1}w^{(1)}_{s_1,s_2}(-1)^{h_1+h_2}\binom{h_1}{h_2}\binom{ck+av_1}{h_1}\left\{\frac{t_{m,h_2}\beta^{\alpha h_2}}{\Gamma(\alpha)^{h_2}}\right\}\underbrace{x^{\alpha h_2+m}h(x;\alpha s_1+s_2+\alpha,\beta)}_{\frac{\Gamma(\alpha s_1+s_2+\alpha+\alpha h_2+m)}{\beta^{\alpha h_2+m}\Gamma(\alpha s_1+s_2+\alpha)}h(x;\alpha s_1+s_2+\alpha+\alpha h_2+m,\beta)} \\
&= \sum_{h_1,h_2,k,m,s_1,s_2=0}^{\infty} W^{(v_1)}_{h_1,h_2,k,m,s_1,s_2}\; h(x;\alpha s_1+s_2+\alpha+\alpha h_2+m,\beta),
\end{aligned} \tag{18}
$$

where

$$W^{(v_1)}_{h_1,h_2,k,m,s_1,s_2} = s_{k,v_1}w^{(1)}_{s_1,s_2}(-1)^{h_1+h_2}\binom{h_1}{h_2}\binom{ck+av_1}{h_1}\left\{\frac{t_{m,h_2}\,\Gamma(\alpha s_1+s_2+\alpha+\alpha h_2+m)}{\beta^m\Gamma(\alpha s_1+s_2+\alpha)\Gamma(\alpha)^{h_2}}\right\}.$$

D. Information Matrix.

The elements of the observed information matrix $J(\theta)$ for the parameters (α, β, a, b, c) are:

$$J_{\alpha\alpha} = -n\psi'(\alpha) + (a-1)\sum_{i=1}^{n} y_{\alpha\alpha 1}(x_i) - c(b-1)\sum_{i=1}^{n} y_{\alpha\alpha 2}(x_i),$$

$$J_{\alpha\beta} = \frac{n}{\beta} + (a-1)\sum_{i=1}^{n} x_i\, y_{\alpha\beta 1}(x_i) - c(b-1)\sum_{i=1}^{n} x_i\, y_{\alpha\beta 2}(x_i),$$

$$J_{\alpha a} = \sum_{i=1}^{n} y_{\alpha a}(x_i),$$

$$J_{\alpha b} = -c\sum_{i=1}^{n} y_{\alpha b}(x_i),$$

$$J_{\alpha c} = -(b-1)\left[\sum_{i=1}^{n} y_{\alpha b}(x_i) + c\sum_{i=1}^{n} y_{\alpha c}(x_i)\right], J_{\beta\beta} =$$

$$J_{\beta a} = \sum_{i=1}^{n} x_i\, y_{\beta a}(x_i),$$

$$J_{\beta b} = -c\sum_{i=1}^{n} x_i\, y_{\beta b}(x_i),$$

$$J_{\beta c} = -(b-1)\left[\sum_{i=1}^{n} x_i\, y_{\beta b}(x_i) + c\sum_{i=1}^{n} x_i\, y_{\beta c2}(x_i)\right],$$

$$J_{aa} = \frac{n}{c^2}\left[\psi'\left(\frac{a}{c}+b\right) - \psi'\left(\frac{a}{c}\right)\right],$$

$$J_{ab} = \frac{n}{c}\psi'\left(\frac{a}{c}+b\right),$$

$$J_{ac} = \frac{n}{c^2}\left[\psi\left(\frac{a}{c}\right) - \psi\left(\frac{a}{c}+b\right)\right] + \frac{na}{c^3}\left[\psi'\left(\frac{a}{c}\right) - \psi'\left(\frac{a}{c}+b\right)\right],$$

$$J_{bb} = n\left[\psi'\left(\frac{a}{c}+b\right) - \psi'(b)\right],$$

$$J_{bc} = -\frac{na}{c^2}\psi'\left(\frac{a}{c}+b\right) - \sum_{i=1}^{n} y_{bc}(x_i),$$

$$J_{cc} = -\frac{n}{c^2} + \frac{2na}{c^3}\left[\psi\left(\frac{a}{c}+b\right) - \psi\left(\frac{a}{c}\right)\right] + \frac{na^2}{c^4}\left[\psi'\left(\frac{a}{c}+b\right) - \psi'\left(\frac{a}{c}\right)\right] - (b-1)\sum_{i=1}^{n} y_{cc}(x_i),$$

where the quantities $y_{\alpha\alpha 1}(x_i)$, $y_{\alpha\alpha 2}(x_i)$, $y_{\alpha\beta 1}(x_i)$, $y_{\alpha\beta 2}(x_i)$, $y_{\alpha a}(x_i)$, $y_{\alpha b}(x_i)$, $y_{\alpha c}(x_i)$, $y_{\beta\beta 1}(x_i)$, $y_{\beta\beta 2}(x_i)$, $y_{\beta a}(x_i)$, $y_{\beta b}(x_i)$, $y_{\beta c}(x_i)$, $y_{bc}(x_i)$ and $y_{cc}(x_i)$ are given in Appendix E.

E. Auxiliary terms for the observed information matrix.

$$y_{\alpha\alpha 1}(x_i) = \frac{\gamma_1(\alpha,\beta x_i)\{\partial^2[\gamma_1(\alpha,\beta x_i)]/\partial\alpha^2\} - \{\partial[\gamma_1(\alpha,\beta x_i)]/\partial\alpha\}^2}{\gamma_1(\alpha,\beta x_i)^2},$$

$$y_{\alpha\alpha 2}(x_i) = \left\{\frac{\partial[\gamma_1(\alpha,\beta x_i)]}{\partial\alpha}\right\}^2 \frac{[\gamma_1(\alpha,\beta x_i)^{c-2}(c-1) + \gamma_1(\alpha,\beta x_i)^{2(c-1)}]}{[1-\gamma_1(\alpha,\beta x_i)^c]^2} + \frac{\partial^2[\gamma_1(\alpha,\beta x_i)]}{\partial\alpha^2}\frac{[\gamma_1(\alpha,\beta x_i)^{c-1} - \gamma_1(\alpha,\beta x_i)^{2c-1}]}{[1-\gamma_1(\alpha,\beta x_i)^c]^2},$$

$$y_{\alpha\beta 1}(x_i) = \frac{\gamma_1(\alpha,\beta x_i)\{\partial^2[\gamma_1(\alpha,\beta x_i)]/\partial\alpha\partial\beta\} - \{\partial[\gamma_1(\alpha,\beta x_i)]/\partial\alpha\}\{\partial[\gamma_1(\alpha,\beta x_i)]/\partial\beta\}}{\gamma_1(\alpha,\beta x_i)^2},$$

$$y_{\alpha\beta 2}(x_i) = \frac{\partial[\gamma_1(\alpha,\beta x_i)]}{\partial\alpha}\frac{\partial[\gamma_1(\alpha,\beta x_i)]}{\partial\beta}\frac{[\gamma_1(\alpha,\beta x_i)^{c-2}(c-1) + \gamma_1(\alpha,\beta x_i)^{2(c-1)}]}{[1-\gamma_1(\alpha,\beta x_i)^c]^2} + \frac{\partial^2[\gamma_1(\alpha,\beta x_i)]}{\partial\alpha\partial\beta}\frac{[\gamma_1(\alpha,\beta x_i)^{c-1} - \gamma_1(\alpha,\beta x_i)^{2c-1}]}{[1-\gamma_1(\alpha,\beta x_i)^c]^2},$$

$$y_{\alpha a}(x_i) = \frac{\partial[\gamma_1(\alpha,\beta x_i)]/\partial\alpha}{\gamma_1(\alpha,\beta x_i)},$$

$$y_{\alpha b}(x_i) = \frac{\gamma_1(\alpha,\beta x_i)^{c-1}\{\partial[\gamma_1(\alpha,\beta x_i)]/\partial\alpha\}}{1-\gamma_1(\alpha,\beta x_i)^c},$$

$$y_{\alpha c}(x_i) = \frac{\gamma_1(\alpha,\beta x_i)^{c-1}\log[\gamma_1(\alpha,\beta x_i)]\{\partial[\gamma_1(\alpha,\beta x_i)]/\partial\alpha\}}{[1-\gamma_1(\alpha,\beta x_i)^c]^2},$$

$$y_{\beta\beta 1}(x_i) = \frac{\gamma_1(\alpha,\beta x_i)\{\partial^2[\gamma_1(\alpha,\beta x_i)]/\partial\beta^2\} - \{\partial[\gamma_1(\alpha,\beta x_i)]/\partial\beta\}^2}{\gamma_1(\alpha,\beta x_i)^2},$$

$$y_{\beta\beta 2}(x_i) = \left\{\frac{\partial[\gamma_1(\alpha,\beta x_i)]}{\partial\beta}\right\}^2 \frac{[\gamma_1(\alpha,\beta x_i)^{c-2}(c-1)+\gamma_1(\alpha,\beta x_i)^{2(c-1)}]}{[1-\gamma_1(\alpha,\beta x_i)^c]^2} + \frac{\partial^2[\gamma_1(\alpha,\beta x_i)]}{\partial\beta^2}\frac{[\gamma_1(\alpha,\beta x_i)^{c-1}-\gamma_1(\alpha,\beta x_i)^{2c-1}]}{[1-\gamma_1(\alpha,\beta x_i)^c]^2},$$

$$y_{\beta a}(x_i) = \frac{\partial[\gamma_1(\alpha,\beta x_i)]/\partial\beta}{\gamma_1(\alpha,\beta x_i)},$$

$$y_{\beta b}(x_i) = \frac{\gamma_1(\alpha,\beta x_i)^{c-1}\{\partial[\gamma_1(\alpha,\beta x_i)]/\partial\beta\}}{1-\gamma_1(\alpha,\beta x_i)^c},$$

$$y_{\beta c}(x_i) = \frac{\gamma_1(\alpha,\beta x_i)^{c-1}\log[\gamma_1(\alpha,\beta x_i)]\{\partial[\gamma_1(\alpha,\beta x_i)]/\partial\beta\}}{[1-\gamma_1(\alpha,\beta x_i)^c]^2},$$

$$y_{bc}(x_i) = \frac{\gamma_1(\alpha,\beta x_i)^c\log[\gamma_1(\alpha,\beta x_i)]}{1-\gamma_1(\alpha,\beta x_i)^c},$$

$$y_{cc}(x_i) = \frac{\gamma_1(\alpha,\beta x_i)^c\log[\gamma_1(\alpha,\beta x_i)]^2}{[1-\gamma_1(\alpha,\beta x_i)^c]^2}.$$

References

Ascher, H., & Feingold, H. (1984). *Repairable systems reliability*. New York: Marcel Dekker.

Akinsete, A., Famoye, F., & Lee, C. (2008). The beta Pareto distribution. *Statistics, 42*, 547-563. http://dx.doi.org/10.1080/02331880801983876

Alexander, C., Cordeiro, G. M., Ortega, E. M. M., & Sarabia, J. M. (2011). Generalized beta-generated distributions. *Computational statistics and data analysis*. http://dx.doi.org/10.1016/j.csda.2011.11.015

Amoroso, L. (1925). Ricerche intono alla curva dei redditi. *Annali di matematica pura ed applicata, 2*, 123-159. http://dx.doi.org/10.1007/BF02409935

Barreto-Souza, W., Santos, A. H. S., & Cordeiro, G. M. (2010). The beta generalized exponential distribution. *Journal of statistical computation and simulation, 80*, 159-172. http://dx.doi.org/10.1080/00949650802552402

Bobee, B., & Ashkar, F. (1991). *The gamma family and derived distributions applied in hydrology*. Littleton, CO: Water Resources Publications.

Bowley, A. L. (1920). *Elements of statistics*. New York: C. Scribner's sons.

Cordeiro, G. M., & Brito, R. (2012). The beta power distribution. *Brazilian journal of probability and statistics, 26*, 88-112. http://dx.doi.org/10.1214/10-BJPS124

Cordeiro, G. M., & de Castro, M. (2011). A new family of generalized distributions. *Journal of statistical computation and simulation, 81*, 883-898. http://dx.doi.org/10.1080/00949650903530745

Cordeiro, G. M., & Nadarajah, S. (2011). Closed-form expressions for moments of a class of beta generalized distributions. *Brazilian journal of probability and statistics, 25*, 14-33. http://dx.doi.org/10.1214/09-BJPS109

Cordeiro, G. M., Ortega, E. M. M., & Silva, G. O. (2010). The exponentiated generalized gamma distribution with application to lifetime data. *Journal of statistical computation and simulation, 81*, 827-842. http://dx.doi.org/10.1080/00949650903517874

Cox, C. (2008). The generalized f distribution: An umbrella for parametric survival analysis. *Statistics in medicine, 27*, 4301-4312. http://dx.doi.org/10.1002/sim.3292

de Pascoa, M. A. R., Ortega, E. M. M., & Cordeiro, G. M. (2011). The Kumaraswamy generalized gamma distribution with application in survival analysis. *Statistical methodology, 8*, 411-433. http://dx.doi.org/10.1016/j.stamet.2011.04.001

Esteban, J. (1981). *Income-share elasticity, density functions and the size distribution of income*. Mimeographed manuscript, University of Barcelona, Spain.

Eugene, N., Lee, C., & Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in statistics - Theory and methods, 31*, 175-195. http://dx.doi.org/10.1081/STA-120003130

Gilchrist, W. (2000). Statistical modelling with quantile functions. Boca Raton, FL: CRC Press.

Giorgi, G. M., & Nadarajah, S. (2010). Bonferroni and Gini indices for various parametric families of distributions. *Metron - International journal of statistics*, LXVIII, 23-46. ftp://metron.sta.uniroma1.it/RePEc/articoli/2010-1-2.pdf

Gradshteyn, I. S., & Ryzhik, I. M. (1980). *Table of integrals, series, and products*. New York: Academic Press.

Jambunathan, M. (1954). Some properties of beta and gamma distributions. *The annals of mathematical statistics, 25*, 401-405. http://dx.doi.org/10.1214/aoms/1177728800

Jones, M. (2004). Families of distributions arising from distributions of order statistics. *Test, 13*, 1-43. http://dx.doi.org/10.1007/BF02602999

Lee, C., Famoye, F., & Olumolade, O. (2007). Beta-Weibull distribution: Some properties and applications to censored data. *Journal of modern applied statistical methods, 6*, 173-186.

Meeker, W. Q., & Escobar, L. A. (1998). *Statistical methods for reliability data*. New York: John Wiley & Sons.

McDonald, J. (1984). Some generalized functions for the size distribution of income. *Econometrica, 52*, 647-663. http://www.jstor.org/stable/1913469

Moors, J. J. A. (1986). The meaning of kurtosis: Darlington reexamined. *The American Statistician, 40*, 283-284. http://dx.doi.org/10.1080/00031305.1986.10475415

Moors, J. J. A. (1988). A quantile alternative for kurtosis. *Journal of the royal statistical society (Series D), 37*, 25-32. http://www.jstor.org/stable/2348376

Nadarajah, S. (2008). On the use of the generalized gamma distribution. *International journal of electronics, 95*, 1029-1032. http://dx.doi.org/10.1080/00207210802354981

Nadarajah, S., & Gupta, A. K. (2004). The beta Fréchet distribution. *Far east journal of theoretical statistics, 14*, 15-24.

Nadarajah, S., & Gupta, A. K. (2007). A generalized gamma distribution with application to drought data. *Mathematics and computers in simulation, 74*, 1-7. http://dx.doi.org/10.1016/j.matcom.2006.04.004

Nadarajah, S., & Kotz, S. (2004). The beta Gumbel distribution. *Mathematical problems in engineering, 10*, 323-332. http://dx.doi.org/10.1155/S1024123X04403068

Nadarajah, S., & Kotz, S. (2005). The beta exponential distribution. *Reliability engineering & system safety, 91*, 689-697. http://dx.doi.org/10.1016/j.ress.2005.05.008

Nadarajah, S., & Kotz, S. (2006). The exponentiated type distributions. *Acta applicandae mathematicae, 92*, 97-111. http://dx.doi.org/10.1007/s10440-006-9055-0

Pescim, R. R., Demétrio, C. G. B., Cordeiro, G. M., Ortega, E. M. M., & Urbano, M. R. (2006). The beta generalized half-normal distribution. *Computational statistics & data analysis, 54*, 945-957. http://dx.doi.org/10.1016/j.csda.2009.10.007

Sen, P. K., & Singer, J. M. (1993). *Large sample methods in statistics: An introduction with applications*. New York: Chapman and Hall.

Stacy, E. W. (1962). A generalization of the gamma distribution. *The annals of mathematical statistics, 33*, 1187-1192. http://dx.doi.org/10.1214/aoms/1177704481

Steinbrecher, G. (2002). *Taylor expansion for inverse error function around origin*. Working Paper, University of Craiova, Romania.

Steinbrecher, G., & Shaw, W. T. (2008). Quantile mechanics. *European journal of applied mathematics, 19*, 87-112. http://dx.doi.org/10.1017/S0956792508007341

R Development Core Team. (2008). *R: A Language and environment for statistical computing*. R foundation for statistical computing, Vienna. ISBN 3-900051-07-0

Yevjevich, V. (1972). *Probability and statistics in hydrology*. Littleton, CO: Water Resources Publications.

Table 1. Descriptive statistics for diesel engine data from the USS Halfbeak

Mean	Median	Std. Desv.	Variance	Skewness	Kurtosis	Min	Max
19.3997	21.4610	5.8165	33.8322	-1.5104	4.3305	1.3820	25.5180

Table 2. MLEs and Goodness-of-fit measures

Model	Estimates (standard errors)					Goodness-of-fit measures		
	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{a}$	$\widehat{b}$	$\widehat{c}$	AIC	BIC	CAIC
Mc-Γ	99.8654 (0.2942)	2.0299 (0.0221)	0.0421 (0.0058)	200.0400 (60.1782)	0.2796 (0.0135)	444.7	456.1	445.7
Kw-Γ	6.3844 (0.9922)	0.1996 (0.0404)	1 ×	2.4034 (0.0011)	0.0013 (0.0000)	486.4	503.5	495.0
$\mathcal{B}$-Γ	46.4007 (0.0311)	3.4952 (0.0017)	0.1536 (0.0253)	0.0851 (0.0115)	1 ×	496.8	505.9	497.4
Γ	5.8339 (0.9525)	0.3007 (0.0513)	1 ×	1 ×	1 ×	492.9	497.4	493.0

Table 3. LR tests under Mc-Γ parameters based on real data

Model	Hypotheses	Statistic Λ	p-value
Mc-Γ vs Kw-Γ	$H_0 : a = 1$ vs $H_1 : a \neq 1$	51.7	<0.0001
Mc-Γ vs $\mathcal{B}$-Γ	$H_0 : c = 1$ vs $H_1 : c \neq 1$	54.1	<0.0001
Mc-Γ vs Γ	$H_0 : a = b = c = 1$ vs H_1 : not H_0	54.2	<0.0001

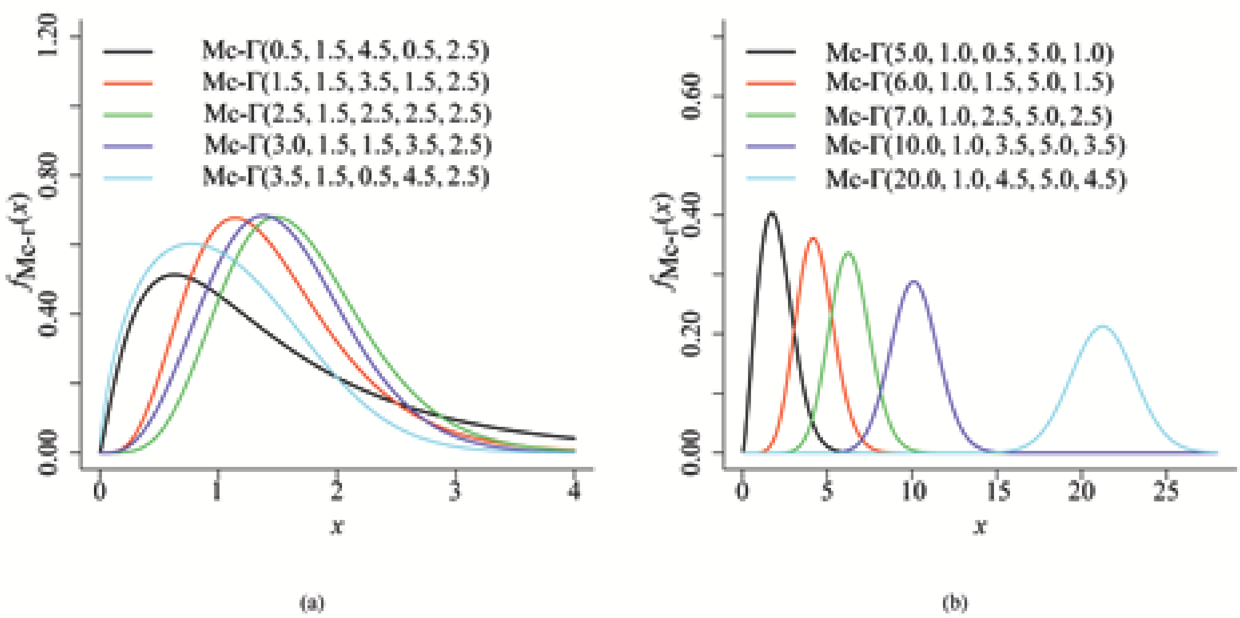


Figure 1. The Mc-Γ density function for some parameter values

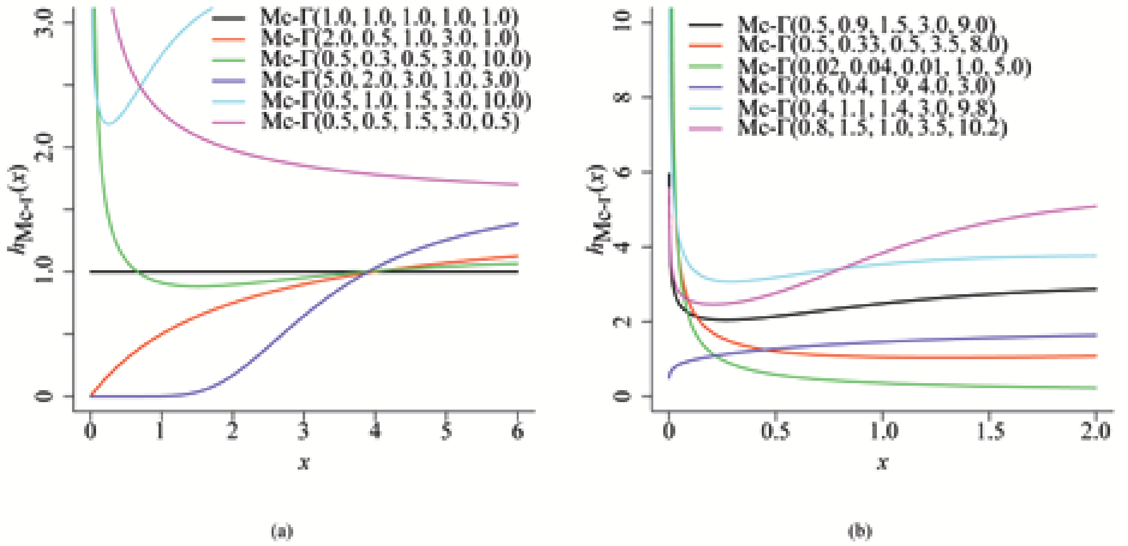


Figure 2. Plots of the Mc-Γ hazard function for some parameter values

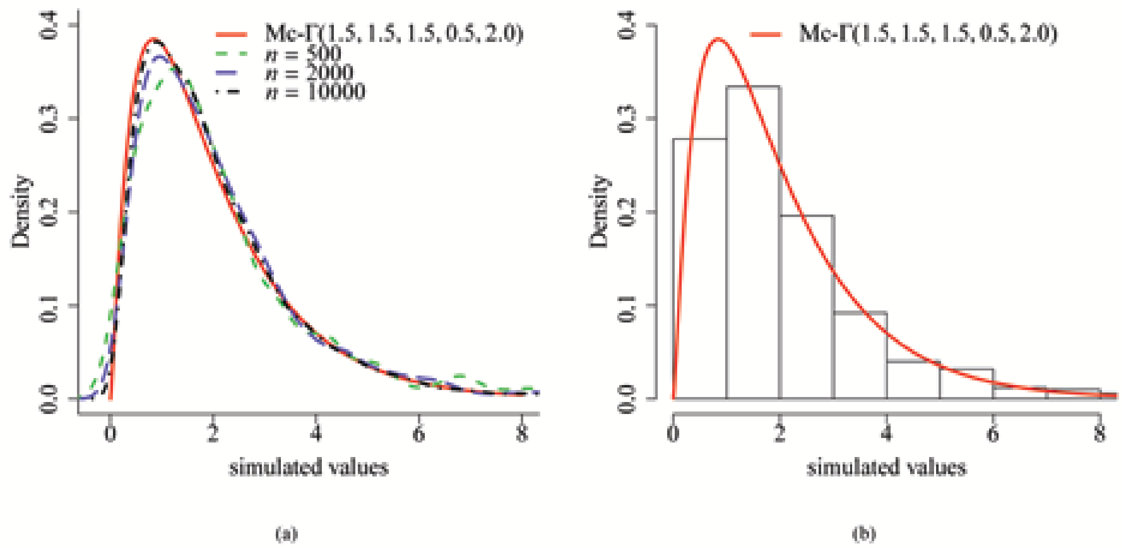


Figure 3. Approximate density (a) and histogram (b) of the values generated from the theoretical density

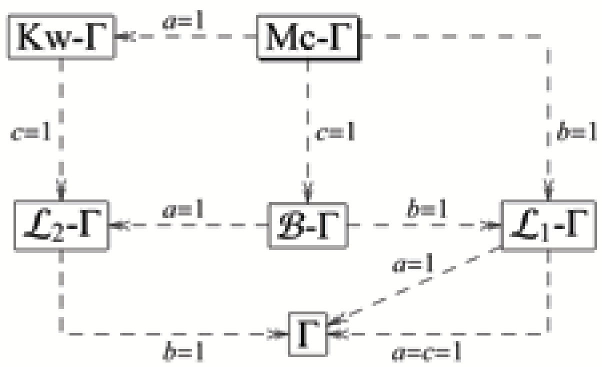


Figure 4. Relationships of the Mc-Γ sub-models

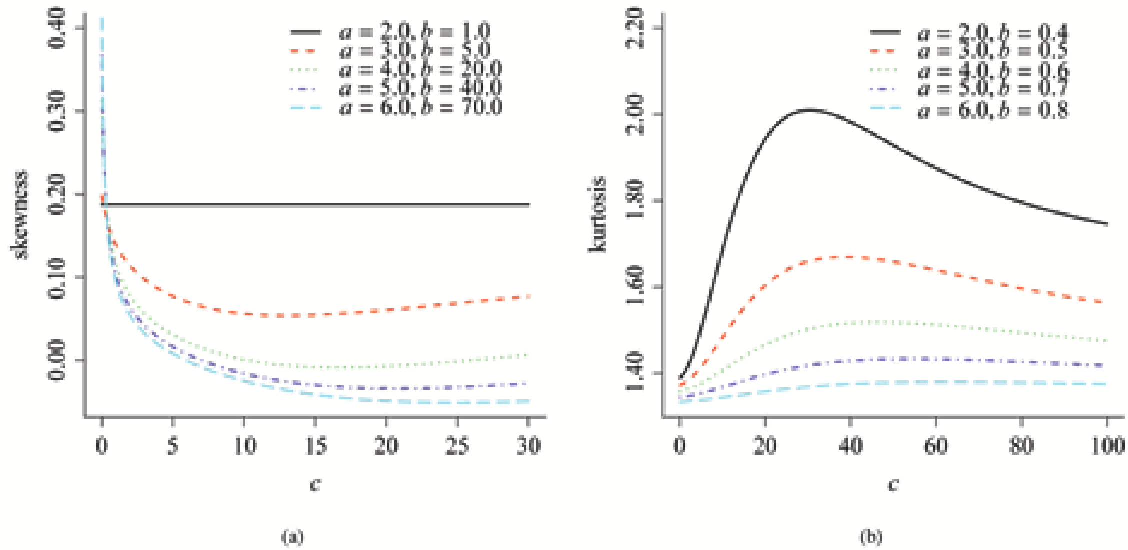


Figure 5. Skewness and kurtosis of the Mc-Γ distribution for different values of c : (a) For $\alpha = 1.0$ and $\beta = 10$ (skewness), (b) For $\alpha = 0.2$ and $\beta = 10$ (kurtosis)

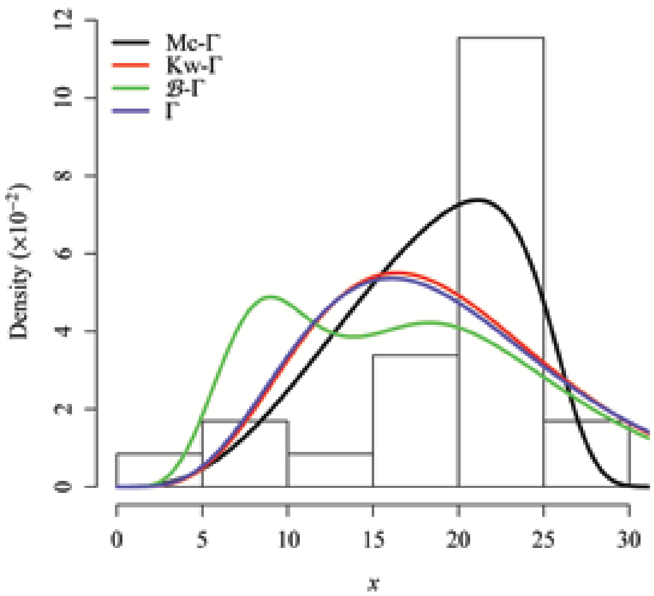


Figure 6. Estimated densities of the Mc-Γ, Kw-Γ, $\mathcal{B}$-Γ and Γ models for USS Halfbeak diesel engine data

Notes

Note 1. This author also is a doctoral student at the Federal University of Pernambuco.

Note 2. This result can be obtained in *Wolfram Alpha* website http://www.wolframalpha.com/

Measure of Departure from Extended Bradley-Terry Model for Paired Comparisons

Kouji Tahata

Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science

Noda City, Chiba 278-8510, Japan

E-mail: kouji_tahata@is.noda.tus.ac.jp

Keigo Kozai

Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science

Noda City, Chiba 278-8510, Japan

E-mail: keigo14@hotmail.co.jp

Sadao Tomizawa

Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science

Noda City, Chiba 278-8510, Japan

E-mail: tomizawa@is.noda.tus.ac.jp

Received: February 20, 2012 Accepted: March 2, 2012 Published: May 1, 2012

doi:10.5539/ijsp.v1n1p72 URL: http://dx.doi.org/10.5539/ijsp.v1n1p72

Abstract

For paired comparisons, we propose a measure to represent the degree of departure from the extended Bradley-Terry model. The measure is expressed by using the Kullback-Leibler information and it ranges between 0 and 1. The measure is applied to the win-loss standings of professional baseball league in Japan.

Keywords: Extended Bradley-Terry model, Kullback-Leibler information, Power-divergence

1. Introduction

Consider the athletic competitions with the outcome for the play of any two teams of R teams, namely a set of data from $R(R-1)/2$ paired comparison. Let π_{ij} for $i \neq j$ denote the probability that team i defeats team j when team i plays team j. Note that $\pi_{ji} = 1 - \pi_{ij}$ for $i < j$; that is, a tie cannot occur.

The Bradley-Terry (BT) model is defined by

$$\pi_{ij} = \frac{\delta_i}{\delta_i + \delta_j} \quad \text{for} \quad i \neq j.$$

This may be expressed by

$$G_{ijk} = G_{kji} \quad \text{for} \quad i < j < k,$$

where

$$G_{ijk} = \pi_{ij}\pi_{jk}\pi_{ki}, \quad G_{kji} = \pi_{kj}\pi_{ji}\pi_{ik};$$

see Bradley and Terry (1952), and Tahata, Miyamoto and Tomizawa (2004).

Assume that R teams are arranged in an order, for example, in order of ranking. The extended Bradley-Terry (EBT) model is defined by

$$\pi_{ij} = \frac{\gamma\delta_i}{\gamma\delta_i + \delta_j} \quad \text{for} \quad i < j.$$

This may be expressed as

$$G_{ijk} = \gamma G_{kji} \quad \text{for} \quad i < j < k;$$

see Davidson and Beaver (1977), and Agresti (1990, p. 373). A special case of this model obtained by putting $\gamma = 1$ is the BT model. This model indicates that for the plays of any two teams of teams i, j and k, the probability that team i defeats team j, team j defeats team k, and team k defeats team i, is γ times higher than the probability that k defeats j, j defeats i, and i defeats k.

For square tables with *nominal* categories, in which cells on the main diagonal are empty, Tahata et al. (2004) proposed the measures to represent the degree of departure from the BT model. We are interested in considering a measure which represents the degree of departure from the EBT model for square tables with *ordered* categories.

Section 2 proposes the measure to represent the degree of departure from the EBT model. Section 3 gives the approximate confidence interval for the measure. Section 4 shows examples.

2. Measure

Let for $i < j < k$,

$$G_{ijk}^{(1)} = \frac{G_{ijk}}{\sum_{s<t<u} G_{stu}}, \quad G_{ijk}^{(2)} = \frac{G_{kji}}{\sum_{s<t<u} G_{uts}}$$

where

$$\sum_{s<t<u} G_{stu} \neq 0, \quad \sum_{s<t<u} G_{uts} \neq 0, \quad G_{ijk} + G_{kji} \neq 0.$$

The EBT model may be expressed as

$$G_{ijk}^{(1)} = G_{ijk}^{(2)} \quad \text{for} \quad i < j < k.$$

Denote any probabilities having the structure of EBT by $\{q_{ij}\}$ with $q_{ij} + q_{ji} = 1$. Then denote $\{G_{ijk}^{(t)}\}$ with $\{\pi_{ij}\}$ replaced by $\{q_{ij}\}$, by $\{Q_{ijk}^{(t)}\}$, $t = 1, 2$. Thus

$$Q_{ijk}^{(1)} = Q_{ijk}^{(2)} \left(= Q_{ijk}^{EBT}\right) \quad \text{for} \quad i < j < k.$$

Consider a measure defined by

$$\Psi = \frac{1}{2\log 2} \min_{\{Q_{ijk}^{EBT}\}} \sum_{t=1}^{2} I\left(\left\{G_{ijk}^{(t)}\right\}; \left\{Q_{ijk}^{EBT}\right\}\right), \tag{1}$$

where

$$I\left(\left\{a_{ijk}\right\}; \left\{b_{ijk}\right\}\right) = \sum_{i<j<k} a_{ijk} \log\left(\frac{a_{ijk}}{b_{ijk}}\right),$$

being the Kullback-Leibler information. Then we can see that Q_{ijk}^{EBT} satisfying (1) are $\bar{Q}_{ijk}^{EBT} = (G_{ijk}^{(1)} + G_{ijk}^{(2)})/2$ for $i < j < k$. Thus, the measure can be expressed as

$$\Psi = \frac{1}{2\log 2} \sum_{t=1}^{2} I\left(\left\{G_{ijk}^{(t)}\right\}; \left\{\frac{G_{ijk}^{(1)} + G_{ijk}^{(2)}}{2}\right\}\right).$$

We see that (i) $0 \leq \Psi \leq 1$, (ii) $\Psi = 0$ if and only if the EBT model holds, and (iii) $\Psi = 1$ if and only if the degree of departure from the EBT model is maximum, in the sense that $G_{ijk}^{(1)} = 0$ (then $G_{ijk}^{(2)} > 0$) for some $i < j < k$ and $G_{ijk}^{(2)} = 0$ (then $G_{ijk}^{(1)} > 0$) for the other $i < j < k$. The maximum degree of departure from EBT can also be expressed as $G_{ijk}/(G_{ijk} + G_{kji}) = 0$ for some $i < j < k$ and $G_{kji}/(G_{ijk} + G_{kji}) = 0$ for the other $i < j < k$. Namely, $\Psi = 1$ indicates that for any three teams of R teams, the conditional probability that team i defeats team j, team j defeats team k, and team k defeats team i on conditional that i defeats j, j defeats k and k defeats i, or i defeats k, k defeats j, j defeats i, is 0 or 1. We shall refer to this situation as "strongest stochastic three way deadlock". Note that from the assumption, $G_{ijk}^{(1)} = 0$ for all $i < j < k$ and $G_{ijk}^{(2)} = 0$ for all $i < j < k$ are excluded from the strongest stochastic three way deadlock. Moreover, since $\Psi = 1$ indicates that $G_{ijk}^{(1)}/G_{ijk}^{(2)} = 0$ for some $i < j < k$ and $G_{ijk}^{(1)}/G_{ijk}^{(2)} = \infty$ for the other $i < j < k$, it seems appropriate to consider that then the degree of departure from EBT (i.e., from $G_{ijk}^{(1)}/G_{ijk}^{(2)} = 1$ for $i < j < k$) is the largest.

3. Approximate Confidence Interval for Measure

Consider a set of data from $R(R-1)/2$ paired comparison experiments for R treatments. Let r_{ij} be the number of comparisons for the treatment pair (i, j), and n_{ij} the number that the treatment i exceeds the treatment j in the r_{ij} comparisons. We assume that there is no tie, i.e., $r_{ij} = r_{ji} = n_{ij} + n_{ji}$. The probability for $\{n_{ij}\}$, $i \neq j$, is then the product of $R(R-1)/2$ binomials. The sample version of Ψ, i.e., $\hat{\Psi}$, is given by Ψ with $\{\pi_{ij}\}$ replaced by $\{\hat{\pi}_{ij}\}$, where $\hat{\pi}_{ij} = n_{ij}/r_{ij}$. Using the delta method (Bishop, Fienberg and Holland, 1975, sec.14.6), $\hat{\Psi}$ has asymptotically a normal distribution with mean Ψ and variance $\sigma^2[\hat{\Psi}]$. The $\sigma^2[\hat{\Psi}]$ is given in Appendix 1.

Let $\hat{\sigma}^2[\hat{\Psi}]$ denote $\sigma^2[\hat{\Psi}]$ with $\{\pi_{ij}\}$ replaced by $\{\hat{\pi}_{ij}\}$. Then, $\hat{\sigma}[\hat{\Psi}]/\sqrt{n}$ is an estimated approximate standard error for $\hat{\Psi}$, and $\hat{\Psi} \pm z_{p/2}\hat{\sigma}[\hat{\Psi}]/\sqrt{n}$ is an approximate $100(1-p)$ percent confidence interval for Ψ, where $z_{p/2}$ is the percentage point from the standard normal distribution corresponding to a two-tail probability equal to p.

4. Examples

Table 1 gives the results of professional baseball league in Japan in 2008 and 2011. These data are obtained from the official website of Japan Professional Baseball (http://www.npb.or.jp/). The categories have the ranking in these years. Namely, for the data in Table 1a, the first is Giants, the second is Tigers and so on. For example, from Giants's perspective, the (Giants, Tigers) result in 2008 correspond to 14 successes and 10 failures in 24 trials.

<Table 1>

The estimated measure $\hat{\Psi}$ are 0.137 for the data in Table 1a and 0.081 for the data in Table 1b. The approximate 95% confidence interval for Ψ are $(0.014, 0.259)$ with standard error 0.063 for the data in Table 1a and $(-0.021, 0.184)$ with standard error 0.052 for the data in Table 1b. Since the confidence interval for Ψ applied to the data in Table 1a do not contain zero, this would indicate that there is not a structure of EBT between the teams in Central league in 2008. On the other hand, since the confidence interval for the measure applied to the data in Table 1b contains zero, this would indicate that there is a structure of EBT between the teams in Central league in 2011; or if this is not the case, then it indicates that the degree of departure from the EBT model is slight.

When the degrees of departure from the EBT model in Tables 1a and 1b are compared using the estimated measures $\hat{\Psi}$, it is greater for Table 1a than for Table 1b. Namely, the data in Table 1a rather than in Table 1b is estimated to be close to the *maximum* departure from the EBT model.

5. Discussions

Consider an $R \times R$ square contingency table with same ordinal row and column classifications. Let p_{ij} denote the probability that an observation will fall in the ith row and the jth column of the table ($i = 1, \ldots, R; j = 1, \ldots, R$). Tomizawa (1984) proposed the extended quasi-symmetry (EQS) model defined by

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

where $\psi_{ij} = \gamma\psi_{ji}$ ($i < j$). Let $p^c_{ij} = p_{ij}/(p_{ij} + p_{ji})$ for $i \neq j$. Then the EQS model may also be expressed as

$$p^c_{ij}p^c_{jk}p^c_{ki} = \gamma p^c_{ji}p^c_{kj}p^c_{ik} \quad \text{for} \quad i < j < k.$$

It is seen that the EQS model is essentially equivalent to the EBT model. Thus we shall define the measure ϕ which represents the degree of departure from the EQS model, by Ψ with $\{\pi_{ij}\}$ replaced by $\{p^c_{ij}\}$.

Let x_{ij} denote the observed frequency in the ith row and the jth column of the table ($i = 1, \ldots, R; j = 1, \ldots, R$). We assume that $\{x_{ij}\}$ have a multinomial distribution. Let $\hat{\phi}$ denote ϕ with $\{p_{ij}\}$ replaced by $\{\hat{p}_{ij}\}$ where $\hat{p}_{ij} = x_{ij}/n$ with $n = \sum\sum x_{ij}$. Using delta method, $\hat{\phi}$ has asymptotically a normal distribution with mean ϕ and variance $\sigma^2[\hat{\phi}]$. The measure $\hat{\Psi}$ is applied to the data obtained from independent binomial sampling, and $\hat{\phi}$ is applied to the data obtained from multinomial sampling. So, $\sigma^2[\hat{\Psi}]$ with $\{\pi_{ij}\}$ replaced by $\{p^c_{ij}\}$, $i \neq j$, is not always identical to $\sigma^2[\hat{\phi}]$. Let $\hat{\sigma}^2[\hat{\phi}]$ denote $\sigma^2[\hat{\phi}]$ with $\{p_{ij}\}$ replaced by $\{\hat{p}_{ij}\}$. Noting that $\{\hat{p}_{ij} + \hat{p}_{ji} = (x_{ij} + x_{ji})/n\}$ in $\hat{\sigma}^2[\hat{\phi}]$, we point out that the estimated variance $\hat{\sigma}^2[\hat{\phi}]$ is theoretically identical to the estimated variance $\hat{\sigma}^2[\hat{\Psi}]$. For more detail, see Tahata et al. (2004).

Note that we can consider a generalized measure for representing the degree of departure from the EBT (EQS) model by using the power-divergence (Cressie and Read, 1984) including the Kullback-Leibler information as follows: for $\lambda > -1$,

$$\Psi^{(\lambda)} = \frac{\lambda(\lambda+1)}{2(2^\lambda - 1)} \sum_{t=1}^{2} I^{(\lambda)}\left(\left\{G^{(t)}_{ijk}\right\}; \left\{\frac{G^{(1)}_{ijk} + G^{(2)}_{ijk}}{2}\right\}\right),$$

where

$$I^{(\lambda)}\left(\left\{a_{ijk}\right\}; \left\{b_{ijk}\right\}\right) = \frac{1}{\lambda(\lambda+1)} \sum_{i<j<k} a_{ijk}\left[\left(\frac{a_{ijk}}{b_{ijk}}\right)^\lambda - 1\right],$$

and the value at $\lambda = 0$ is taken to be the limit as $\lambda \to 0$. When $\lambda = 0$, $\Psi^{(0)}$ is identical to Ψ. The approximate variance of estimated measure $\hat{\Psi}^{(\lambda)}$ is given in Appendix 2.

Consider the data in Table 1, again. Since $\hat{\Psi} = 0.137$ for Table 1a, we can see that the degree of departure from EBT is estimated to be 13.7 percent of the maximum degree of departure from EBT. Similarly, we can infer that the degree of departure from EBT is 8.1 percent of the maximum degree of departure from EBT for the data in Table 1b. Also, we point out that the measure proposed in this paper may be useful to analyze the square contingency tables, for example, social mobility data, paired comparison data, and so on.

6. Concluding Remarks

Since the measure Ψ always ranges between 0 and 1 independent of the number of categories and sample size, it may be useful for comparing the degree of departure from the EBT model in several tables.

The proposed measures would be useful when we want to see with single summary measure what degree the departure from EBT is toward the strongest stochastic three way deadlock, although we cannot see it by the test statistic.

The proposed measures are not invariant under the arbitrary similar permutations of row and column categories. Therefore it is possible to apply these measures for analyzing the data on an *ordered* categories.

Finally, for the data having nominal category, if one wants to measure the degree of departure from **BT**, it is appropriate to use the measure proposed by Tahata et al. (2004). On the other hand, for the data having ordinal categories, if one wants to measure the degree of departure from **EBT**, it is appropriate to use the measure proposed.

Acknowledgements

The authors are grateful to the editor and a referee for their helpful comments and suggestions.

Appendix 1

The variance $\sigma^2[\hat{\Psi}]$ is given as follows:

$$\sigma^2[\hat{\Psi}] = \sum_{a=1}^{R-1} \sum_{b=a+1}^{R} \frac{1}{r_{ab}} \left\{ \frac{1}{\pi_{ab}} (A_{ab})^2 + \frac{1}{\pi_{ba}} (B_{ab})^2 - (A_{ab} + B_{ab})^2 \right\}, \tag{A.1}$$

where

$$\begin{aligned} A_{ab} &= \frac{1}{2\log 2} \sum_{i<j<k} \Big[G_{ijk}^{(1)} (\log H_{ijk}^{(1)}) \Big\{ I_{ij} + I_{jk} - \sum_{s<t<u} G_{stu}^{(1)} (I_{st} + I_{tu}) \Big\} \\ &\quad + G_{ijk}^{(2)} (\log H_{ijk}^{(2)}) \{ I_{ik} - \sum_{s<t<u} G_{stu}^{(2)} I_{su} \} \Big], \end{aligned}$$

with

$$H_{ijk}^{(t)} = \frac{G_{ijk}^{(t)}}{G_{ijk}^{(1)} + G_{ijk}^{(2)}}, \quad I_{ij} = \begin{cases} 1 & (\text{when} \quad i = a \quad \text{and} \quad j = b), \\ 0 & (\text{otherwise}), \end{cases}$$

and B_{ab} is defined by A_{ab} obtained by interchanging $G_{ijk}^{(1)}$ and $G_{ijk}^{(2)}$.

Appendix 2

The variance $\sigma^2[\hat{\Psi}^{(\lambda)}]$ is given by (A.1), where for $\lambda > -1$ and $\lambda \neq 0$,

$$\begin{aligned} A_{ab} &= \frac{2^{\lambda-1}}{2^\lambda - 1} \sum_{i<j<k} \Big[G_{ijk}^{(1)} (H_{ijk}^{(1)})^\lambda \Big\{ I_{ij} + I_{jk} - \sum_{s<t<u} G_{stu}^{(1)} (I_{st} + I_{tu}) \Big\} + G_{ijk}^{(2)} (H_{ijk}^{(2)})^\lambda \Big\{ I_{ik} - \sum_{s<t<u} G_{stu}^{(2)} I_{su} \Big\} \\ &\quad + \lambda \Big((H_{ijk}^{(1)})^{\lambda+1} G_{ijk}^{(2)} - (H_{ijk}^{(2)})^{\lambda+1} G_{ijk}^{(1)} \Big) \Big\{ I_{ij} + I_{jk} - I_{ik} - \sum_{s<t<u} \Big(G_{stu}^{(1)} I_{st} + G_{stu}^{(1)} I_{tu} - G_{stu}^{(2)} I_{su} \Big) \Big\} \Big], \end{aligned}$$

and B_{ab} is defined by A_{ab} obtained by interchanging $G_{ijk}^{(1)}$ and $G_{ijk}^{(2)}$. When $\lambda = 0$, $\sigma^2[\hat{\Psi}^{(0)}]$ is identical to $\sigma^2[\hat{\Psi}]$.

References

Agresti, A. (1990). *Categorical Data Analysis.* New York: Wiley.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Massachusetts: The MIT Press.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs I. The method of paired comparisons. *Biometrika, 39*, 324-345.

Cressie, N. A. C., & Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B, 46*, 440-464.

Davidson, R. R., & Beaver, R. J. (1977). On extending the Bradley-Terry model to incorporate within-pair order effects. *Biometrics, 33*, 693-702. http://dx.doi.org/10.2307/2529467

Tahata, K., Miyamoto, N., & Tomizawa, S. (2004). Measure of departure from quasi-symmetry and Bradley-Terry models for square contingency tables with nominal categories. *Journal of the Korean Statistical Society, 33*, 129-147.

Tomizawa, S. (1984). Three kinds of decompositions for the conditional symmetry model in a square contingency table. *Journal of the Japan Statistical Society, 14*, 35-42.

Table 1. Score sheet of the Central League in Japan in 2008 and 2011

(a) 2008

	Giants	*Tigers*	*Dragons*	*Carp*	*Swallows*	*Baystars*	*Total*
Giants	-	14	10	10	18	18	70
Tigers	10	-	17	14	13	13	67
Dragons	14	6	-	13	9	17	59
Carp	12	10	9	-	12	13	56
Swallows	6	10	13	11	-	15	55
Baystars	5	10	7	11	9	-	42
Total	47	50	56	59	61	76	349

(b) 2011

	Dragons	*Swallows*	*Giants*	*Tigers*	*Carp*	*Baystars*	*Total*
Dragons	-	11	10	13	12	15	61
Swallows	10	-	12	10	13	15	60
Giants	12	8	-	11	16	14	61
Tigers	9	14	11	-	12	12	58
Carp	10	9	6	12	-	17	54
Baystars	8	5	10	10	7	-	40
Total	49	47	49	56	60	73	334

Nonlinear Markov Games on a Finite State Space (Mean-field and Binary Interactions)

Vassili N. Kolokoltsov

Department of Statistics, University of Warwick

Coventry CV4 7AL, UK

E-mail: v.kolokoltsov@warwick.ac.uk

Received: February 13, 2012 Accepted: February 27, 2012 Published: May 1, 2012

doi:10.5539/ijsp.v1n1p77 URL: http://dx.doi.org/10.5539/ijsp.v1n1p77

Supported by the AFOSR grant FA9550-09-1-0664 'Nonlinear Markov control processes and games' and by the Institute of Problems in Informatics of the Russian Academy of Science.

Abstract

Managing large complex stochastic systems, including competitive interests, when one or several players can control the behavior of a large number of particles (agents, mechanisms, vehicles, subsidiaries, species, police units, etc), say N_k for a player k, the complexity of the game-theoretical (or Markov decision) analysis can become immense as $N_k \to \infty$. However, under rather general assumptions, the limiting problem as all $N_k \to \infty$ can be described by a well manageable deterministic evolution. In this paper we analyze some simple situations of this kind proving the convergence of Nash-equilibria for finite games to equilibria of a limiting deterministic differential game.

Keywords: Markov control, Large complex systems, Dynamic law of large numbers, Differential games, Rates of convergence, Nonlinear Markov games

1. Introduction

A steady increase in complexity is one of the characteristic features of the modern technological development. It requires an appropriate (or better optimal) management of complex stochastic systems consisting of large number of interacting components (agents, mechanisms, vehicles, subsidiaries, species, police units, etc), which may have competitive or common interests. Carrying out a traditional Markov decision analysis for a large state space is often unfeasible. However, under rather general assumptions, the limiting problem as the number of components tends to infinity can be described by a well manageable deterministic evolution, which represents a performance of a dynamic law of large numbers (LLN). In general, this limiting deterministic evolution is measure-valued (it is an evolution of probability laws on the initial state space), and its probabilistic analysis has led to the notion of a nonlinear Markov process, see monograph (Kolokoltsov, 2010) and references therein. Its controlled version can be naturally called a nonlinear Markov control process or (in case of competitive interests) a nonlinear Markov game (Kolokoltsov, 2009). In case of finite initial state space, the corresponding space of measures is a finite-dimensional Euclidean space (more precisely its positive orthant $\mathbf{R}^d_+$), so that the limiting measure-valued evolution becomes a deterministic control process or a differential game in $\mathbf{R}^d$. This paper is devoted to the analysis of simplest situations of this kind, aiming at the identification of deterministic limit and proof of convergence with explicit rates. More precisely, we shall assume that there is a fixed number of players $\{1, \cdots, K\}$ each controlling a stochastic system consisting of a large number $N_1, \cdots, N_K \to \infty$ components respectively. These can be generals controlling armies, engineers controlling robot swamps, large banks managers controlling subsidiaries, etc. The components can interact between themselves and with agents of other groups. The limit $N_1, \cdots, N_k \to \infty$ will be described by a differential game in R^K_+.

The plan of the paper is as follows. In a preliminary Section 2 we set the stage by describing the dynamic law of large numbers for interacting Markov chains (without control). This topic is rather well developed by now, but we present it on the level of generality needed for what follows, including time nonhomogeneous chains (with discontinuous dependence on time) and rates of convergence resulting from Hölder continuity assumptions on the r.h.s. of the limiting ODE. Section 3 introduces control without competition. A strong progress for the analysis of such systems was made recently in (Gast, Gaujal & Le Boudec, 2010). We present it in a quite different form (see discussion in the last Section), which paves the road for the extension to competitive interests developed further. Main results are presented in Sections 3-5, where we discuss consecutively two player zero-sum games with mean -field interaction, two player zero-sum games with binary interaction and a K player noncooperative game. The proofs are given in Sections 6 and 7. The last section is devoted to a short review of relevant literature and to further perspectives.

The following (rather standard) notations for functional spaces will be used throughout the paper. For a closed subset Ω of a Euclidean space we shall denote by $C(\Omega)$ the Banach space of bounded continuous functions on Ω equipped with the usual sup-norm (which will be denoted simply $\|.\|$ everywhere), and by $C^k(\Omega)$, $k \in \mathbf{N}$, the Banach space of k times continuously differentiable functions in the interior of Ω with f and all its derivatives up to and including order k having continuous and bounded extension to Ω, equipped with norm $\|f\|_{C^k(\Omega)}$ which is the sum of the sup-norms of f and all its derivatives up to and including order k. Finally, for $\alpha \in (0, 1]$, we denote by $C^{k,\alpha}(\Omega)$ the subspace of $C^k(\Omega)$ consisting of functions, whose kth order derivatives are Hölder continuous of index α. The Banach norm on this space is defined as the sum of the norm in $C^k(\Omega)$ plus the minimal Hölder constant. For an operator Φ in a Banach space B we shall denote by $\|\Phi\|_B$ the corresponding operator norm of Φ.

2. Preliminaries: LLN for Interacting Markov Chains

Let us first recall the basic setting of mean-field interacting particle systems with a finite number of types. Suppose our initial state space is a finite set $\{1, ..., d\}$, which can be interpreted as the types of particles (say, possible opinions of individuals on a certain subject, or the levels of fitness in a military unit, or the types of robots in a robot swamp). Let $\{Q(t, x)\} = \{(Q_{ij})(t, x)\}$ be a family of $d \times d$ square Q-matrices or Kolmogorov matrices (i.e. non-diagonal elements of these matrices are non-negative and the elements of each row sum up to one) depending continuously on a vector x from the closed simplex

$$\Sigma_d = \{x = (x_1, ..., x_d) \in \mathbf{R}^d_+ : \sum_{j=1}^{d} x_j = 1\},$$

and piecewise continuously on time $t \geq 0$. For any x, the family $\{Q(., x)\}$ specifies a Markov chain on the state space $\{1, ..., d\}$ with the generator

$$(Q(t, x)f)_n = \sum_{m\neq n} Q_{nm}(t, x)(f_m - f_n), \quad f = (f_1, \cdots, f_d),$$

and with the intensity of jumps being

$$|Q_{ii}(t, x)| = -Q_{ii}(t, x) = \sum_{j\neq i} Q_{ij}(t, x).$$

In other words, the transition matrices $P(s, t, x) = (P_{ij}(s, t, x))^d_{i,j=1}$ of this chain satisfy the Kolmogorov forward equations

$$\frac{d}{dt} P_{ij}(s, t, x) = \sum_{l=1}^{d} Q_{lj}(t, x)P_{il}(s, t, x), \quad s \leq t.$$

Remark 1. Instead of piecewise continuous dependence on t we can assume that Q is uniformly bounded and depends measurably on t. Everything remains the same. This extension is relevant if one is interested in arbitrary discontinuous controls.

Suppose we have a large number of particles distributed arbitrary among the types $\{1, ..., d\}$. More precisely our state space S is $\mathbf{Z}^d_+$, the set of sequences of d non-negative integers $N = (n_1, ..., n_d)$, where each n_i specifies the number of particles in the state i. Let $|N|$ denote the total number of particles in state N: $|N| = n_1 + ... + n_d$. For $i \neq j$ and a state N with $n_i > 0$ denote by N^{ij} the state obtained from N by removing one particle of type i and adding a particle of type j, that is n_i and n_j are changed to $n_i - 1$ and $n_j + 1$ respectively. The mean-field interacting particle system specified by the family $\{Q\}$ is defined as the Markov process on S specified by the generator

$$L_t f(N) = \sum_{i,j=1}^{d} n_i Q_{ij}(t, N/|N|)[f(N^{ij}) - f(N)]. \tag{1}$$

Probabilistic description of this process is as follows. Starting from any time and current state N one attaches to each particle a $|Q_{ii}|(N/|N|)$-exponential random waiting time (where i is the type of this particle). If the shortest of the waiting times τ turns out to be attached to a particle of type i, this particle jumps to a state j according to the distribution $(Q_{ij}/|Q_{ii}|)(N/|N|)$. Briefly, with this distribution and at rate $|Q_{ii}|(N/|N|)$, any particle of type i can turn (migrate) to a type j.After any such transition the process starts again from the new state N^{ij}. Notice that since the number of particles $|N|$ is preserved by any jump, this process is in fact a Markov chain with a finite state space.

Remark 2. Yet another way of describing the chain generated by L_t is via the forward Kolmogorov (or master) equation for its transition probabilities $P_{MN}(s, t)$:

$$\frac{d}{dt} P_{MN}(s, t) = \sum_{i,j=1}^{d} (n_i + 1)Q_{ij}(t, \frac{N^{ji}}{|N|})P_{MN^{ij}}(s, t) - \sum_{i,j=1}^{d} n_i Q_{ij}(t, \frac{N^{ij}}{|N|})P_{MN}(s, t), \quad s \leq t.$$

To shorten the formulas, we shall denote the inverse number of particles by h, that is $h = 1/|N|$. Normalizing the states to $N/|N| \in \Sigma_d^h$, where Σ_d^h is a subset of Σ_d with coordinates proportional to h, leads to the generator of the form

$$L_t^h f(N/|N|) = \sum_{i=1}^d \sum_{j=1}^d \frac{n_i}{|N|}|N|Q_{ij}(t, N/|N|)[f(N^{ij}/|N|) - f(N/|N|)], \tag{2}$$

or equivalently

$$L_t^h f(x) = \sum_{i=1}^d \sum_{j=1}^d x_i Q_{ij}(t,x)\frac{1}{h}[f(x - he_i + he_j) - f(x)], \quad x \in h\mathbf{Z}_+^d, \tag{3}$$

where $e_1, ..., e_d$ denotes the standard basis in $\mathbf{R}^d$. With some abuse of notation, let us denote by $hN^{t,h}$ the corresponding Markov chain. The transition operators of this chain will be denoted by $\Psi_{s,t}^h$:

$$\Psi_{s,t}^h f(hN) = \mathbf{E}_{s,hN} f(hN(t,h)), \quad s \le t, \tag{4}$$

where $\mathbf{E}_{s,x}$ denotes the expectation of the chain started at x at time s. These operators are known to form a propagator, i.e. they satisfy the chain rule (or Chapman-Kolmogorov equation)

$$\Psi_{s,t}^h \Psi_{t,r}^h = \Psi_{s,t}^h, \quad s \le t \le r.$$

We shall be interested in the asymptotic behavior of these chains as $h \to 0$. To this end, let us observe that, for $f \in C^1(\Sigma_d)$,

$$\lim_{|N|\to\infty,\, N/|N|\to x} |N|[f(N^{ij}/|N|) - f(N/|N|)] = \frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_i}(x),$$

so that

$$\lim_{|N|\to\infty,\, N/|N|\to x} L_t^h f(N/|N|) = \Lambda_t f(x),$$

where

$$\Lambda_t f(x) = \sum_{i=1}^d \sum_{j\ne i} x_i Q_{ij}(t,x)[\frac{\partial f}{\partial x_j} - \frac{\partial f}{\partial x_i}](x) = \sum_{k=1}^d \sum_{i\ne k}[x_i Q_{ik}(t,x) - x_k Q_{ki}(t,x)]\frac{\partial f}{\partial x_k}(x). \tag{5}$$

The limiting operator $\Lambda_t f$ is a first-order PDO with characteristics solving the equation

$$\dot{x}_k = \sum_{i\ne k}[x_i Q_{ik}(t,x) - x_k Q_{ki}(t,x)] = \sum_{i=1}^d x_i Q_{ik}(t,x), \quad k = 1, ..., d, \tag{6}$$

called the *kinetic equations* for the process of interaction described above. The characteristics specify the dynamics of the deterministic time-nonhomogeneous Markov Feller process in Σ_d defined via the generator Λ_t. The corresponding transition operators act on $C(\Sigma_d)$ as

$$\Phi_{s,t} f(x) = f(X_{s,x}(t)), \quad s \le t, \tag{7}$$

where $X_{s,x}(t)$ is the solution to (6) with the initial condition x at time s. These operators form a Feller propagator (i.e. $\Phi_{s,t}$ depend strongly continuous on s, t and satisfy the chain rule $\Phi_{s,t}\Phi_{t,r} = \Phi_{s,r}$, $s \le t \le r$). Of course in case of Q that do not depend on time t explicitly, $\Phi_{s,t}$ depend only on the difference $t - s$ and the operators $\Phi_t = \Phi_{0,t}$ form a Feller semigroup.

Remark 3. It is easy to see that if $x_k \ne 0$, then $(X_{s,x}(t))_k \ne 0$ for any $t \ge s$. Hence the boundary of Σ_d is not attainable for this semigroup, but, depending on Q, it can be glueing or not. For instance, if all elements of Q never vanish, then the points $X_{s,x}(t)$ never belong to the boundary of Σ_d for $t > s$, even if the initial point x does so.

The convergence of the Markov chains with generators of type (2) to a deterministic evolution and various versions of this result are well known, see e.g. (Darling & Norris, 2008; Kolokoltsov, 2010; Benaïm & Le Boudec, 2008) and references therein.

We present here an extension (for time-nonhomogeneous chains with discontinuous time dependence) of a result from (Kolokoltsov, 2011, Sect. 5.11), on a level of generality which allows us to get the corresponding convergence results for controlled problems as more or less straightforward corollaries.

Theorem 1. *(i) Let all the elements $Q_{ij}(t, .)$ belong to $C^{1,\alpha}(\Sigma)$, $\alpha \in (0, 1]$, with norms uniformly bounded in t. Then, if for some $s > 0$ and $x \in \mathbf{R}^d$, the initial data hN_s converge to x in $\mathbf{R}^d$, as $h \to 0$, the Markov chains $hN(t, h)$ with the initial data*

hN_s (generated by L_t^h and with transitions $\Psi_{s,t}$) converge in distribution and in probability to the deterministic characteristic $X_{s,x}(t)$. For the corresponding converging propagators of transition operators the following rates of convergence hold:

$$\sup_{0\le s\le t\le T}\ \sup_{N\in\mathbf{Z}_+^d:|N|=1/h}\left[\Psi_{s,t}^h f(hN)-\Phi_{s,t}f(hN)\right]\le C(T)(t-s)h^\alpha\|f\|_{C^{1,\alpha}(\Sigma_d)}, \tag{8}$$

for $f\in C^{1,\alpha}(\Sigma)$ and

$$\sup_{0\le s\le t\le T}\left[\mathbf{E}_{s,hN}f(hN(t,h))-f(X_{s,x}(t))\right]\le C(T)\left((t-s)h^\alpha\|f\|_{C^{1,\alpha}(\Sigma_d)}+\|f\|_{C^1(\Sigma_d)}|hN-x|\right), \tag{9}$$

where $C(T)$ depends only on the supremum in t of $C^{1,\alpha}(\Sigma)$-norm of the functions $Q(t,x)$.

(ii) Assuming a weaker regularity condition, namely that $Q_{ij}(t,.)$ belong to $C^1(\Sigma)$ uniformly in t, the convergence of Markov chains $hN(t,h)$ in distribution and in probability to the deterministic characteristics still holds, but instead of (8)*, we have weaker rates in terms of the modulus of continuity w_h of ∇f and Q (see* (39) *for the definition):*

$$\sup_{0\le s\le t\le T}\ \sup_{N\in\mathbf{Z}_+^d:|N|=1/h}\left[\Psi_{s,t}^h f(hN)-\Phi_{s,t}f(hN)\right]$$

$$\le C(T)(t-s)\left(w_{hC(T)}(\nabla f)+w_{hC(T)}(\nabla Q)\|f\|_{C^1(\Sigma_d)}\right),0 \tag{10}$$

where $C(T)$ depends on the $C^1(\Sigma)$-norm of Q. A similar modification of (9) *holds.*

Our objective is to extend this result to interacting and competitively controlled families of Markov chains.

3. Mean Field Markov Control

As a warm-up, let us start with mean-field controlled Markov chains without competition. Suppose we are given a family of Q-matrices $\{Q(t,u,x)\}=\{(Q_{ij})(t,u,x),\ i,j=1,\cdots d\}$, depending on $x\in\Sigma_d$, $t\ge 0$ and a parameter u from a metric space interpreted as control. The main assumption will be that $Q\in C^{1,\alpha}(\Sigma_d)$ as a function of x with the norm bounded uniformly in t,u, and Q depends continuously on t and u.

Any given bounded measurable curve $u(t)$, $t\in[0,T]$, defines a Markov chain on Σ_d^h with the time-dependent family of generators of type (2), that is

$$L_{t,u(t)}f\left(\frac{N}{|N|}\right)=\sum_{i,j}^d n_iQ_{ij}\left(t,u(t),\frac{N}{|N|}\right)\left[f\left(\frac{N^{ij}}{|N|}\right)-f\left(\frac{N}{|N|}\right)\right], \tag{11}$$

or equivalently

$$L_{t,u(t)}^h f(x)=\sum_{i=1}^d\sum_{j=1}^d x_iQ_{ij}(t,u(t),x)\frac{1}{h}[f(x-he_i+he_j)-f(x)]. \tag{12}$$

For simplicity (and effectively without loss of generality), we shall stick further to controls $u(.)$ from the class $C_{pc}[0,T]$ of piecewise-continuous curves (with a finite number of discontinuities).

Again for $f\in C^1(\Sigma_d)$,

$$\lim_{h=1/|N|\to 0,\,N/|N|\to x} L_{t,u(t)}^h f(N/|N|)=\Lambda_{t,u(t)}f(x),$$

where

$$\Lambda_{t,u(t)}f(x)=\sum_{k=1}^d\sum_{i\ne k}[x_iQ_{ik}(t,u(t),x)-x_kQ_{ki}(t,u(t),x)]\frac{\partial f}{\partial x_k}(x), \tag{13}$$

with the corresponding controlled characteristics governed by the equations

$$\dot{x}_k=\sum_{i\ne k}[x_iQ_{ik}(t,u(t),x)-x_kQ_{ki}(t,u(t),x)]=\sum_{i=1}^d x_iQ_{ik}(t,u(t),x),\quad k=1,...,d. \tag{14}$$

For a given $T>0$ and continuous functions J (current payoff) and V_T (terminal payoff), let $\Gamma(T,h)$ denote the problem of a centralized controller of the chain with $|N|=1/h$ particles, aiming at maximizing the payoff

$$\int_0^T J\left(s,u(s),\frac{N(s,h)}{|N|}\right)ds+V_T\left(\frac{N(T,h)}{|N|}\right). \tag{15}$$

The optimal payoff will be denoted by $V^h(t,x)$:

$$V^h(t,x) = \sup_{u(.)\in C_{pc}[t,T]} \mathbf{E}_{t,x}^{u(.)} \left[\int_t^T (J(s,u(s),hN(s,h))ds + V_T(hN(T,h)) \right], \tag{16}$$

where $E_{t,x}^{u(.)}$ denotes the expectation with respect to the Markov chain on Σ_d^h generated by (11) and started at $x = hN$ at time t.

We are aiming at approximating $V^h(t,x)$ by the optimal payoff

$$V(t,x) = \sup_{u(.)\in C_{pc}[t,T]} \left[\int_t^T J(s,u(s),X_{t,x}(s))\,ds + V_T(X_{t,x}(T)) \right] \tag{17}$$

for the controlled dynamics (14).

We can also obtain approximate optimal synthesis for problems $\Gamma(T,h)$ with large $|N| = 1/h$, at least if regular enough synthesis is available for the limiting system. Let us recall that a function $\gamma(t,x)$ is called an optimal synthesis (or an adaptive policy) for the problem $\Gamma(T,h)$ if

$$V^h(t,x) = \mathbf{E}_{t,x}^{\gamma} \left[\int_t^T (J(s,\gamma(s,hN(s,h)),hN(s))ds + V_T(hN(T,h)) \right] \tag{18}$$

for all $t \le T$ and $x \in \Sigma_d^h$, where $E_{t,x}^{\gamma}$ denotes the expectation with respect to the Markov chain on Σ_d^h generated by (11) with $u(t) = \gamma(t,x)$ and starting at $x = hN$ at time t. A function $\gamma(t,x)$ is called an ϵ-optimal synthesis or an ϵ-adaptive policy, if the r.h.s. of (18) differs from its l.h.s. by not more than ϵ. Similarly an optimal synthesis or an adaptive policy are defined for the limiting deterministic system.

Theorem 2. *(i) Assume that Q, J depend continuously on t, u and $Q, J \in C^{1,\alpha}(\Sigma_d)$, $\alpha \in (0,1]$, as functions of x, with the norms bounded uniformly in t, u, and finally $V_T \in C^{1,\alpha}(\Sigma_d)$. Then*

$$\sup_{0\le t\le T} [V^h(t,hN) - V(t,x)]$$

$$\le C(T)((T-t)h^{\alpha} + |hN - x|) \left(\|V_T\|_{C^{1,\alpha}(\Sigma_d)} + \sup_{s,u} \|J(t,u,.)\|_{C^{1,\alpha}(\Sigma_d)} \right), \tag{19}$$

with $C(T)$ depending only on the bounds of the norms of Q in $C^{1,\alpha}(\Sigma_d)$. Moreover, if $u(t)$ is an ϵ-optimal control for deterministic dynamics (14)*, that is the payoff obtained by using $u(.)$ differs by ϵ from $V(t,x)$, then $u(.)$ is also an $(\epsilon + C(T)h^{\alpha})$-optimal control for $|N| = 1/h$ particle system.*

(ii) Suppose additionally that u belong to a convex subset of a Euclidean space and that $Q(t,u,x)$ depends Lipschitz continuously on u. Let $\epsilon \ge 0$, and let $\gamma(t,x)$ be a Lipschitz continuous function of x uniformly in t that represents an ϵ-optimal synthesis for the limiting deterministic control problem. Then, for any $\delta > 0$, there exists h_0 such that, for $h \le h_0$, $\gamma(t,x)$ is an $(\epsilon + \delta)$-optimal synthesis for the approximate optimal problem $\Gamma(T,h)$ on Σ_d^h.

Remark 4. As in Theorem (ii), there is a version of Theorem with $Q \in C^1(\Sigma_d)$ and $V_T \in C^1(\Sigma_d)$. We omit the detail. The same remark concerns other theorems given below.

Notice finally that by the standard dynamic programming, see e.g. (Fleming & Soner, 2006; McEneaney, 2006), the optimal payoff $V(t,x)$ given by (17) represents the unique viscosity solution of the HJB-Isaacs equation

$$\frac{\partial V}{\partial t}(t,x) + \max_u \left[J(t,u,x) + \sum_{i,k=1}^{d} x_i Q_{ik}(u,x) \frac{\partial V}{\partial x_k}(t,x) \right] = 0, \tag{20}$$

and the optimal payoff $V^h(t,x)$ given by (16) solves the HJB equation

$$\frac{\partial V^h}{\partial t}(t,x) + \max_u [J(t,u,x) + L_{t,u}^h V^h(t,x)] = 0. \tag{21}$$

Thus, as a corollary of Theorem , we have proved the convergence of the solutions of the Cauchy problem for equation (21) to the viscosity solution of (20).

4. Two Players with Mean-field Interaction

Let us turn to a game-theoretic setting starting with a simplest model of two competing mean-field interacting Markov chains. Suppose we are given two families of Q-matrices $\{Q(t,u,x) = (Q_{ij})(u,x)\}$ and $\{P(t,v,x) = (P_{ij})(v,x)\}$, $i,j = 1,\cdots d$, depending on $x \in \Sigma_d$ and parameters u and v from two subsets U and V of Euclidean spaces. Any given bounded measurable curves $u(t), v(t)$, $t \in [0,T]$, define a Markov chain on $\Sigma_d^{1/|N|} \times \Sigma_d^{1/|M|}$, specified by the generator

$$L_{t,u(t),v(t)} f(\frac{N}{|N|}, \frac{M}{|M|}) = \sum_{i,j}^{d} n_i Q_{ij}(t,u(t),\frac{N}{|N|})[f\left(\frac{N^{ij}}{|N|},\frac{M}{|M|}\right) - f\left(\frac{N}{|N|},\frac{M}{|M|}\right)]$$

$$+\sum_{i,j}^{d} m_i P_{ij}(t,v(t),\frac{M}{|M|})[f\left(\frac{N}{|N|},\frac{M^{ij}}{M}\right) - f\left(\frac{N}{|N|},\frac{M}{|M|}\right)], \tag{22}$$

where $N = (n_1,\cdots,n_d)$, $M = (m_1,\cdots,m_d)$.

We shall assume for simplicity that $|N| = |M| = 1/h$.

Then (22) rewrites as

$$L^h_{t,u(t),v(t)} f(x,y) = \sum_{i=1}^{d}\sum_{j=1}^{d} x_i Q_{ij}(t,u(t),x)\frac{1}{h}[f(x-he_i+he_j,y) - f(x,y)]$$

$$+\sum_{i=1}^{d}\sum_{j=1}^{d} y_i P_{ij}(t,v(t),y)\frac{1}{h}[f(x,y-he_i+he_j) - f(x,y)], \quad x,y \in h\mathbf{Z}^d_+. \tag{23}$$

For $f \in C^1(\Sigma_d \times \Sigma_d)$,

$$\lim_{h\to 0,\, N/|N|\to x, M/|M|\to y} L^h_{t,u(t),v(t)} f(N/|N|, M/|M|) = \Lambda_{t,u(t),v(t)} f(x,y),$$

where

$$\Lambda_{t,u(t),v(t)} f(x,y) = \sum_{k=1}^{d}\sum_{i\neq k}[x_i Q_{ik}(t,u(t),x) - x_k Q_{ki}(t,u(t),x)]\frac{\partial f}{\partial x_k}(x)$$

$$+\sum_{k=1}^{d}\sum_{i\neq k}[y_i P_{ik}(t,u(t),x) - y_k P_{ki}(t,v(t),y)]\frac{\partial f}{\partial y_k}(y). \tag{24}$$

The corresponding controlled characteristics are governed by the equations

$$\dot{x}_k = \sum_{i\neq k}[x_i Q_{ik}(t,u(t),x) - x_k Q_{ki}(t,u(t),x)] = \sum_{i=1}^{d} x_i Q_{ik}(t,u(t),x), \quad k = 1,...,d, \tag{25}$$

$$\dot{y}_k = \sum_{i\neq k}[y_i P_{ik}(t,v(t),y) - y_k P_{ki}(t,v(t),y)] = \sum_{i=1}^{d} y_i P_{ik}(t,v(t),y), \quad k = 1,...,d. \tag{26}$$

For a given $T > 0$, let us denote by $\Gamma(T,h)$ the stochastic game with the dynamics specified by the generator (22) and with the objective of the player I (controlling Q via u) to maximize the payoff

$$\int_0^T J\left(s,u(s),v(s),\frac{N(s,h)}{|N|},\frac{M(s,h)}{|M|}\right) ds + V_T\left(\frac{N(T,h)}{|N|},\frac{M(T,h)}{|M|}\right) \tag{27}$$

for given functions J (current payoff) and V_T (terminal payoff), and with the objective of player II (controlling P via v) to minimize this payoff (zero-sum game). As previously we want to approximate it by the deterministic zero-sum differential game $\Gamma(T)$, defined by dynamics (25), (26) and the payoff of player I given by

$$\int_0^T J(s,u(s),v(s),X_{t,x}(s),Y_{t,y}(s))\,ds + V_T(X_{t,x}(T),Y_{t,y}(T)). \tag{28}$$

Recall the basic notions of the upper and lower values for a game $\Gamma(T)$, see e.g. (Fleming & Soner, 2006) or (Malaeyev, 2000). As above, we shall use controls $u(.)$ and $v(.)$ from the classes $C_{pc}([0,T];U)$ and $C_{pc}([0,T];V)$ of piecewise-continuous curves with values in U and V respectively. A progressive strategy of player I is defined as a mapping β

from $C_{pc}([0,T];V)$ to $C_{pc}([0,T];U)$ such that if $v_1(.)$ and $v_2(.)$ coincide on some initial interval $[0,t]$, $t < T$, then so do $u_1 = \beta(v_1(.))$ and $u_2 = \beta(v_2(.))$. Similarly progressive strategies are defined for player II. Let us denote the sets of progressive strategies for players I and II by $S_p([0,T];U)$ and $S_p([0,T];V)$. Then the upper and the lower values for the game $\Gamma(T)$ are defined as

$$V_+(t,x,y) = \sup_{\beta\in S_p([0,T];U)} \inf_{v(.)\in C_{pc}([0,T];V)}$$

$$\left[\int_t^T J(s,(\beta(v))(s),v(s),X_{t,x}(s),Y_{t,x}(s))\,ds + V_T(X_{t,x}(T),Y_{t,x}(T))\right], \tag{29}$$

$$V_-(t,x,y) = \inf_{\beta\in S_p([0,T];V)} \sup_{u(.)\in C_{pc}([0,T];U)}$$

$$\left[\int_t^T J\left(s,u(s),(\beta(u))(s),X_{t,x}(s),Y_{t,x}(s)\right)\,ds + V_T\left(X_{t,x}(T),Y_{t,x}(T)\right)\right].$$

If the so called Isaac's condition holds, that is, for any p_k, q_k,

$$\max_u \min_v \left[J(t,u,v,x,y) + \sum_{i,k=1}^d x_i Q_{ik}(t,v,x)q_k + \sum_{i,k=1}^d y_i P_{ik}(t,v,x)p_k\right]$$

$$= \min_v \max_u \left[J(t,u,v,x,y) + \sum_{i,k=1}^d x_i Q_{ik}(t,v,x)q_k + \sum_{i,k=1}^d y_i P_{ik}(t,v,x)p_k\right], \tag{30}$$

then the upper and lower values coincide: $V_+(t,x,y) = V_-(t,x,y)$.

Similarly the upper and the lower values $V_+^h(t,x,y)$ and $V_-^h(t,x,y)$ for the stochastic game $\Gamma(T,h)$ are defined.

Theorem 3. *Assume that Q,P,J depend continuously on t,u and $Q,P,J,V_T \in C^{1,\alpha}(\Sigma_d)$, $\alpha \in (0,1]$, as functions of x, with the norms bounded uniformly in t,u,v. Then*

$$\sup_{0\le t\le T} [V_\pm^h(t,hN) - V_\pm(t,x)]$$

$$\le C(T)((T-t)h^\alpha + |hN-x|)\left(\|V_T\|_{C^{1,\alpha}(\Sigma_d)} + \sup_{s,u} \|J(t,u,v,.)\|_{C^{1,\alpha}(\Sigma_d)}\right), \tag{31}$$

with $C(T)$ depending only on the bounds of the norms of Q in $C^{1,\alpha}(\Sigma_d)$. Moreover, if $\beta \in S_p([0,T];U)$ and $v(.) \in C_{pc}([0,T];V)$ are ϵ-optimal for the minimax problem (29)*, then this pair is also $(\epsilon+C(T)h^\alpha)$-optimal for the corresponding stochastic game $\Gamma(T,h)$.*

As in Theorem (ii), one can also approximate optimal (equilibrium) adaptive polices for $\Gamma(T,h)$, if regular enough (i.e. Lipschitz continuous) equilibrium adaptive policies exist for the limiting game $\Gamma(T)$. In fact, as is known from differential games, see e.g. (Fleming & Soner, 2006; Malaeyev, 2000) or (Petrosjan & Zenkevich, 1996), the upper value $V_+(t,x,y)$ represents the unique viscosity solution of the upper Isaac's equation

$$\frac{\partial V_+}{\partial t}(t,x,y) + \min_v \max_u \left[J(t,u,v,x,y) + \Lambda_{t,u,v}V_+(t,x,y)\right], \quad V_+(T,x,y) = V_T(x,y), \tag{32}$$

and $V_-(t,x,y)$ of the lower Isaac's equation (with min and max placed in a different order). Similar equations are satisfied by the values of stochastic games $V_\pm^h(t,x,y)$, see e.g. (Fleming & Souganidis, 1989). Now, if V^* is a solution to the Cauchy problem (32) and there exist Lipschitz continuous functions $v^*(t,x,y)$ and $u^*(t,v,x,y)$ such that

$$u^*(t,v,x,y) \in \; argmax[J(t,u,v,x,y) + \Lambda_{t,u,v}V^*(t,x,y)],$$

$$v^*(t,x,y) \in \; argmin \; \max_v [J(t,u,v,x,y) + \Lambda_{t,u,v}V^*(t,x,y)],$$

then V^* is a saddle point for the differential game $\Gamma^+(T)$ giving the information advantage to maximizing player I, see e.g. (Fleming & Soner, 2006, Theorem 3.1). Analogously to Theorem (ii), we can conclude by Theorem that the policies $v^*(t,x,y)$ and $u^*(t,v,x,y)$ represent ϵ-equilibria for the corresponding stochastic game $\Gamma^+(T,h)$.

5. Two Players with Binary Interaction

In a slightly different setting one can assume that changes in a competitive control process occur as a result of group interactions, and are not determined just by the overall mean field distribution. Let us discuss a simple situation with binary interaction.

As in the previous section, assume we have two groups of d states (of objects or agents) controlled by players I and II respectively. Suppose now that any particle from a state i of the first group can interact with any particle from a state j of the second group (binary interaction) producing changes i to l and j to r with certain rates $Q_{ij}^{lr}(t,u,v)$ that may depend on controls u and v of the players. Assuming, as usual, that our particles are indistinguishable (any particle from a state is selected for interaction with equal probability), leads to the process, generated by the operators

$$L_{t,u(t),v(t)}f(N,M) = \sum_{i,j,l,r=1}^{d} n_i m_j Q_{ij}^{lr}(t,u(t),v(t),\frac{N}{|N|},\frac{M}{|M|})[f(N^{il},M^{jr}) - f(N,M)].$$

Again let us assume for simplicity that $|M| = |N|$ and define $h = 1/|N| = 1/|M|$. To get a reasonable scaling limit, it is necessary to scale time by factor h leading to the generators

$$L_{t,u(t),v(t)}^{h}f(\frac{N}{|N|},\frac{M}{|M|}) = h\sum_{i,j,l,r=1}^{d} n_i m_j Q_{ij}^{lr}(t,u(t),v(t),\frac{N}{|N|},\frac{M}{|M|})[f(N^{il},M^{jr}) - f(N,M)], \tag{33}$$

which, for $x = hN$, $y = hM$ and $h \to 0$, tends to

$$\Lambda_{t,u(t),v(t)}f(x,y) = \sum_{i,j,l,r=1}^{d} x_i y_j Q_{ij}^{lr}(t,u(t),v(t),x,y)\left[\frac{\partial f}{\partial x_l} + \frac{\partial f}{\partial y_r} - \frac{\partial f}{\partial x_i} - \frac{\partial f}{\partial y_j}\right](x,y). \tag{34}$$

The corresponding kinetic equations (characteristics of this first order partial differential operator) have the form

$$\dot{x}_k = \sum_{i,j,r=1}^{d} y_j\left[x_i Q_{ij}^{kr}(t,u(t),v(t)) - x_k Q_{kj}^{ir}(t,u(t),v(t))\right],$$

$$\dot{y}_k = \sum_{i,j,l=1}^{d} x_i\left[y_j Q_{ij}^{lk}(t,u(t),v(t)) - y_k Q_{ik}^{lj}(t,u(t),v(t))\right],$$

As in the previous section, we are interested in the zero-sum stochastic game, which will again be denoted by $\Gamma(T,h)$, with the dynamics specified by generator (33) and with the objective of the player I (controlling Q via u) to maximize the payoff of the same type (27), and in an approximation of this game by the limiting deterministic zero-sum differential game $\Gamma(T)$, defined by the payoff (28) of player I.

Theorem 4. *Assume that Q, J depend continuously on t,u,v and $Q,J,V_T \in C^{1,\alpha}(\Sigma_d)$, $\alpha \in (0,1]$, as functions of x, with the norms bounded uniformly in t,u,v. Then the same estimate* (31) *holds for the difference of upper and lower values of limiting and approximating games.*

Moreover, literally the same approximations (as in the game of the previous section) hold for optimal strategies and policies.

6. Several Players with Coupled Mean-field Interaction

Suppose now that there are K players, each one controlling a mean-field interacting Markov chain, which are coupled via the joint mean-field distribution. Namely, suppose we are given K families of Q-matrices $\{Q^k = (Q_{ij}^k)(t,u_k,\mu_1,\cdots,\mu_K)\}$, $k = 1,\cdots K$, $i,j = 1,\cdots d$, depending on a parameter u_k from a metric space U_k (control space of the player k) and K vectors $\mu_k \in \Sigma_d$. Any bounded measurable vector-valued curve $u(t) = (u_1(t),\cdots u_K(t))$, $t \in [0,T]$, and natural numbers $|N_1|,\cdots,|N_k|$ define a Markov chain on $S_{1/|N_1|} \times \cdots \times S_{1/|N_k|}$ (each S_k consists of sequences of d negative integers $N_k = (n_1^k,\cdots,n_d^k)$ totting up to a given number $|N_k|$), specified by the generator

$$L_{t,u(t)}f(\frac{N_1}{|N_1|},\cdots,\frac{N_K}{|N_K|}) = \sum_{k=1}^{K}\sum_{i_k,j_k=1}^{d} n_{i_k}^k Q_{i_k j_k}^k\left(t,u_k(t),\frac{N_1}{|N_1|},\cdots,\frac{N_K}{|N_K|}\right)$$

$$\times\left[f(\frac{N_1}{|N_1|},\cdots,\frac{N_{k-1}}{|N_{k-1}|},\frac{N_k^{i_k j_k}}{|N_k|},\frac{N_{k+1}}{|N_{k+1}|},\cdots,\frac{N_K}{|N_K|}) - f(\frac{N_1}{|N_1|},\cdots,\frac{N_K}{|N_K|})\right], \tag{35}$$

where $N_k = (n_1^k, \cdots, n_d^k)$ with elements totting up to $|N_k|$. For

$$x_k = \frac{N_k}{|N_k|} = \left(\frac{n_1^k}{|N_1|}, \cdots, \frac{n_d^k}{|N_k|}\right) \in h\mathbf{Z}_+^d, \quad k = 1, \cdots K,$$

this rewrites as

$$L^h_{t,u(t)} f(x_1, \cdots, x_k) = \frac{1}{h} \sum_{k=1}^{K} \sum_{i_k, j_k=1}^{d} x_k^{i_k} Q^k_{i_k j_k}(t, u_k(t), x_1, \cdots, x_K)$$

$$\times[f(x_1, \cdots, x_{k-1}, x_k - he_{i_k} + he_{j_k}, x_{k+1}, \cdots x_K) - f(x_1, \cdots, x_K)], \tag{36}$$

where we denote by h the maximum of all $|N_k|^{-1}$.

For $f \in C^1(\Sigma_d \times \cdot \times \Sigma_d)$,

$$\lim_{h\to 0,\, N_k/|N_k| \to x_k, k=1,\cdots K} L^h_{t,u(t)} f(\frac{N_1}{|N_1|}, \cdots, \frac{N_K}{|N_K|}) = \Lambda_{t,u(t)} f(x_1, \cdots, x_K),$$

where

$$\Lambda_{t,u(t)} f(x) = \sum_{k=1}^{K} \sum_{j,i=1}^{d} x_i Q_{ij}(t, u_k(t), x_1, \cdots, x_K) \frac{\partial f}{\partial x_j^k}(x_1, \cdots, x_K). \tag{37}$$

The corresponding controlled characteristics are governed by the equations

$$\frac{d}{dt} x_k^j = \sum_{i=1}^{d} x_i Q_{ij}(t, u_k(t), x_1, \cdots, x_K), \quad j = 1, ..., d, \quad k = 1, \cdots, K. \tag{38}$$

For a given $T > 0$, suppose the objective of player k is to maximize the payoff

$$V_k^h(t, x_1, \cdots, x_k, u(.))$$

$$= \mathbf{E}^{u(.)}_{t,x_1,\cdots,x_k} \int_0^T J_k\left(s, u_k(s), \frac{N_1(s,h)}{|N_1|}, \cdots, \frac{N_K(s,h)}{|N_K|}\right) ds + V_T^k\left(\frac{N_1(T,h)}{|N_1|}, \cdots, \frac{N_K(T,h)}{|N_K|}\right)$$

with given functions J_k (current payoffs) and V_T^k (terminal payoffs). Let us denote by $\Gamma_K(T, h)$ the corresponding stochastic game. In the limit $h \to 0$ with all ratios N_k/N_j uniformly bounded we get the deterministic differential game $\Gamma_K(T)$ of K players with separated dynamics (38) and the payoffs

$$V(t, x_1, \cdots, x_k, u(.)) = \int_0^T J_k(s, u_k(s), x_1(s), \cdots, x_K(s))\, ds + V_T^k(x_1(T), \cdots, x_K(T)).$$

K-player differential games are much less understood as two-player zero-sum games, see e.g. (Friedman, 1971; Malafeyev, 2000; Olsder, 2001; Ramasubraanian, 2007; Tolwinski, Haurie & Leitmann, 1986) for informative discussions, including links with viscosity solutions of the systems of HJB equations. It is not our objective here to contribute to this development. We just want to stress that, as our method shows, most of the natural equilibria of various kinds that can be analyzed for the limiting differential game $\Gamma_K(T)$ do approximate the corresponding equilibria for games $\Gamma_K(T, h)$. As simplest examples, let us consider open-loop equilibria and K-player analogs of upper and lower values of zero-sum games. A vector curve $u^*(t) = (u_1^*(t), \cdots, u_K^*(t))$ is called an open-loop Nash equilibrium for the game $\Gamma_K(T)$, if for any $k = 1, \cdots, K$,

$$V(t, x_1, \cdots, x_k, u^*(.)) = \max_{u_k(.)} V(t, x_1, \cdots, x_k, u_1^*(.), \cdots u_{k-1}^*(.), u_k(.), u_{k+1}^*(.), \cdots u_K(.))).$$

Similarly open-loop Nash equilibria are defined for the games $\Gamma_K(T, h)$.

On the other hand, for any permutation π of K players one can define a vector-value V_π arising from information discrimination specified by π. That is, the set of strategies of player $\pi(1)$ is just $C_{pc}([0, T]; U_{\pi(1)})$, and the progressive strategies for a player $\pi(k) \neq \pi(1)$ are defined as mapping

$$\beta_k : \times_{i=1}^{k-1} C_{pc}([0, T]; U_{\pi_i}) \to C_{pc}([0, T]; U_{\pi_k})$$

such that if $u^1 = (u_{\pi(1)}(.), \cdots, u_{\pi(k-1)}(.))$ and $u^2 = (u_{\pi(1)}(.), \cdots, u_{\pi(k-1)}(.))$ coincide on some initial interval $[0, t]$, $t < T$, then so do $\beta_k(u^1)$ and $\beta_k(u^2)$. Let us denote the sets of progressive strategies for players $\pi(k)$ by $S_p([0, T]; U_{\pi(k)})$.

The discriminated vector-value $V_\pi(t, x_1, \cdots, x_K) = (V_\pi^1, \cdots, V_\pi^K)(t, x_1, \cdots, x_K)$ of the game $\Gamma_K^\pi(T)$ is defined as a Nash-equilibrium payoff of the game (in normal form) with these strategy spaces.

Theorem 5. *Assume that Q_k, J_k depend continuously on all parameters and $Q_k, J_k, V_T^K \in C^{1,\alpha}(\Sigma_d)$, $\alpha \in (0, 1]$, as functions of $x_1, \cdots, x_k$, with the norms bounded uniformly in t, u, v. Then the same estimate* (31) *holds for the difference of payoffs in open-loop Nash equilibria of limiting and approximating games, as well as for the difference of payoffs in Nash equilibria of the limiting and approximating discriminatory games specified by any permutation π.*

There does not seem to exits any general results on the regularity of adaptive policies for N player game. On the other hand, in many examples, see (Case, 1967), and in some sense in a general position, see (Malafeyev, 2000), the state space can be decomposed into a finite number of open sets, where equilibrium adaptive policies are smooth, with these sets being separated by lower dimensional manifolds, where switching occurs. Under this condition, one could possibly prove the convergence of adaptive policies for large $|N|$-approximations to adaptive policies of the limiting deterministic differential games, but such analysis is beyond the present contribution.

7. Auxiliary Lemmas

We shall need two simple lemmas (possibly well known for experts), one on the bounds for semigroups arising from deterministic processes, and another on the coupling of jump-type Markov processes.

Let us recall that the modulus of continuity for a continuous function ϕ on Σ_d is defined as follows:

$$w_h(\phi) = \sup\{|\phi(x_1) - \phi(x_2)| : |x_1 - x_2| \le h\}, \quad h > 0. \tag{39}$$

Below we denote by ∇ the derivative operator with respect to variable $x \in \Sigma_d$.

Lemma 1. *(i) Let $Q(t, x) \in C^1(\Sigma_d)$ as a function of x uniformly in t and depends measurably on t. Let $\Phi_{s,t}$ denote the linear operators* (7)*, where $X_{s,x}(t)$ is the solution to* (6) *with the initial condition x at time s. Then $\Phi_{s,t}$ preserve the space $C^1(\Sigma_d)$ and*

$$\|\Phi_{s,t}\|_{C^1(\Sigma_d)} \le e^{\omega(t-s)}, \tag{40}$$

or more explicitly

$$\|\Phi_{s,t} f\|_{C^1(\Sigma_d)} \le e^{\omega(t-s)} \|f\|_{C^1(\Sigma_d)},$$

where ω is the supremum over t of the norms of the functions $xQ(t, x)$ (as functions of x) in $C^1(\Sigma_d)$. Moreover, let $\tilde{w}_h$ denote the supremum over t of the modulus of continuity of the functions $\nabla(xQ(t, x))$ (as functions of x). Then

$$w_h(\nabla(\Phi_{s,t} f)) \le e^{\omega(t-s)} \left[w_{he^{\omega(t-s)}}(\nabla f) + \tilde{w}_{he^{\omega(t-s)}} \|f\|_{C^1(\Sigma_d)} \right]. \tag{41}$$

(ii) Supposes additionally that $Q(t, .) \in C^{1,\alpha}(\Sigma_d)$, $\alpha \in (0, 1]$, uniformly in t, and let ω_α denote any uniform upper bound for the corresponding Hölder constants. Then the space $C^{1,\alpha}(\Sigma_d)$ is preserved by $\Phi_{s,t}$ and

$$\|\Phi_{s,t}\|_{C^{1,\alpha}(\Sigma_d)} \le [1 + (t - s)\omega_\alpha] e^{\omega(t-s)}. \tag{42}$$

Proof. (i) The derivative of $X_{s,x}(t)$ with respect to x solves the linear ODE

$$\frac{d}{dt} \frac{\partial X_{s,x}(t)}{\partial x} = \frac{\partial}{\partial y} (yQ(t, y)|_{y=X_{s,x}(t)} \frac{\partial X_{s,x}(t)}{\partial x}$$

implying

$$\|\frac{\partial X_{s,x}(t)}{\partial x}\| \le e^{\omega(t-s)}, \tag{43}$$

and hence (40). The proof of (41) is the same as of (42) below, and therefore is omitted.

(ii) The function

$$\phi_{s,t}(x_1, x_2) = \left.\frac{\partial X_{s,x}(t)}{\partial x}\right|_{x=x_1} - \left.\frac{\partial X_{s,x}(t)}{\partial x}\right|_{x=x_2}$$

satisfies the linear ODE

$$\frac{d}{dt} \phi_{s,t}(x_1, x_2) = \frac{\partial}{\partial y} (yQ(t, y)|_{y=X_{s,x_1}(t)} \left.\frac{\partial X_{s,x}(t)}{\partial x}\right|_{x=x_1} - \frac{\partial}{\partial y} (yQ(t, y)|_{y=X_{s,x_2}(t)} \left.\frac{\partial X_{s,x}(t)}{\partial x}\right|_{x=x_2}.$$

The r.h.s. of this equation rewrites as

$$\left(\frac{\partial}{\partial y} (yQ(t, y)|_{y=X_{s,x_1}(t)} - \frac{\partial}{\partial y} (yQ(t, y)|_{y=X_{s,x_2}(t)} \right) \left.\frac{\partial X_{s,x}(t)}{\partial x}\right|_{x=x_1}$$

$$+\frac{\partial}{\partial y}(yQ(t,y)|_{y=X_{s,x_2}(t)}\left(\frac{\partial X_{s,x}(t)}{\partial x}\bigg|_{x=x_1}-\frac{\partial X_{s,x}(t)}{\partial x}\bigg|_{x=x_2}\right)$$

and hence, by (40), is bounded in magnitude by the expression

$$\omega_\alpha e^{\omega(t-s)}|x_1-x_2|^\alpha+\omega|\phi_{s,t}(x_1,x_2)|.$$

By Gronwall's lemma,

$$|\phi_{s,t}(x_1,x_2)|\le(t-s)\omega_\alpha e^{\omega(t-s)}|x_1-x_2|^\alpha.$$

Hence the Hölder constant (of index α) for the function $\partial X_{s,x}(t)/\partial x$, as a function of x, is bounded by $\omega_\alpha e^{\omega(t-s)}$ implying (42), because

$$\frac{\partial}{\partial x}f(X_{s,x}(t))|_{x=x_1}-\frac{\partial}{\partial x}f(X_{s,x}(t))|_{x=x_2}$$
$$=\left[\frac{\partial f}{\partial x}(X_{s,x_1}(t))-\frac{\partial f}{\partial x}(X_{s,x_2}(t))\right]\frac{\partial X_{s,x}(t)}{\partial x}\bigg|_{x=x_1}$$
$$+\frac{\partial f}{\partial x}(X_{s,x_2}(t))\left(\frac{\partial X_{s,x}(t)}{\partial x}\bigg|_{x=x_1}-\frac{\partial X_{s,x}(t)}{\partial x}\bigg|_{x=x_2}\right).$$

The next lemma will be used only for the proof of the second part of Theorem 2. It is needed to compare the effects of applying different adaptive policies.

Lemma 2. *Let $Z_z(t)$ and $W_w(t)$ be two jump-type Markov processes in $\mathbf{R}^d$ (z and w stand for the initial points) with integral generators*

$$L^Zf(x)=\int(f(x+y)-f(x))\nu(x,y)M(dy),\quad L^Wf(x)=\int(f(x+y)-f(x))\mu(x,y)M(dy),$$

with a certain probability measure M on $\mathbf{R}^d$ and uniformly bounded non-negative functions $\nu(x,y)$, $\mu(x,y)$ such that $\int|y|\nu(x,y)M(dy)$ and $\int|y|\mu(x,y)M(dy)$ are bounded for any x and

$$\int|y||\nu(x_1,y)-\nu(x_2,y)|M(dy)\le\varkappa|x_1-x_2|,$$

$$\int|y||\nu(x,y)-\mu(x,y)|M(dy)\le\epsilon$$

for a $\varkappa>0$ and an $\epsilon>0$.

Then there exists a Markov process $X_{z,w}(t)$ on $\mathbf{R}^d\times\mathbf{R}^d$ that couples $Z_z(t)$ and $W_w(t)$ in the sense that the distribution of the first (respectively second) coordinate of $X_{z,w}(t)$ coincides with the distribution of $Z_z(t)$ (respectively $W_w(t)$), and such that, with respect to this coupling,

$$\mathbf{E}|Z_z(t)-W_w(t)|\le e^{\varkappa t}(|z-w|+t\epsilon). \tag{44}$$

Proof. Let $X_{z,w}(t)$ be specified by the so called marching coupling, that is by the generator

$$Lf(x_1,x_2)=\int[f(x_1+y,x_2+y)-f(x_1,x_2)]m(x_1,x_2,y)M(dy)$$
$$+\int[f(x_1+y,x_2)-f(x_1,x_2)][\nu(x_1,y)-m(x_1,x_2,y)]M(dy)$$
$$+\int[f(x_1,x_2+y)-f(x_1,x_2)][\mu(x_2,y)-m(x_1,x_2,y)]M(dy).$$

where

$$m(x_1,x_2,y)=\min(\nu(x_1,y),\mu(x_2,y)).$$

Clearly, if f does not depend on the second argument, then $Lf(x_1,x_2)=L^Zf(x_1)$, and if f does not depend on the first argument, then $Lf(x_1,x_2)=L^Wf(x_2)$, so that $X_{z,w}(t)$ is really a coupling of $Z_z(t)$ and $Z_w(t)$. Moreover, as one sees by inspection, for $f(x_1,x_2)=|x_1-x_2|$,

$$Lf(x_1,x_2)\le\int|y|[\max(\nu(x_1,y),\mu(x_2,y))-m(x_1,x_2,y)]M(dy).$$

Consequently,

$$Lf(x_1,x_2) \le \int |y||\nu(x_1,y) - \mu(x_2,y)|M(dy) \le \varkappa|x_1 - x_2| + \epsilon.$$

By Dynkin's formula, the process

$$f(X_{x_1,x_2}(t)) - \int_0^t Lf(X_{x_1,x_2}(s))ds$$

is a martingale. Hence

$$\mathbf{E}|X^1_{x_1,x_2}(t) - X^2_{x_1,x_2}(t)|$$
$$\le |x_1 - x_2| + \varkappa \int_0^t \mathbf{E}|X^1_{x_1,x_2}(s) - X^2_{x_1,x_2}(s)|ds + t\epsilon$$

(where $X = (X^1, X^2)$), which by Gronwall's lemma implies

$$\mathbf{E}|X^1_{x_1,x_2}(t) - X^2_{x_1,x_2}(t)| \le e^{\varkappa t}(|x_1 - x_2| + t\epsilon),$$

that is (44).

8. Proof of the Theorems

Proof of Theorem 1.

Let us start by noting that the transition operators $\Psi^h_{s,t}$ of the Markov chains $hN^{t,h}$ satisfy (outside the finite number of discontinuity points of $Q(t,x)$) the standard Kolmogorov equations

$$\frac{d}{dt}\Psi^h_{s,t}f = \Psi^h_{s,t}L_t f, \quad \frac{d}{ds}\Psi^h_{s,t}f = -L_s\Psi^h_{s,t}f, \quad 0 \le s \le t, \tag{45}$$

for any function f on Σ^h_d. Similarly, the transition operators $\Phi_{s,t}$ of the deterministic Markov process $X_{s,x}(t)$ satisfy the equations

$$\frac{d}{dt}\Phi_{s,t}f = \Phi^h_{s,t}\Lambda_t f, \quad \frac{d}{ds}\Phi_{s,t}f = -\Lambda_s\Phi_{s,t}f, \quad 0 \le s \le t \tag{46}$$

(again outside the discontinuity points of $Q(t,x)$), for any $f \in C^1(\Sigma_d)$, which is of course easily checked.

To compare these propagators, we shall use the following standard trick. We write

$$\Psi^h_{s,t}f - \Phi_{s,t}f = \Psi^h_{s,r}\Phi_{r,t}|^t_{r=s}f = \int_s^t \frac{d}{dr}(\Psi^h_{s,r}\Phi_{r,t})f\,dr = \int_s^t \Psi^h_{s,r}(L^h_r - \Lambda_r)\Phi_{r,t}f\,dr,$$

if only $\Phi_{r,t}f \in C^1(\Sigma_d)$ for any r. Let us apply this equation to an $f \in C^{1,\alpha}(\Sigma_d)$. By Lemma 1 (ii), $\Phi_{r,t}f \in C^{1,\alpha}(\Sigma_d)$ for all r with a uniform bound, so that the above equation applies. Moreover, from (3) and (5) we conclude that

$$\|(L^h_t - \Lambda_t)f\| \le C(T)h^\alpha\|f\|_{C^{1,\alpha}(\Sigma_d)}, \tag{47}$$

with a constant $C(T)$ depending on the sup-norm of $Q(t,x)$ only. Consequently, by (42),

$$\|\Psi^h_{s,t}f - \Phi_{s,t}f\| \le (t-s)C(T)h^\alpha[1 + (t-s)\omega_\alpha]e^{\omega(t-s)}\|f\|_{C^{1,\alpha}(\Sigma_d)},$$

implying (8). The estimate (9) is a straightforward corollary, taking into account (40). The convergence of the chains in distribution follows from the convergence of transition operators by a well known general result, see e.g. (Kallenberg, 2002), or (Kolokoltsov, 2011). Moreover, weak convergence of random variables to a constant implies their convergence in probability.

Finally, (10) is proved analogously, if instead of (47), one uses the estimate

$$\|(L^h_t - \Lambda_t)f\| \le C(T)w_h(\nabla f) \tag{48}$$

together with (41).

Proof of Theorem 2.

(i) By Theorem 1 (and taking into account that the integral of the function J can be approximated by Riemannian sums), for any $u(.) \in C_{pc}(t,T)$,

$$V^h(t,x,u(.)) = \mathbf{E}^{u(.)}_{t,x}\left[\int_t^T (J(s,u(s),hN(s,h))ds + V_T(hN(T,h))\right]$$

differs from

$$V(t, x, u(.)) = \int_t^T J(s, u(s), X_{t,x}(s))\, ds + V_T(X_{t,x}(T))$$

by

$$C(T)((T-t)h^\alpha + |hN - x|)\left(\|V_T\|_{C^{1,\alpha}(\Sigma_d)} + \sup_{s,u} \|J(t, u, .)\|_{C^{1,\alpha}(\Sigma_d)}\right).$$

Hence the same estimate holds for the difference of the suprema of these two functions of $u(.)$.

(ii) First of all approximating $\tilde{J}$ by a smooth function (and noting that all payoffs will be then uniformly approximated) reduces the problem to the case of a smooth J. Next, we can approximate $\gamma(t, x)$ arbitrary close by a smooth function $\tilde{\gamma}(t, x)$ in x, so that the corresponding $V(t, x, \gamma)$ and $V(t, x, \tilde{\gamma})$ differ by arbitrary small amounts. By Theorem 1, $V^h(t, x, \tilde{\gamma})$ will converge to $V(t, x, \tilde{\gamma})$ as $h \to 0$, and hence $\tilde{\gamma}$ becomes an $(\epsilon + \delta)$-optimal policy of $\Gamma(T, h)$ for small enough h. It remains to compare $V^h(t, x, \tilde{\gamma})$ and $V^h(t, x, \gamma)$. But they are close by Lemma 2.

Proof of Theorems 3, 4, 5. It is the same as the proof of Theorem 2, being based on Theorem 1 and an evident observation that if two families of functions are uniformly close, then so are also their minimax values.

9. Conclusion and Bibliographical Comments

The work on deterministic LLN limits for interacting particles and its representation in terms of nonlinear kinetic equations goes back to (Leontovich, 1935) and (Bogolyubov, 1946). A deduction of kinetic equations in a very general setting of kth order (binary, ternary, etc) interaction can be found in (Maslov & Tariverdiev, 1982) or (Belavkin & Kolokoltsov, 2003). In particular, for the corresponding limit in a game-theoretic setting (replicator dynamics), we can refer to (Benaim & Weibull, 2003) or the last section of (Kolokoltsov & Malafeyev, 2010).

For the introduction to nonlinear Markov processes we refer to (Kolokoltsov, 2007, 2010) and references therein, or, for more physical point of view, to (Frank, 2008) or (Zak, 2002).

Mean-field control is a rapidly expanding area, see e.g. (Andersson & Djehiche, 2003; Buckdahn, B. Djehiche, J. Li, & S. Peng, 2009) and references therein, for diffusion based models, and (Le Boudec, McDonald, & Mundinger, 2007; Gast & B. Gaujal, 2009; Gast, Gaujal, & Le Boudec, 2010; Bordenave, McDonald, & Proutiere, 2007; Benaïm & Le Boudec, 2008; Milutinovic & Lima, 2006) for discrete models, more engineering application oriented. In these papers one can find various concrete applications (from robot swamps to transportation theory and networks), which are also relevant to the mathematical models discussed in the present paper.

The closest to our setting seems to be the recent work (Gast, Gaujal, & Le Boudec, 2010), which is devoted to a convergence result similar to our Theorem 2. However, in (Gast, Gaujal, & Le Boudec, 2010) a continuous time mean-field control model is obtained as a limit of discrete-time models, and we work directly with continuous time models. More essentially, under a slightly stronger regularity assumptions on the model (basically continuous differentiability of coefficients instead of Lipschitz continuity) we obtain explicit rate of convergence for the averages over empirical measures, and in some cases even for adaptive control policies. Moreover, our main objective was to develop a general framework to treat competitive control problems presented in Sections 4-6. Even further, we developed a method to treat not only the simplest mean-field type interactions, but more involved binary or ternary ones.

As was mentioned, our initial state space was finite, resulting in the corresponding measure-valued limit being a finite-dimensional differential game. In more general setting (which will be discussed elsewhere), for an arbitrary initial state space of a single particle (in our large ensemble), the corresponding limit becomes truly measure-valued controlled nonlinear evolution (controlled nonlinear Markov process) specified by kinetic equations of rather general type. Other possible extensions can include models with variable number of particles (robots becoming out of order, or military units destroyed, etc).

A slightly different class of models describe the situations when each agent in a large group pursue his/her own interests. These class of models is often referred to as mean-field games. Related equilibrium concept was called the Nash certainty equivalence principle, see (Lasry & Lions, 2006; Achdou & Capuzzo-Dolcetta, 2010; Gomes, Mohr, & Souza, 2010; Huang, Caines, & Malhamé, 2003; Huang, Malhamé, & Caines, 2006; Huang, 2010; Kolokoltsov, Li, & Yang, 2011) and references therein. Unlike our centralized setting above, analysis of these models does not lead to the control of distributions in the limit, but rather to certain consistency condition on homogeneous individual controls (Nash certainty equivalence). However, recent paper (Huang, Malhamé, & Caines, 2007) aims at linking individual and centralized controls.

The main objective of this paper was to approximate discrete stochastic systems with large number of particles by a simpler continuous state limit. One can also look at these results from an opposite point of view: as approximating differential games by discrete Markov chains. This point of view would relate our results with Kushner (2002), Kushner and Dupuis (2001).

References

Andersson, D., & Djehiche, B. (2011). A maximum principle for SDEs of mean-field type. *Appl. Math. Optim., 63,* 341-356. http://dx.doi.org/10.1007/s00245-010-9123-8

Belavkin, V. P., & Kolokoltsov, V. N. (2003). On general kinetic equation for many particle systems with interaction, fragmentation and coagulation. *Proc. Royal Soc. Lond. A, 459*, 727-748. http://dx.doi.org/10.1098/rspa.2002.1026

Bogolyubov, N. N. (1946). Problems of the dynamic theory in statistical physics. Moscow, Nauka (in Russian).

Bordenave, C., McDonald, D., & Proutiere A. (2007). A particle system in interaction with a rapidly varying environment: Mean field limits and applications.

Buckdahn, R., Djehiche, B., Li, J., & Peng, S. (2009). Mean-field backward stochastic differential equations: a limit approach. *Ann. Prob, 37,* 1524-1565. http://dx.doi.org/10.1214/08-AOP442

Case, J. H. (1967). Toward a Theory of Many Player Differential Games. *SIAM Journal on Control, 7,* 179-197. http://dx.doi.org/10.1137/0307013

Darling, R. W. R., & Norris, J. R. (2008). Differential equation approximations for Markov chains. *Probab. Surv., 5*, 3779.

Fleming, W. H., & Soner, H. M. (2006). Controlled Markov Processes and Viscosity Solutions. Sec. Edition, NY: Springer.

Fleming, W. H., & Souganidis, P. E. (1989). On the existence of value functions of two-player, zero-sum stochastic differential games. *Indiana Univ. Math. J., 38,* 293-314. http://dx.doi.org/10.1512/iumj.1989.38.38015

Frank, T. D. (2008). Nonlinear Markov processes. *Phys. Lett. A, 372,* 4553-4555. http://dx.doi.org/10.1016/j.physleta.2008.04.027

Friedman, A. (1971). *Differential Games*. New York: Wiley-Interscience.

Gast, N., & Gaujal, B. (2009). A Mean Field Approach for Optimization in Particle Systems and Applications. Proc. 4th Intern. ICST Conf. on Performance Evaluation Methodologies and Tools.

Gast, N., Gaujal, B., & Le Boudec, J.-Y. (2010). Mean field for Markov Decision Processes: from Discrete to Continuous Optimization. INRIA report No. 7239.

Gomes, D. A., Mohr, J., & Souza, R. R. (2010). Discrete time, finite state space mean field games. *J. Math. Pures Appl., 9,* 308-328. http://dx.doi.org/10.1016/j.matpur.2009.10.010

Huang, M., Caines, E., & Malhamé, R. P. (2003). Individual and mass behaviour in large population stochastic wireless power control problems: centralized and Nash equilibrium solutions. Proc. the 42nd IEEE Conference on Decision and Control, Maui, Hawaii, 98-103.

Huang, M., Malhamé, R. P., & Caines, P. E. (2006). Large population stochastic dynamic games: closed-loop Mckean-Vlasov systems and the Nash certainty equivalence principle. *Communications in information and systems, 6,* 221-252.

Huang, M. (2010). Large-population LQG games involving a major player: the Nash certainty equivalence principle. *SIAM J. Control Optim., 48,* 3318-3353. http://dx.doi.org/10.1137/080735370

Huang, M, Caines, P. E. & Malhamé, R. P. (2007). Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized ϵ?-Nash equilibria. *IEEE Trans. Automat. Control, 52,* 1560-1571. http://dx.doi.org/10.1109/TAC.2007.904450

Kallenberg, O. (2002). Foundations of Modern Probability. 2ed., Springer.

Kolokoltsov, V. N. (2007). Nonlinear Markov Semigroups and Interacting Lévy Type Processes. *Journ. Stat. Physics, 126,* 585-642. http://dx.doi.org/10.1007/s10955-006-9211-y

Kolokoltsov, V. N. (2009). Nonlinear Markov games. Talk on 'Adverserial and Stochastic Elements in Autonomous Systems' Workshop, Washington DC, http://maeresearch.ucsd.edu/mceneaney/ASEAS/Kolokoltsov.pdf

Kolokoltsov, V. N. (2010). *Nonlinear Markov processes and kinetic equations.* Cambridge University Press.

Kolokoltsov, V. N. (2011). Markov processes, semigroups and generators. De Gruyter, Studies in Mathematics v. 38.

Kolokoltsov, V. N., Li, J., & Yang, W. (2011) Mean Field Games and Nonlinear Markov Processes.

Kolokoltsov, V. N., & Malafeyev, O. A. (2010). Understanding Game Theory. World Scientific, Singapore.

Kushner, H. J. (2002). Numerical approximations for stochastic differential games. *SIAM Journal of Control and Optimization, 41,* 457-486. http://dx.doi.org/10.1137/S0363012901389457

Kushner, H. J., & Dupuis, P. (2001). Numerical methods for stochastic control problems in continuous time. 2ed., New York: Springer.

Lasry, J.-M., & Lions, L. Jeux Champ Moyen. *C. R. Math. Acad. Sci. Paris, 343,* 619-625 and 679-684.

Le Boudec, J.-Y., McDonald, D., & Mundinger, J. (2007). A Generic Mean Field Convergence Result for Systems of Interacting Objects. *QEST 2007 (4th INternationalConference on Quantitative Evaluation of SysTems)*, 3-18.

Leontovich, M. A. (1935). Main equations of the kinetic theory from the point of view of random processes. *Journal of Experimantal and Theoretical Physics , 5,* 211-231.

Malafeyev, O. A. (2000). Controlled conflict systems (In Russian). St. Petersburg State University Publication, ISBN 5-288-01769-7.

Maslov, V. P., & Tariverdiev, C. E. (1982). Asymptotics of the Kolmogorov-Feller equation for systems with the large number of particles. *Itogi Nauki i Techniki. Teoriya veroyatnosti, v.19, VINITI*, Moscow (in Russian), 85-125.

Benaïm, M., & Le Boudec, J.-Y. (2008). A class of mean field interaction models for computer and communication systems. *Performance Evaluation, 65,* 823-838. http://dx.doi.org/10.1016/j.peva.2008.03.005

Benaïm, M., & Weibull, J. (2003). Deterministoc approximation of stochastic evolution in games. *Econometrica, 71,* 873-903. http://dx.doi.org/10.1111/1468-0262.00429

McEneaney, W. (2006). *Max-plus methods for nonlinear control and estimation. Systems and Control: Foundations and Applications*. Boston, MA: Birkhäuser Boston, Inc.

Milutinovic, D., & Lima, P. (2006). Modeling and Optimal Centralized Control of a Large-Size Robotic Population. *Robotics, IEEE Transactions on Robotics, 22,* 1280-1285. http://dx.doi.org/10.1109/TRO.2006.882941

Olsder, G. J. (2001). On open- and closed-loop bang-bang control in nonzero-sum differential games. *SIAM J. Control Optim., 40,* 1087-1106. http://dx.doi.org/10.1137/S0363012900373252

Petrosjan, L. A., & Zenkevich, N. A. (1996). Game Theory, World Scientific, Singapore.

Ramasubraanian, S. (2007). A d-person Differential Game with State Space Constraints. *Appl. Math. Optim., 56,* 312-342. http://dx.doi.org/10.1007/s00245-007-9011-z

Song, Q. S. (2008). Convergence of Markov chain approximation on generalized HJB equation and its applications. *Automatica J. IFAC, 44,* 761-766. http://dx.doi.org/10.1016/j.automatica.2007.07.014

Tolwinski, B., Haurie, A., & Leitmann, G. (1986). Cooperative Equilibra in Differential Games. *J. Math. Anal. Appl., 119,* 182-202. http://dx.doi.org/10.1016/0022-247X(86)90152-6

Zak, M. (2000). Quantum Evolution as a Nonlinear Markov Process. *Foundations of Physics Letters, 15,* 229-243. http://dx.doi.org/10.1023/A:1021079403550

Nonparametric Tests of Trend for Proportions

Mayer Alvo

Department of Mathematics and Statistics, University of Ottawa

585 King Edward, Ottawa K1N 6N5, Canada

Tel: 1-613-562-5800 ext. 3534 E-mail: malvo@uottawa.ca

Marie-Pascal Berthelot

Department of Mathematics and Statistics, University of Ottawa

585 King Edward, Ottawa K1N 6N5, Canada

E-mail: mbert074@uottawa.ca

Received: February 16, 2012 Accepted: February 29, 2012 Published: May 1, 2012

doi:10.5539/ijsp.v1n1p92 URL: http://dx.doi.org/10.5539/ijsp.v1n1p92

This research was supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada.

Abstract

A general method is proposed for constructing nonparametric tests of trend for proportions. Such alternatives arise in situations where it is of interest to test for monotonicity in rates of growth. The class of tests is based on the ranks of the observations. The general approach consists of defining two sets of rankings: the first describes the time and the other the binary data itself. The test statistics measures the similarity between the two sets. The asymptotic null distributions are determined for similarity measures due to Spearman, Kendall and Hamming. A limited simulation study shows that the Spearman test has greater power.

Keywords: Tests of trend for proportions, Ranks, Spearman, Kendall, Hamming similarity measure, Asymptotic Distributions

1. Introduction

There are several instances in practice when one is interested in testing for a trend in proportions. For instance, one may be interested in the trend in birth rates, in mortality rates or in the incidence of a certain disease. As an example, we will consider the mortality statistics in South Africa during the period 2000-2008 in Table 1. One may ask if there is an increasing trend in the mortality rates. We refer the reader for other examples of applications to Chen et al. (1997), to Arase et al. (2001) for controlled clinical trials and to Ku et al. (2001) for community based surveys.

In such problems, one usually observes at time values $t_1 < t_2 < ... < t_k$ a random sample of n_i binary variables $\left\{y_{ij}\right\}$ with y_{ij} taking values 1 or 0 with unknown probabilities p_i and $1 - p_i$ respectively. The observed data may be viewed as :

	t_1	t_2	$\cdots$	t_k	$Total$
	y_1	y_2	$\cdots$	y_k	y
	$n_1 - y_1$	$n_2 - y_2$	$\cdots$	$n_k - y_k$	$n - y$
$Total$	n_1	n_2	$\cdots$	n_k	n

where $y_i = \sum_j y_{ij}$, $y = \sum_i y_i$. Let $\bar{p}_i = y_i/n_i$ and $\bar{p} = y/n$. Without loss in generality, we shall be interested in detecting a monotone increasing trend in p_i. There are a number of different approaches to the problem as discussed in Terpestra (1952) and Armitage (1955). One approach founded on regression and analysis of variance and discussed below leads to the Cochran-Armitage test (Cochran 1954; Armitage 1955). Williams (1988) and Chen et al. (1997) have noted that when the expected values $n_i p_i$, or $n_i (1 - p_i)$ are small, the normal approximation may become unreliable. As a consequence the Cochran-Armitage test becomes conservative and may lead to a type I error rate greater than to the prescribed significance level Portier and Hoel (1984), Hothorn and Bretz (2000). Corcoran et al. (2000) have noted that the Cochran-Armitage test is very sensitive to the choice of scores and conclude that this “makes its general use suspect”. Williams (1988) has proposed an exact conditional test whereby the test statistic is calculated for all possible 2 x k tables with identical marginal totals. This is the randomization model in which the groups or treatments have the experimental units assigned to them at random. Randomization models are predominantly chosen in biomedical research. Neuhäuser (2006) proposed a modification of the Baumgartner-Weiβ-Schindler (1998) two-sample statistic which utilizes ranks instead of scores. A simulation study is inconclusive and does not reveal a clear winner. In section 2, we consider the problem using ranking methods.

Let $\{x_i\}$ represent arbitrary pre-selected scores, $x_1 < x_2 < ... < x_k$ which mimic a monotone increasing time trend. The linear model regression of y_{ij} expressed by

$$p_i = \alpha + \beta (x_i - \bar{x}_k) \tag{1}$$

and subject to the constraint that $\sum_i n_i (x_i - \bar{x}_k) = 0$ where $\bar{x}_k = \sum n_i x_i / n$, yields estimates

$$\hat{\alpha} = \frac{\sum_i n_i \bar{p}_i}{\sum_i n_i} = \bar{p}, \ \ \hat{\beta} = \frac{\sum_i n_i (x_i - \bar{x}_k)(\bar{p}_i - \bar{p})}{\sum_i n_i (x_i - \bar{x}_k)^2}.$$

The hypothesis of homogeneity is

$$H_0 : p_i = p, i = 1, ..k$$

(or equivalently $\beta = 0$). Possible alternatives are

$$\begin{aligned} H_1 &: \ p_1 \le p_2 \le ... \le p_k. \text{with at least one strict inequality} \\ H_2 &: \ p_i \ne p_j, \ \text{for at least one pair } (i, j) \end{aligned}$$

Hypothesis H_2 can also be expressed as

$$\sum_{i<j} \left(p_i - p_j\right) = \sum_i (k + 1 - 2i)\, p_i \ne 0 \tag{2}$$

(Van Eeden & Hemelrijk, 1955).

Under the null hypothesis of homogeneity, the estimate of variance of $\hat{\beta}$ is given by

$$V\left(\hat{\beta}\right) = \frac{\bar{p}(1 - \bar{p})}{\sum_i n_i (x_i - \bar{x}_k)^2}$$

and consequently we reject H_0 in favor of H_1 (or equivalently $\beta > 0$) for large values of the statistic

$$\hat{\beta} / \sqrt{V\left(\hat{\beta}\right)} = \frac{\sum_i n_i (x_i - \bar{x}_k)(\bar{p}_i - \bar{p})}{\sqrt{\bar{p}(1 - \bar{p})} \sqrt{\sum_i n_i (x_i - \bar{x}_k)^2}}$$

which for large samples has a standard normal. If we suppose further that $\bar{p}$ is small (of the order of $1\% - 2\%$) so that $\bar{p}^2$ is negligible, then the statistic becomes

$$\frac{\sum_i n_i (x_i - \bar{x}_k)(\bar{p}_i - \bar{p})}{\sqrt{\sum_i n_i \bar{p} (x_i - \bar{x}_k)^2}}.$$

The difference between observed and expected frequencies $n_i (\bar{p}_i - \bar{p})$ are multiplied by the score effect $(x_i - \bar{x}_k)$ in the numerator whereas the square of the score effects are weighted by the expected frequencies in the denominator. The statistic provides a test of trend in frequencies as opposed to a test of trend on the proportions.

Since the sample correlation between $\{\bar{p}_i - \bar{p}\}$ and $\{x_i - \bar{x}_k\}$ is given by

$$\begin{aligned} & \frac{\sum_i n_i (x_i - \bar{x}_k)(\bar{p}_i - \bar{p})}{\sqrt{\sum_i n_i (\bar{p}_i - \bar{p})^2 \sum_i n_i (x_i - \bar{x}_k)^2}} \\ = & \ \hat{\beta} \sqrt{\frac{\sum_i n_i (x_i - \bar{x}_k)^2}{\sum_i n_i (\bar{p}_i - \bar{p})^2}} \end{aligned}$$

we may view the test of monotonicity as equivalent to a test that the correlation is 0.

Alternatively, the test of homogeneity may be conducted by treating the data as coming from a 2 x k contingency table. That tests rejects H_0 in favor of H_2 for large values of the statistic

$$\begin{aligned} & \sum_i \frac{(n_i \bar{p}_i - n_i \bar{p})^2}{n_i \bar{p}} + \sum_i \frac{(n_i (1 - \bar{p}_i) - n_i (1 - \bar{p}))^2}{n_i (1 - \bar{p})} \\ = & \ \frac{\sum_i n_i (\bar{p}_i - \bar{p})^2}{\bar{p}(1 - \bar{p})} \end{aligned}$$

which for large samples $n_i \to \infty$, has a chi square distribution with $(k-1)$ degrees of freedom. In that case there is no need to define scores. We note that even though H_1 is included in H_2, "one-sided" tests which focus strictly on H_1 will in general be more powerful.

The regression model may be extended to apply to two or more groups of individuals. For example, to model the birth rates of men and women in the population, the functional regression model becomes

$$p_i = \alpha + \beta(x_i - \bar{x}_k) + \gamma(\alpha' + \beta'(x_i - \bar{x}_k)) \tag{3}$$

where γ takes value 1 for the first group and 0 otherwise.

In section two, we propose a general approach for constructing nonparametric test statistics based on the ranks of the observations to test H_0 against H_2. We obtain the asymptotic null distributions for these test statistics whenever either the sample sizes gets large or the number of time points k gets large. In section three, we report on the results of some simulation studies and note that the test statistics perform well under a variety of different underlying patterns. We present an application and conclude with some final remarks.

2. The Construction of the Test Statistics

In the previous section we considered the problem of testing for trend in proportions when scores provide a proxy for time. In this section we consider the problem from an entirely new perspective using methods based on the ranks of the data. The approach by-passes the use of scores. We first provide an introduction to statistical methods based on ranks.

A complete ranking of n objects is a permutation of the integers $(1,...,n)$. For any two rankings $\mu = (\mu(1),...,\mu(n))'$, $\nu = (\nu(1),...,\nu(n))'$, we may define the following measures of similarities due to Spearman, Kendall and Hamming respectively:

$$\begin{aligned}
\mathcal{A}_S(\mu,\nu) &= \sum_{i=1}^{n}\left(\mu(i) - \frac{n+1}{2}\right)\left(\nu(i) - \frac{n+1}{2}\right)\\
\mathcal{A}_K(\mu,\nu) &= \sum_{i<j} sgn(\mu(j)-\mu(i))\, sgn(\nu(j)-\nu(i))\\
\mathcal{A}_H(\mu,\nu) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left(I[\mu(i)=j] - \frac{1}{n}\right)\left(I[\nu(i)=j] - \frac{1}{n}\right)
\end{aligned}$$

where $sgn(x)$ is either 1 or -1 depending on whether $x > 0$ or $x < 0$ and where $I[\cdot]$ is the indicator function which is 1 or 0 depending on whether the statement in brackets holds or not. Measures of similarity can be used to define rank correlations to provide tests of trend and of independence. A review of related results may be found in Alvo and Cabilio (1992).

In what follows, we make use of the notion of tie compatibility (see Alvo & Cabilio, 1999) to extend the measures of similarity defined above to the case where ties occur in the data.

Definition 1. A tied ordering of n objects is a partition into e sets, $1 \le e \le n$, each of which contains d_i objects, $d_1 + d_2 + ... + d_e = n$, so that the d_i objects in each set share the rank i, $1 \le i \le e$. Such a tie pattern is denoted by $\delta = (d_1, d_2, ..., d_e)$. The ranking denoted by $\nu_\delta = (\nu_\delta(1), \nu_\delta(2), ..., \nu_\delta(n))$, resulting from such an ordering, is a tied ranking, and is one of $n!/(d_1!d_2!...d_e!)$ possible permutations.

Example 1. Suppose that $n = 7$ objects have the ranking (3172465), that is object 1 is ranked 3, object 2 is ranked 1, object 3 is ranked 7, and so on. Suppose that ties are allowed and that the ordering assumes the form $\langle(24)(157)(6)(3)\rangle$, where now the parentheses indicate members of the same tie class, so that objects 2 and 4 receive rank 1, objects 1, 5, and 7 receive rank 2, and objects 6 and 3 receive ranks 3 and 4 respectively. The tied ranking then becomes (2141232). In this case $e = 4, d_1 = 2, d_2 = 3, d_3 = d_4 = 1$.

The ranking which describes the time points may be viewed as a tied ranking with tie pattern

$$\delta_1 = (n_1, n_2, ..., n_k) \tag{4}$$

and with ordering

$$\left\langle (1,...,n_1)(n_1+1,...,n_1+n_2)...\left(\sum_{i=1}^{k-1} n_i + 1, ..., \sum_{i=1}^{k} n_i\right)\right\rangle \tag{5}$$

On the other hand, the binary variables y_{ij} have with $e = 2$ the simple tie pattern

$$\delta_2 = (y, n-y) \tag{6}$$

We now define tie compatibility whereby a tie ranking could be conceived as having arisen from a complete ranking in which some objects are grouped as being of equivalent standing. A re-ranking will then produce the tied ranking. All complete rankings which could give rise in this way to the specified tied ranking are then said to be compatible to it. More precisely, we have the following.

Definition 2. A complete ranking of n objects is compatible with a tied ranking of these objects with tie pattern $\delta = (d_1, d_2, ..., d_e)$, if every pair of objects which receive distinct ranks is given the same relative ranking in both rankings. We shall denote by $C(\mu)$ and $C(\nu)$ the class of complete rankings compatible to μ, ν respectively.

Example 2. Suppose that $k = 2, n_1 = 3, n_2 = 2$, and we observe 2 successes 1 failure at time t_1, and 1 success 1 failure at time t_2. The compatibility class corresponding to the time ranking ν contains the 12 permutations obtained by permuting ranks $1, 2, 3$ among themselves and ranks $4, 5$ among themselves

$$C(\nu) = \left\{ \begin{array}{c} (123|45), (132|45), (213|45), (231|45), (321|45), (312|45), \\ (123|54), (132|54), (213|54), (231|54), (321|54), (312|54) \end{array} \right\}$$

On the other hand, there are a total of 6 patterns of rankings with (2 $successes$, 1 $failure$) at time t_1, and (1 $success$, 1 $failure$) at time t_2 as follows:

$$(341|52), (351|42), (451|32), (342|51), (352|41), (452|31).$$

Permuting the entries in blocks 1 and 2 respectively, we obtain a total of 72 compatible rankings in the class $C(\mu)$.

Definition 3. The measure of similarity for the case of tied rankings $\mu_{\delta_1}, \nu_{\delta_2}$ is defined to be the conditional expectation

$$A(\mu_{\delta_1}, \nu_{\delta_2}) = E[\mathcal{A}(\mu, \nu)|C(\mu), C(\nu)]$$

where the expectation is computed by averaging over the complete rankings compatible to $\mu_{\delta_1}, \nu_{\delta_2}$.

In the next section, we compute the statistics corresponding to tied rankings for each of the similarity measures defined above.

2.1 The Test Statistic Corresponding to Spearman Similarity

To that end we note that

$$\begin{aligned} & \mathcal{A}_S(\mu_{\delta_1}, \nu_{\delta_2}) \\ = & \sum_{i=1}^{n} E\left[\left(\mu(i) - \frac{n+1}{2}\right)|C(\mu)\right] E\left[\left(\nu(i) - \frac{n+1}{2}\right)|C(\nu)\right] \end{aligned}$$

The average of the compatible ranks at time t_1 is $g_1 = \left(\frac{n_1+1}{2}\right)$, at time t_2 it is $g_2 = n_1 + \left(\frac{n_2+1}{2}\right)$ and so on. In general at time t_l the average rank is $g_l = \sum_{i=1}^{l-1} n_i + \left(\frac{n_l+1}{2}\right), l = 1, .., k$. Clearly, $\sum_{i=1}^{k} n_i g_i = \frac{n(n+1)}{2}$. Hence, the conditional expectation

$$E[\nu(l)|C(\nu)] = g_l, l = 1, ..., n_i$$

Turning attention now to the binary observations, the average rank for the $n - y$ which take value 0 is $l_1 = \frac{n-y+1}{2}$ whereas the average rank for the y observations which take value 1 is $l_2 = n - y + \left(\frac{y+1}{2}\right) = n - \frac{y}{2} + \frac{1}{2}$ It follows that at time t_i,

$$E[\mu(l)|C(\mu)] = \left\{ \begin{array}{ll} l_1 & l = 1, ..., n_i - y_i, \\ l_2 & l = n_i - y_i + 1, ..., n_i, \end{array} \right.$$

Since $l_2 - l_1 = \frac{n}{2}$ and $\sum_{i=1}^{k} n_i g_i = \frac{n(n+1)}{2}$, we have that

$$\begin{aligned} \mathcal{A}_S(\mu_{\delta_1}, \nu_{\delta_2}) &= \sum_{i=1}^{k} \left[g_i - \frac{n+1}{2}\right] [l_2 y_i + (n_i - y_i) l_1] \\ &= \frac{n}{2} \sum_{i=1}^{k} \left[g_i - \frac{n+1}{2}\right] y_i \end{aligned}$$

We shall define the Spearman statistic to be

$$\begin{aligned} S &= \sum_{i=1}^{k}\left[g_i - \frac{n+1}{2}\right] y_i \\ &= \sum_{i=1}^{k} n_i c_i \left(\overline{p_i} - \overline{p}\right) \end{aligned} \quad (7)$$

where $c_i = g_i - \frac{n+1}{2}$ and $\sum_{i=1}^{k} n_i c_i = 0$.

2.2 The Test Statistic Corresponding to Kendall Similarity

We now consider the test statistic when using Kendall's similarity measure. Note that within a given time period, the difference between ties is zero and hence there is no contribution to the distance. Between different time periods we have

$$\begin{aligned} & \sum_{i<j}^{n} E\left[sgn\left(\mu(j) - \mu(i)\right)|C(\mu)\right] E\left[sgn\left(\nu(j) - \nu(i)\right)|C(\nu)\right] \\ =& \sum_{i<j}^{k}\left[(n_i - y_i)\, y_j - \left(n_j - y_j\right) y_i\right] \\ =& \sum_{i<j}^{k}\left[n_i y_j - n_j y_i\right] \end{aligned}$$

Now

$$\begin{aligned} \sum_{j=i+1}^{k} n_j &= \left[n - n_i - \sum_{j=1}^{i-1} n_j\right] \\ &= \left[n - g_i - \frac{n_i - 1}{2}\right] \end{aligned}$$

Hence,

$$\begin{aligned} \mathcal{A}_K\left(\mu_{\delta_1}, \nu_{\delta_2}\right) &= \sum_{i=1}^{k} y_i\left[g_i - \frac{n_i+1}{2}\right] - \sum_{i=1}^{k} y_i\left[n - g_i - \frac{n_i - 1}{2}\right] \\ &= 2\sum_{i=1}^{k}\left[g_i - \frac{n+1}{2}\right] y_i \end{aligned}$$

Hence, the Kendall and Spearman statistics are equivalent.

2.3 The Test Statistic Corresponding to Hamming Similarity

In this section, we consider a test statistic based on the Hamming similarity. Returning to Example 2, we may define a matrix of scores given by

$$\begin{array}{c} t_1 \\ t_1 \\ t_1 \\ t_2 \\ t_2 \end{array} \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

where rows indicate time periods and columns indicate rank. At time t_1, there are 3 observations and hence ranks $1, 2, 3$ will occur $1/3$ of the time. Similarly, at time t_2, ranks $4, 5$ occur with probability $1/2$. Entries outside the block diagonals are 0.

Hence, in general we have for time t_i, $l = \sum_{q=1}^{i-1} n_q + 1, ..., \sum_{q=1}^{i} n_q$

$$E\left[I\left(\nu(l) = j\right)|C(\nu)\right] = \left\{ \begin{array}{l} \frac{1}{n_i},\ \sum_{q=1}^{i-1} n_q + 1 \le j \le \sum_{q=1}^{i} n_q \\ \\ 0,\ \text{otherwise} \end{array} \right\}$$

We now consider the data ranking in accordance with the model. First the complete rankings corresponding to successes (and correspondingly failures) can be permuted among themselves in

$$\frac{y!}{\Pi y_i!}\frac{(n-y)!}{\Pi\,(n_i-y_i)!}$$

ways. Since entries in time blocks can be permuted in $\Pi n_i!$ ways, we have that the total number of compatible rankings is given by the product

$$\frac{y!\,(n-y)!}{\Pi y_i!\,(n_i-y_i)!}\Pi n_i!$$

Define the set of integers

$$\begin{aligned} B_i &= \left\{\sum_{q=1}^{i-1} n_q+1,...,\sum_{q=1}^{i} n_q\right\}, i=1,...,k \\ A_{0y} &= \{1,2,...,n-y\} \\ A_{1y} &= \{n-y+1,...,n\} \end{aligned}$$

and let

$$\begin{aligned} w_{0i} &= Card\ \ (A_{0Y}\cap B_i) \\ w_{1i} &= Card\ \ (A_{1Y}\cap B_i) \end{aligned}$$

where $w_{0i}=0$ if $y=n$ and $w_{1i}=0$ if $y=0$.

It follows that at time t_i and $l=\sum_{q=1}^{i-1} n_q+1,...,\sum_{q=1}^{i} n_q$,

$$\begin{aligned} E\left[I\left(\mu(l)=j\right)|C\left(\mu\right)\right] &= P\left(\mu\left(l\right)=j|C\left(\mu\right)\right) \\ &= \begin{cases} \frac{y_i}{n_i y}, & j\epsilon A_{1y} \\ \frac{(n_i-y_i)}{n_i(n-y)}, & j\epsilon Ay \end{cases} \end{aligned}$$

In fact, at time t_i, $j\epsilon A_{0y}$, we have

$$P\left(\mu\left(l\right)=j|C\left(\mu\right)\right)=\frac{\frac{y!(n-y-1)!(n_i-1)!}{\Pi y_i!\Pi_{j\neq i}\left(n_j-y_j\right)!(n_i-y_i-1)!}\Pi_{j\neq i}n_j!}{\frac{y!(n-y)!}{\Pi y_i!(n_i-y_i)!}\Pi n_i!}$$

In Example 2, we obtain a total of 72 compatible rankings. Averaging the frequency of occurrences of each ranking leads to the following matrix of scores:

$$\begin{array}{c} t_1 \\ t_1 \\ t_1 \\ t_2 \\ t_2 \end{array}\begin{pmatrix} 1/6 & 1/6 & 2/9 & 2/9 & 2/9 \\ 1/6 & 1/6 & 2/9 & 2/9 & 2/9 \\ 1/6 & 1/6 & 2/9 & 2/9 & 2/9 \\ 1/4 & 1/4 & 1/6 & 1/6 & 1/6 \\ 1/4 & 1/4 & 1/6 & 1/6 & 1/6 \end{pmatrix}$$

Noting that each row at time t_i is repeated n_i times, and since $w_{0i}+w_{1i}=n_i$, the Hamming similarity measure for tied data becomes

$$\begin{aligned} & \mathcal{A}_H\left(\mu_{\delta_1},\nu_{\delta_2}\right) \\ =& \sum_{i,j}^{n}\left(E\left[I\left(\mu(i)=j\right)|C\left(\mu\right)\right]E\left[I\left(\nu(i)=j\right)|C\left(\nu\right)\right]\right)-1 \\ =& \sum_{i=1}^{k}\left[w_{1i}\frac{y_i}{n_i y}+w_{0i}\frac{(n_i-y_i)}{n_i\left(n-y\right)}\right]-1 \\ =& \sum_{i=1}^{k}w_{1i}\left[\frac{\bar{p}_i}{y}-\frac{(1-\bar{p}_i)}{(n-y)}\right] \\ =& \frac{1}{n\bar{p}\left(1-\bar{p}\right)}\sum_{i=1}^{k}\left[n_i w_i\left(\bar{p}_i-\bar{p}\right)\right] \end{aligned}$$

where $w_i = \frac{w_{1i}}{n_i}$. We may define the Hamming statistic as

$$H = \frac{1}{\bar{p}(1-\bar{p})} \sum_{i=1}^{k} \left[n_i w_i (\bar{p}_i - \bar{p}) \right] \tag{8}$$

The weights $\{w_i\}$ which depend on y may also be expressed as

$$w_i = \begin{cases} 0 & n - y + 1 > \sum_{q=1}^{i} n_q \\ \frac{1}{n_i} \left(\sum_{q=1}^{i} n_q - (n-y) \right) & \sum_{q=1}^{i-1} n_q + 1 \le n - y + 1 \le \sum_{q=1}^{i} n_q \\ 1 & n - y + 1 < \sum_{q=1}^{i-1} n_q + 1 \end{cases} \tag{9}$$

and in this form we see that most of the weight is assigned to later time points. In the next section, we determine the asymptotic null distribution of the Spearman and Hamming statistics.

3. The Asymptotic Distribution of the Test Statistics Under the Null Hypothesis

In this section, we consider the asymptotic distribution of the Spearman and Hamming test statistics under the null hypothesis. Let $\{y_i\}$ be k independent binomials (n_i, p_i) and suppose we would like to test H_0 vs H_1. In most applications the asymptotic situation of interest occurs when

$$n_i \to \infty, with\ \frac{n_i}{n} \to \lambda_i > 0, i = 1, ..., k. \tag{10}$$

In Theorem 1, we shall show that the Spearman statistic has an asymptotic normal distribution under either (10) or under the condition that the $\{n_i\}$ are bounded while $k \to \infty$.

Theorem 1. *Suppose that both $y \to \infty$ and $n - y \to \infty$ as $n \to \infty$. Under H_o the Spearman test statistic has asymptotically a standard normal distribution, i.e.*

$$\frac{\sum_{i=1}^{k} n_i c_i (\bar{p}_i - \bar{p})}{\sqrt{\sum_{i=1}^{k} c_i^2 n_i p_i (1 - p_i)}} \to_d N(0, 1)$$

under either i) (10) or ii) the $\{n_i\}$ are bounded and $k \to \infty$.

We may estimate p_i either by $\bar{p}_i$ or by $\bar{p}$. In the first case, the test rejects whenever

$$\frac{\sum_{i=1}^{k} c_i y_i}{\sqrt{\sum_{i=1}^{k} c_i^2 n_i \bar{p}_i (1 - \bar{p}_i)}} \ge z_\alpha$$

where z_α is the upper $100(1-\alpha)\%$ percentage point from a standard normal distribution. The expression for the estimate of the asymptotic power becomes

$$\begin{aligned} & 1 - \Phi\left(z_\alpha - \frac{\sum_{i=1}^{k} n_i c_i p_i}{\sqrt{\sum c_i^2 n_i p_i (1 - p_i)}} \right) \\ = \quad & \Phi\left(\frac{\sum_{i=1}^{k} n_i c_i p_i}{\sqrt{\sum_{i=1}^{k} c_i^2 n_i p_i (1 - p_i)}} - z_\alpha \right) \end{aligned}$$

It is seen that the power converges to 1 with increasing n.

Alternatively, we may use the statistic

$$\frac{\sum_{i=1}^{k} n_i c_i (\bar{p}_i - \bar{p})}{\sqrt{\bar{p}(1-\bar{p})} \sqrt{\sum_{i=1}^{k} n_i c_i^2}} \tag{11}$$

which in the simulation studies reported appears to more closely attain the prescribed significance level. In the case of equal sample sizes, $n_i = n_0$ say, the test statistic (11) takes the simpler form

$$\frac{\sum_{i=1}^{k} \left(i - \frac{k+1}{2} \right) (\bar{p}_i - \bar{p})}{\sqrt{n_0} \sqrt{\bar{p}(1-\bar{p})} \sqrt{\frac{k(k^2-1)}{12}}}$$

Theorem 2. *Under the null hypothesis and under* (10), *the conditional mean and variance of* $\sum_{i=1}^{k} w_i y_i$ given *that* $\sum_{i=1}^{k} y_i = y$, *are given respectively by* $y\bar{p}$ *and*

$$\bar{p}(1-\bar{p}) \sum_{i=1}^{k} n_i (w_i - \bar{p})^2$$

In the next theorem we demonstrate that asymptotically the Hamming statistic converges to a normal distribution.

Theorem 3. *Suppose that both* $y \to \infty$ *and* $n - y \to \infty$ *as* $n \to \infty$. *Under* H_o, *the Hamming test statistic has asymptotically a standard normal distribution, i.e.*

$$\frac{\sum_{i=1}^{k} [n_i w_i (\bar{p}_i - \bar{p})]}{\sqrt{\bar{p}(1-\bar{p}) \sum_{i=1}^{k} n_i (w_i - \bar{p})^2}} \to_d N(0,1)$$

under either i) (10) *or ii*) *the* $\{n_i\}$ *are bounded and* $k \to \infty$.

Both test statistics share the same general form as the regression statistic, namely

$$\frac{\sum_i n_i (x_i - \bar{x}_k)(\bar{p}_i - \bar{p})}{\sqrt{\sum_i n_i (x_i - \bar{x}_k)^2} \sqrt{\bar{p}(1-\bar{p})}}, \bar{x}_k = \sum n_i x_i / n \tag{12}$$

where for Spearman, $x_i = g_i$ and for Hamming $x_i = w_i$.

We may also consider the asymptotic null distribution of the Spearman and Hamming test statistics under the condition that $k \to \infty$.in lieu of (10).

4. Simulation Study

Comparing Table 2 with Table 3, it is seen that the significance level actually attained by the Spearman similarity measure when the $\{p_i\}$ in the variance expressions are estimated by $\overline{p}$ are closer to the prescribed 5% level than when $\{\overline{p_i}\}$ are used. From Table 4 we note that the level actually attained by the Hamming similarity measure is closer to the prescribed 5% level than Spearman's.

In Table 5 we report on simulations for the power when $k = 5$. Three cases were considered: proportions which are strictly increasing, proportions which are non-decreasing and some which have no particular pattern. It can be seen that the Spearman measure is clearly only superior in the first two cases. Predictably the power is smaller when the $\{p_i\}$ are closer together than when they are further apart.

5. Applications

Returning to the example on mortality rates in South Africa, we calculated values of 260.1 and 70.3 for the Spearman and Hamming similarity measures respectively. These yielded p-values $< 10^{-4}$.

Another application deals with water quality at Hong Kong beaches. Table 6 shows the geometric E. Coli count for each of 6 beaches in the Sai Kung district of Hong Kong during the period 1986-2009. A beach is classified as good if the count is at most 24. The Spearman test statistic yielded a value of 22.98 which points to strong evidence of an upward trend in the annual proportion of good beaches. The Hamming test statistic on the other hand yielded a value of 2.90 which has a p-value equal to 0.0018.

6. Conclusion

An approach has been proposed for constructing nonparametric tests of trend in proportions. Similarity measures due to Spearman, Kendall and Hamming led to new test statistics. It was shown that the Spearman and Kendall measures led to identical test statistics. The asymptotic null distributions for the Spearman and Hamming test statistics were shown to be normal. Simulations were performed to check on the significance level attained. It was seen that the Hamming measure was closer to the prescribed level. On the other hand, the Spearman similarity measure attained greater power under a variety of alternatives. Two applications were considered.

Acknowledgment

The authors are grateful to Professor Philip Yu for suggesting the application to water quality data.

This research was supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada.

Appendix

Proof of Theorem 1.

i) The Spearman statistic (7) is expressible as a linear combination of independent binomials, $\sum_{i=1}^{k} c_i y_i$ where $c_i = g_i - \frac{n+1}{2}$ and $g_i = \sum_{i=1}^{i-1} n_i + \left(\frac{n_i+1}{2}\right)$. Hence, under (10), we have approximately $y_i \approx_d N(n_i p_i, n_i p_i (1-p_i))$. In view of the independence of the $\{y_i\}$,

$$\sum c_i y_i \approx_d N\left(\sum c_i n_i p_i, \sum c_i^2 n_i p_i (1-p_i)\right)$$

and hence

$$\frac{\left(\sum c_i y_i - \sum c_i n_i p_i\right)}{\sqrt{\sum c_i^2 n_i p_i (1-p_i)}} \to_d N(0,1).$$

Under H_0, $\sum c_i n_i p_i = 0$ and hence

$$\frac{\sum c_i y_i}{\sqrt{\sum c_i^2 n_i p_i (1-p_i)}} \to_d N(0,1).$$

The theorem follows.

ii) For this part, note that the Spearman statistic is expressible as

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} c_i y_{ij}$$

where the $\left\{y_{ij}\right\}$ are Bernoulli (p_i). We need to show that

$$\frac{\max_{1 \le i \le k} c_i^2}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} c_i^2} \to 0, as\ k \to \infty$$

Note that if $n_i \le M$, for all i,

$$\begin{aligned} c_i^2 &\le g_i^2 + \frac{(n+1)^2}{4} \\ &\le M^2 \left(i - \frac{1}{2}\right)^2 + \frac{1}{4} + \frac{(n+1)^2}{4} = O\left(k^2\right) \end{aligned}$$

Moreover, since $n = \sum_{i=1}^{k} n_i \le Mk$

$$\begin{aligned} \sum_{i=1}^{k} \sum_{j=1}^{n_i} c_i^2 &= \sum_{i=1}^{k} n_i g_i^2 - \frac{n(n+1)^2}{4} \\ &\ge \sum_{i=1}^{k} n_i g_i^2 - \frac{M^3 k^3}{4} \\ &= \sum_{i=1}^{k} n_i \left(\sum_{j=1}^{i-1} n_j\right)^2 + \sum_{i=1}^{k} (n_i + 1) \left(\sum_{j=1}^{i-1} n_j\right) + \sum_{i=1}^{k} n_i \left(\frac{n_i+1}{2}\right)^2 - \frac{M^3 k^3}{4} \\ &\ge \sum_{i=1}^{k} n_i (i-1)^2 + 2 \sum_{i=1}^{k} (i-1) = O\left(k^3\right) \end{aligned}$$

It follows that

$$\frac{\max_{1 \le i \le k} c_i^2}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} c_i^2} = O(\frac{1}{k}) \to 0, as\ k \to \infty$$

and the theorem is proved.

Proof of Theorem 2.

Turning attention to the Hamming statistic (8), we note that under H_0 and conditional on $\sum_{i=1}^{k} y_i = y$, the joint probability distribution of the $\{y_i\}$ is the multivariate hypergeometric

$$\frac{\Pi_{i=1}^{k}\binom{n_i}{y_i}}{\binom{n}{y}} \tag{13}$$

It follows that

$$\begin{aligned} E\left(y_i|H_0, y\right) &= \frac{n_i}{n} y \\ Var\left(y_i|H_0, y\right) &= \frac{n_i\left(n-n_i\right)}{n^2\left(n-1\right)} y\left(n-y\right) \\ Cov\left(y_i, y_j|H_0, y\right) &= -\frac{n_i n_j}{n^2\left(n-1\right)} y\left(n-y\right), i \neq j \end{aligned}$$

Hence,

$$\begin{aligned} E\left(\sum_{i=1}^{k} w_i y_i|H_0, y\right) &= \left(\sum_{i=1}^{k} w_i \frac{n_i}{n} y\right) \\ &= y\bar{p} \end{aligned}$$

and

$$\begin{aligned} Var\left(\sum_{i=1}^{k} w_i y_i|H_0, y\right) &= \sum_i w_i^2 Var\left(y_i|H_0, y\right) + \sum_{i\neq j} w_i w_j Cov\left(y_i, y_j|H_0, y\right) \\ &= \frac{y\left(n-y\right)}{n^2\left(n-1\right)}\left[\sum_i n_i\left(n-n_i\right) w_i^2 - \sum_{i\neq j} n_i n_j w_i w_j\right] \\ &= \frac{y\left(n-y\right)}{n^2\left(n-1\right)}\left[n\sum_i n_i w_i^2 - \left(\sum n_i w_i\right)^2\right] \\ &= \bar{p}\left(1-\bar{p}\right)\sum_{i=1}^{k} n_i\left(w_i-\bar{p}\right)^2 \end{aligned}$$

Proof of Theorem 3.

Given (10), it follows that (13) converges to a multinomial distribution independent of p given by

$$\binom{y}{y_1 y_2 \ldots y_k}\lambda_1^{y_1}\lambda_2^{y_2}\ldots\lambda_k^{y_k}$$

Suppose now that the observed weights $\{w_i\}$ are the values obtained from an i.i.d. sample from a population with mean μ and variance σ^2. Consider a bootstrap random sample of size y, say $W_1^*, \ldots, W_y^*$ obtained with replacement from that population of observed weights $\{w_i\}$ in accordance with the distribution

$$P\left(W^* = w_i\right) = \lambda_i, i = 1, \ldots, k$$

The central limit theorem then asserts that for large y, the sum of the bootstrap sample

$$S^* = \sum_{i=1}^{y} W_i^*$$

is asymptotically normal. Now y will get large a.s. since $n \to \infty$ for otherwise $y/n \nrightarrow p$ a.s. and thus contradict the strong law of large numbers.

To relate the Hamming distribution to that of the bootstrap, let x_i denote the number of times w_i is selected in the resampling. It follows that the vector $(x_1, \ldots, x_k)$ has a multinomial distribution $(y; \lambda_1, \ldots, \lambda_k)$ and S^* can be equivalently written as

$$S^* = \sum_{i=1}^{k} w_i x_i =_d \sum_{i=1}^{k} w_i y_i.$$

Consequently, under the null hypothesis for large n and conditional on y

$$\frac{\sum_{i=1}^{k}\left[n_i w_i\left(\bar{p}_i - \bar{p}\right)\right]}{\sqrt{\bar{p}\left(1-\bar{p}\right)\sum_{i=1}^{k} n_i\left(w_i - \bar{p}\right)^2}} \rightarrow_d N(0,1)$$

and hence this is also true unconditionally. Note that

$$\sum_{i=1}^{k} n_i\left(w_i - \bar{p}\right) = \left(\sum_{i=1}^{k} w_{1i}\right) - n\bar{p} = 0.$$

ii) For this part, it follows that since $0 \le w_i \le 1$, for all i and

$$\sum_{i=1}^{k} n_i w_i^2 \ge y \tag{14}$$

$$\frac{\max_{1\le i\le k} w_i^2}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} w_i^2} = O\left(\frac{1}{y}\right) \rightarrow 0$$

and the theorem is proved.

References

Alvo, M., & Cabilio, P. (1995). Testing ordered alternatives in the presence of incomplete data. *Journal of the American Statistical Association, 90*, 1015-1024. http://dx.doi.org/10.2307/2291338

Alvo, M., & Cabilio, P. (1999). A general rank based approach to the analysis of block data. *Commun. Statist.-Theory Meth, 28*(1), 197-215. http://dx.doi.org/10.1080/03610929908832291

Alvo, M., & Pan, J. (1997). A general theory of hypothesis testing based on rankings. *Journal of Statistical Planning and Inference, 61*, 219-248. http://dx.doi.org/10.1016/S0378-3758(97)89436-0

Arase, Y., Chayama, K., & Ikeda, K. (2001). Randomized controlled clinical trial of lymphoblastoid interferon-alpha for chronic hepatitis C. *Hepathology, 21*, 55-66.

Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics, 11*(3), 375-386. http://dx.doi.org/10.2307/3001775

Baumgartner, W., Weiβ, P., & Schindler, H. (1998). A nonparametric test for the general two sample problem. *Biometrics, 54*, 1129-1135. http://dx.doi.org/10.2307/2533862

Chen, J. J., Kodell, R. L., & Pearce, B. A. (1997). Significance levels of randomization trend tests in the event of rare occurrences. *Biometrical Journal, 39*, 327-337. http://dx.doi.org/10.1002/bimj.4710390307

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics, 10*, 417-451. http://dx.doi.org/10.2307/3001616

Corcoran, C., Mehta, C., & Senchaudhuri, P. (2000). Power comparisons for tests of trend in dose-response studies. *Statistics in Medicine, 19*, 3037-3050. http://dx.doi.org/10.1002/1097-0258(20001130)19:22<3037::AID-SIM601>3.0.CO;2-7

Hong Kong water quality data from 1996 to 2000. [Online] Available: http://www.epd.gov.hk/epd/english/environmentinhk/water/beach_quality/bwq _report8600_appendixa.htm

Hong Kong water quality data from 1997 to 2009. [Online] Available: http://www.epd.gov.hk/epd/english/environmentinhk/water/beach _quality/files/bwq_report2009.pdf

Hothorn, L. A., & Bretz, F. (2000). Evaluation of animal carcinogenicity studies : Cochran-Armitage trend test vs multiple contrast tests. *Biometrical Journal, 42*, 553-567. http://dx.doi.org/10.1002/1521-4036(200009)42:5<553::AID-BIMJ553>3.0.CO;2-R

Ku, J. H., Kim, M. E., Lee, N. K., & Park, Y. H. (2001). The prevalence of chronic prostatitis-like symptoms in young men: a community-based survey. *Urological Research, 29*, 108-112. http://dx.doi.org/10.1007/s002400100172

Neuhäuser, M. (2006). An exact test for trend among binomial proportions based on a modified Baumgartner-Wei β-Schindler statistic. *Journal of Applied Statistics, 33*, 79-88.

Portier, C., & Hoel, D. (1984). Type I error of trend tests in proportions and the design of cancer screens. *Commun. Statist.-Theory Meth., 13*, 1-14. http://dx.doi.org/10.1080/03610928408828657

Terpstra, T. J. (1952). The asymptotic normality and consistency of Kendall's test against trend, when ties are present. *Indagationes Mathematicae, 14*, 327-333.

Van Eeden, C., & Hemelrijk, J. (1955). A test for the equality of probabilities against a class of specified alternative hypotheses including tesnt I. *Proc. Kon. Ned. Akad. van Wet., A*(58), 191-198.

Williams, D. A. (1998). Tests for differences between several small proportions. *Applied Statistics, 37*, 421-434. http://dx.doi.org/10.2307/2347316

Table 1. Mortality statistics for South Africa 2000-2008

Year	Number of deaths	Population size
2000	416, 155	43, 789, 115
2001	454, 882	43, 997, 828
2002	502, 050	44, 187, 637
2003	556, 779	44, 344, 136
2004	576, 709	42, 718, 530
2005	598, 131	42, 768, 678
2006	612, 778	43, 647, 658
2007	603, 094	43, 586, 097
2008	592, 073	43, 421, 021

Table 2. Significance level for Spearman's similarity when p_i is estimated by $\overline{p_i}$

p	10	20	30	50
0.1	0.018	0.035	0.045	0.053
0.2	0.048	0.057	0.055	0.055
0.3	0.062	0.060	0.058	0.054
0.4	0.067	0.057	0.057	0.055
0.5	0.071	0.055	0.059	0.053

Table 3. Significance level for Spearman's similarity when p_i is estimated by $\overline{p}$

p	10	20	30	50
0.1	0.044	0.045	0.050	0.053
0.2	0.051	0.054	0.052	0.053
0.3	0.052	0.053	0.052	0.052
0.4	0.052	0.049	0.052	0.052
0.5	0.048	0.052	0.049	0.051

Table 4. Significance level for Hamming's similarity

p	10	20	30	50
0.1	0.056	0.064	0.061	0.056
0.2	0.054	0.056	0.056	0.055
0.3	0.054	0.052	0.052	0.052
0.4	0.054	0.055	0.052	0.053
0.5	0.053	0.053	0.050	0.050

Table 5. Power for Spearman (S) and Hamming (H) similarity measures using a 5% significance level

					n_i = 10		n_i = 30	
p_1	p_2	p_3	p_4	p_5	H	S	H	S
0.1	0.2	0.4	0.5	0.6	0.77	0.89	0.99	1.00
0.3	0.6	0.7	0.8	0.9	0.88	0.94	1.00	1.00
0.5	0.6	0.7	0.8	0.9	0.58	0.72	0.94	0.99
0.2	0.25	0.3	0.35	0.4	0.25	0.29	0.50	0.60
0.1	0.3	0.3	0.4	0.5	0.51	0.64	0.88	0.97
0.3	0.5	0.5	0.5	0.9	0.58	0.81	0.94	1.00
0.4	0.5	0.6	0.8	0.8	0.62	0.75	0.95	0.99
0.2	0.5	0.3	0.8	0.4	0.51	0.39	0.98	0.80
0.6	0.2	0.4	0.8	0.5	0.46	0.18	0.88	0.39

Table 6. Annual Geometric Mean E. Coli level (per 100ml) in the Sai Kung District. Beaches: Clear Water Bay First (1), Clear Water Bay Second (2), Hap Mun Bay (3), Kiu Tsui (4), Silverstrand (5), Trio (6); number of good beaches (7)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1986	102	69	9	18	255	49	2
1987	133	52	6	9	62	32	2
1988	39	35	4	3	129	35	2
1989	80	38	3	5	192	23	3
1990	51	42	4	5	89	31	2
1991	30	14	2	4	106	14	4
1992	52	42	2	5	94	32	2
1993	31	16	3	4	56	20	4
1994	30	35	3	3	72	14	3
1995	55	39	6	3	226	16	3
1996	34	43	5	5	126	29	2
1997	62	66	3	5	148	30	2
1998	41	44	2	4	99	21	3
1999	11	12	2	4	32	17	5
2000	16	26	2	5	61	10	4
2001	28	22	1	5	100	12	4
2002	28	14	2	4	133	6	4
2003	17	21	4	5	97	10	5
2004	9	10	3	17	74	2	5
2005	16	19	4	14	67	6	5
2006	20	13	4	11	30	5	5
2007	14	9	3	6	33	2	5
2008	11	19	5	12	35	12	5
2009	15	27	3	19	31	5	4

Scale Parameter Estimation of the Laplace Model Using Different Asymmetric Loss Functions

Sajid Ali (Corresponding author)

Department of Statistics, Quaid-i-Azam University, Islamabad 45320, Pakistan

&

Department of Decision Sciences, Bocconi University, Milan, Italy

E-mail: sajidali.qau@hotmail.com

Muhammad Aslam

Department of Statistics, Quaid-i-Azam University, Islamabad 45320, Pakistan

E-mail: aslamsdqu@yahoo.com

Nasir Abbas

Department of Statistics, Government Post Graduate College, Jhang, Pakistan

E-mail: nabbasgcj@yahoo.com

Syed Mohsin Ali Kazmi

Sustainable Development Policy Institute Islamabad, 44000, Pakistan

E-mail: mohsinstats@yahoo.com

Received: January 29, 2012 Accepted: February 20, 2012 Published: May 1, 2012

doi:10.5539/ijsp.v1n1p105 URL: http://dx.doi.org/10.5539/ijsp.v1n1p105

Abstract

In the last few decades, there has been an emergent interest in the construction of flexible parametric classes of probability distributions in Bayesian as compared to Classical approach. In present study Bayesian Analysis of Laplace model using Inverted Gamma, Inverted Chi-Squared informative, Levy and Gumbel Type-II priors is discussed. The properties of posterior distribution, credible interval, highest posterior density region (HPDR) and Bayes Factor are discussed in current study. Bayes estimators are derived under squared error loss function (SELF), precautionary loss function, weighted squared error loss function and modified (quadratic) squared error loss function. Hyperparameters are determined through Empirical Bayes method. The estimates are also compared using the posterior risks (PRs) under the said loss functions. The priors and loss functions are compared using a real life data set.

Keywords: Censored sampling, Squared error loss function (SELF), Precautionary loss function, Weighted squared error loss function (WSELF), Modified squared error loss function (MSELF), Credible interval (CI), Highest posterior density region (HPDR), Fixed test termination time, Informative prior, Posterior risk, Empirical Bayes

1. Introduction

Very few real world events that we need to statistically study are symmetrical. Thus the popular normal model would not be a useful model for studying every phenomenon. The normal model at times is a poor description of observed phenomena. Skewed models, which exhibit varying degrees of asymmetry, are a necessary component of the modeler's tool kit. A variety of different forms of the Laplace distribution have been introduced and applied in several areas of real world problems. So Laplace distribution is getting popularity due to simplicity of its characteristic function as an alternative to Normal model. In particular, various forms of the Laplace distribution have been introduced and applied in several areas. In this article Laplace model is considered as a lifetime model using complete and censored data because in many experiment we cannot continue experiment up to last item failure due to cost or time constraints. So in these cases censoring is unavoidable.

It needs to be mentioned here Laplace distribution has been considered before in literature. Kappenman (1975) obtain the conditional confidence interval for the parameters of a double exponential distribution by finding the conditional distributions of the pivotal quantities for location and scale parameters. Kappenman (1977) propose a procedure for obtaining and lower probability tolerance interval for a proportion of a population for the unknown two parameters of Laplace (double exponential) distribution. Balakrishnan and Chandramouleeswaran (1996) present an estimators for the reliability function based on the best linear unbiased estimators (BLUEs) for the location and scale parameters of Laplace distribution based on Type-II censored samples, and they show that obtained estimator is almost unbiased at varying level of reliability, Childs and Balakrishnan (2000) consider the progressively type-II right censored sample for analysis of Laplace distribution. Childs and Balakrishnan (1997) constrain the maximum likelihood estimators for the parameters of Laplace distribution under the general Type-II censored samples, which are simple linear functions of the order statistics. They also examine the asymptotic variance through the Fisher information matrix. Nadarajah (2009) utilize the Laplace distribution random variables with application to price indices. Kozubowski and Nadarajah (2010) motivated by the recent popularity of Laplace distribution, provide a comprehensive review of the known Laplace distributions along with their properties and applications Nadarajah (2010) obtain two posterior distributions for the mean of the Laplace distribution by deriving the distributions of the product XY and the ratio X/Y when X and Y are Student's t and Laplace random variables distributed independently of each other.

The rest of the article is organized as follows. The Laplace model likelihood, posterior distribution using different informative priors and their properties are defined in Section 2 for complete and censored data. Empirical Bayes method, predictive distributions (prior and posterior) are provided in Section 3. Simulation study of different properties defined in section 2 is conducted using simulated data and real data in Section 4. Section 5 contains discussion and derivation of Bayesian Interval Estimation using real data. Bayesian hypotheses testing discussed in Section 6, while Bayes Estimates and their respective Posterior Risks are evaluated using different loss functions in Section 7 for real and simulated data. A model comparison framework is presented in Section 8. Some concluding remarks and further research proposal are given in last Section 8.

2. Posterior Distribution and Likelihood Function

The posterior distribution summarizes available probabilistic information on the parameters in the form of prior distribution and the sample information contained in the likelihood function. The likelihood principle suggests that the information on the parameter should depend only on its posterior distribution. Bayesian scientist's job is to assist the investigator to extract features of interest from the posterior distribution. In this section we will use the Laplace distribution as sampling distribution mingles with the informative and noninformative priors for the derivation of posterior distribution. A random variable x is said to possess a Laplace distribution if it has the following form:

$$f(x) = \frac{1}{2\lambda}\exp\left(-\frac{|x|}{\lambda}\right), \lambda > 0, -\infty < x < \infty,$$

here λ is scale parameter and location parameter is zero. The likelihood function for a random sample x_1, x_2,.... x_n which is taken from Laplace distribution is:

$$L(.) = \frac{1}{2^n\lambda^n}\exp\left(-\sum_{i=1}^{n}|X_i|/\lambda\right) \tag{1}$$

The likelihood function for censored data as described Mendenhall and Hader (1958) is:

$$L(\lambda|\mathbf{x}) \propto \left[\prod_{j=1}^{r} f(x_j)\right][1 - F(T)]^{(n-r)}$$

$$L(\lambda, \mathbf{x}) \propto \left(\frac{1}{\lambda}\right)^{r} e^{-\frac{1}{\lambda}\left[\sum_{j=1}^{r}|x_j|+(n-r)T\right]}$$

Where T is time, r is for censored observations and n-r are uncensored observations.

2.1 Posterior Distribution Using Inverted Gamma Prior

Informative priors are those that deliberately insert information that researchers have at hand. This seems like a reasonable and reasoned approach since previous scientific knowledge should play a role in statistical inference. The author is deliberately manipulating prior information to obtain a desired posterior result. An informative prior provides more information than the non-informative priors, therefore the analysis using these prior more accurate and informative than classical approach.

The Inverted Gamma prior of λ is defined as:

$$f(\lambda) = \frac{b^a}{\Gamma(a)}\left(\frac{1}{\lambda}\right)^{(a+1)} \exp(-b/\lambda), a, b, \lambda > 0 \tag{2}$$

The posterior distribution of parameter λ for the given data $(x = x_1, x_2, ..., x_n)$ using equation (1) and (2) is:

$$p(\lambda|\mathbf{x}) = \frac{\left(b + \sum_{i=1}^n |x_i|\right)^{(a+n)}}{\Gamma(a+n)} \frac{1}{\lambda^{(a+n+1)}} \exp\left(-\left(b + \sum_{i=1}^n |x_i|\right)/\lambda\right), \lambda > 0$$

which is the density similar to Inverted-Gamma distribution. So $\lambda|x \sim G^{-1}(\alpha, \beta)$ where G^{-1} for Inverted Gamma distribution and $\alpha = a + n, \beta = b + \sum_{i=1}^n |x_i|$.

For Censored data the posterior distribution of parameter λ

$$p(\lambda|\mathbf{x}) = \frac{\left(\sum_{j=1}^r \left|x_j\right| + (n-r)T + b\right)^{(r+a)}}{\Gamma(r+a)} \left(\frac{1}{\lambda}\right)^{(r+a+1)} e^{-\frac{1}{\lambda}\left[\sum_{j=1}^r |x_j| + (n-r)T + b\right]}, \lambda > 0$$

which is the density similar to Inverted-Gamma distribution. So $\lambda|x \sim G^{-1}\left(a + r, \sum_{j=1}^r \left|x_j\right| + (n-r)T + b\right)$

2.2 Posterior Distribution Using the Inverted Chi-squared (ICS) Prior

Using ICS an informative prior with hyperparameters 'a' and 'b', this is defined by the following density:

$$f(\lambda) = \frac{(b/2)^{\frac{a}{2}}}{2^{\frac{a}{2}}\Gamma(a/2)}\left(\frac{1}{\lambda}\right)^{\left(\frac{a}{2}+1\right)} \exp(-b/2\lambda), a, b, \lambda > 0 \tag{3}$$

We obtain the posterior distribution of λ for the given data $(x = x_1, x_2, ..., x_n)$ using equation (1) and (3) as:

$$p(\lambda|x) = \frac{\left((b/2) + \sum_{i=1}^n |x_i|\right)^{((a/2)+n)}}{\Gamma((a/2)+n)} \frac{1}{\lambda^{((a/2)+n+1)}} \exp\left(-\left((b/2) + \sum_{i=1}^n |x_i|\right)/\lambda\right), \lambda > 0$$

Which is similar to inverted-gamma distribution, so $\lambda|x \sim G^{-1}(\alpha, \beta)$, where $\alpha = \frac{a}{2} + n, \beta = \frac{b}{2} + \sum_{i=1}^n |x_i|$.

For Censored, the posterior distribution as follows

$$p(\lambda|x) = \frac{\left(\sum_{j=1}^r \left|x_j\right| + (n-r)T + \frac{b}{2}\right)^{\left(r+\frac{a}{2}\right)}}{\Gamma\left(r + \frac{a}{2}\right)} \left(\frac{1}{\lambda}\right)^{\left(r+\frac{a}{2}+1\right)} e^{-\frac{1}{\lambda}\left[\sum_{j=1}^r |x_j| + (n-r)T + \frac{b}{2}\right]}, \lambda > 0$$

Which is similar to inverted-gamma distribution, so $\lambda|x \sim G^{-1}(\alpha, \beta)$, where $\alpha = \frac{a}{2} + r, \beta = \sum_{j=1}^r \left|x_j\right| + (n-r)T + \frac{b}{2}$.

2.3 Posterior Distribution Using Levy Prior

The Levy prior of λ is defined as:

$$f(\lambda) = \sqrt{\frac{b}{2\pi}}\left(\frac{1}{\lambda}\right)^{(1.5)} \exp(-b/2\lambda), b, \lambda > 0 \tag{4}$$

The posterior distribution of parameter λ for the given data $(x = x_1, x_2, ..., x_n)$ using equation (1) and (4) is:

$$p(\lambda|\mathbf{x}) = \frac{\left(\frac{b}{2} + \sum_{i=1}^n |x_i|\right)^{\left(\frac{1}{2}+n\right)}}{\Gamma\left(\frac{1}{2}+n\right)} \frac{1}{\lambda^{\left(\frac{3}{2}+n\right)}} \exp\left(-\left(\frac{b}{2} + \sum_{i=1}^n |x_i|\right)/\lambda\right), \lambda > 0$$

which is the density similar to Inverted-Gamma distribution and we can write it as $\lambda|x \sim G^{-1}(\alpha, \beta)$, where $\alpha = 0.5 + n, \beta = 0.5b + \sum_{i=1}^n |x_i|$ and here G^{-1} for Inverted Gamma distribution.

For Censored data the posterior distribution of parameter λ

$$p(\lambda|\mathbf{x}) = \frac{\left(\sum_{j=1}^r \left|x_j\right| + (n-r)T + \frac{b}{2}\right)^{\left(r+\frac{1}{2}\right)}}{\Gamma\left(r + \frac{1}{2}\right)} \left(\frac{1}{\lambda}\right)^{\left(r+\frac{3}{2}\right)} e^{-\frac{1}{\lambda}\left[\sum_{j=1}^r |x_j| + (n-r)T + \frac{b}{2}\right]}, \lambda > 0$$

Here also $\lambda|x \sim G^{-1}\left(0.5 + r, \sum_{j=1}^{r}\left|x_j\right| + (n-r)T + 0.5b\right)$.

2.4 Posterior Distribution Using Gumbel Type-II (GTII) Prior

Using GTII an informative prior with hyperparameter 'b', this is defined by the following density:

$$f(\lambda) = b\left(\frac{1}{\lambda}\right)^{(2)} \exp\left(-b/\lambda\right), b, \lambda > 0 \tag{5}$$

We obtain the posterior distribution ofλ for the given data ($x = x_1, x_2, ..., x_n$) using equation (1) and (5) as:

$$p(\lambda|\mathbf{x}) = \frac{\left(b + \sum_{i=1}^{n}|x_i|\right)^{(1+n)}}{\Gamma(1+n)} \frac{1}{\lambda^{(2+n)}} \exp\left(-\left(b + \sum_{i=1}^{n}|x_i|\right)/\lambda\right), \lambda > 0$$

which is inverted-Gamma distribution, $\lambda|x \sim G^{-1}(\alpha, \beta)$, where$\alpha = 1 + n, \beta = b + \sum_{i=1}^{n}|x_i|$.

For Censored, the posterior distribution as follows

$$p(\lambda|\mathbf{x}) = \frac{\left(\sum_{j=1}^{r}\left|x_j\right| + (n-r)T + b\right)^{(r+1)}}{\Gamma(r+1)} \left(\frac{1}{\lambda}\right)^{(r+2)} e^{-\frac{1}{\lambda}\left[\sum_{j=1}^{r}\left|x_j\right| + (n-r)T + b\right]}, \lambda > 0$$

So $\lambda|x \sim G^{-1}(\alpha, \beta)$, where $\alpha = 1 + r, \beta = \sum_{j=1}^{r}\left|x_j\right| + (n-r)T + b$.

2.5 Properties of Posterior Distribution Assuming Different Priors

Since our Posterior distribution is Inverted-Gamma distribution, here we give general forms of properties for Inverted-Gamma distribution.

Mean $= \beta/(\alpha - 1)$, $\alpha > 1$, Mode $= \beta/(\alpha + 1)$. Variance $= \beta^2/\left((\alpha-1)^2(\alpha-2)\right)$, $\alpha > 2$. Coefficient of Skewness $= \left(4\sqrt{\alpha-1}\right)/(\alpha-3)$, $\alpha > 3$. Excess Kurtosis $= (30\alpha - 66)/((\alpha-3)(\alpha-4))$, $\alpha > 4$, where α and β are respective posterior distribution parameters.

3. Elicitation of Hyperparameters through Empirical Bayes

Empirical Bayes procedures utilize past data as a means for bypassing the necessity of identifying a completely unknown and unspecified prior distribution having frequency interpretation. Grabski and Sarhan (1996), Sarhan (2003) used empirical Bayes estimation in the case of exponential reliability while Ahn et al. (2006) consider this procedure for the hazard rate estimation of a mixture model with censored lifetimes. The empirical Bayes approach may be considered as a two-stage estimation procedure in which the hyperparameter is estimated from the marginal distribution and then the parameter is estimated using 'pseudo-prior' where the hyperparameters are replaced by the estimation of first stage. Robbins (1955) sought a representation of the desired Bayes rule in terms of the marginal distribution of data (called prior predictive distribution) and then uses the data to estimate it rather than the prior distribution. For excellent discussion about empirical Bayes see Kass and Steffey (1989), Bansal (2007). We use the Inverted Gamma, levy, Gumbel type-II and Inverted Chi-Squared informative priors for the cradle of prior predictive distribution which will be used for empirical Bayes procedure. Followings are the prior predictive equations used for Empirical Bayes procedure. Let Y be the random variable from Laplace distribution with unknown parameter λ.

$$f(y|\lambda) = \frac{1}{2\lambda}\exp\left(-\frac{|y|}{\lambda}\right), \lambda > 0, -\infty < y < \infty$$

The prior predictive distribution is obtained by using the following equation:

$$p(y) = \int_0^\infty p(\lambda) f(y, \lambda) d\lambda$$

3.1 Prior Predictive Distribution Using Inverted Gamma Prior

The prior predictive distribution using Inverted Gamma prior is:

$$p(y) = \frac{ab^a}{2\left(|y| + b\right)^{(a+1)}}, -\infty < y < \infty$$

We use above equation of prior predictive distribution for the elicitation of hyperparameters 'a' and 'b'.

3.2 Prior Predictive Distribution Using Inverted-chi Prior

The prior predictive distribution using Inverted Chi-Squared prior is:

$$p(y) = \frac{ab^{(a/2)}}{2^{(a/2)+1}\left(|y| + (b/2)\right)^{((a/2)+1)}}, -\infty < y < \infty$$

We use above equation of prior predictive distribution for the elicitation of hyperparameters 'a' and 'b'.

3.3 Prior Predictive Distribution When Prior is Levy Distribution

The prior predictive distribution using Levy prior is:

$$p(y) = \frac{\sqrt{b}}{2^{(5/2)}\left(|y| + (b/2)\right)^{(3/2)}}, -\infty < y < \infty$$

3.4 Prior Predictive Distribution When Prior is Gumbel Type-II Distribution

The prior predictive distribution using Gumbel Type-II prior is:

$$p(y) = \frac{b}{2\left(|y| + b\right)^{(2)}}, -\infty < y < \infty$$

3.5 Predictive Distribution

Often it is necessary to make *predictions* of future observations, based on our best inferences on parameters determined through observations already made. Posterior predictive distribution is the product of the posterior distribution and (conditional) independence (given the parameters) of the new observation from the "learning sample".

$$p(y\,|\mathbf{x}) = \int_{-\infty}^{\infty} p(\lambda\,|x)p(y\,|\lambda)\; d\lambda$$

Where y = x_{n+1} be the future observation given the sample information $x = x_1, x_2, ...x_n$, from of the model with unknown parameter λ.

Predictive Interval can be obtained as:

$$\int_{-\infty}^{\infty} p(y|\mathbf{x})dy = \int_{-\infty}^{L} p(y|x)dy = \frac{\alpha}{2} = \int_{U}^{\infty} p(y|x)dy$$

3.6 Predictive Distribution and Predictive Intervals Using Inverted-Gamma Prior

The posterior predictive distribution for a future random variable 'y' given that data $x = x_1, x_2, ...x_n$ is:

$$p(y|\mathbf{x}) = \frac{(n+a)\left[b + \sum_{i=1}^{n}|x_i|\right]^{(a+n)}}{2\left(\sum_{i=1}^{n}|x_i| + |y| + b\right)^{(a+n+1)}}, -\infty < y < \infty \tag{6}$$

And Predictive interval is:

$$\frac{\alpha}{2} = \frac{\left[b + \sum_{i=1}^{n}|x_i|\right]^{(a+n)}}{2\left(\sum_{i=1}^{n}|x_i| + L + b\right)^{(a+n)}}, \quad \frac{\alpha}{2} = \frac{\left[b + \sum_{i=1}^{n}|x_i|\right]^{(a+n)}}{2\left(\sum_{i=1}^{n}|x_i| + U + b\right)^{(a+n)}}$$

For Censored Data

$$p(y|\mathbf{x}) = \frac{\left[\sum_{j=1}^{r}\left|x_j\right| + (n-r)T + b\right]^{(r+a)}(r+a)}{2\left[\sum_{j=1}^{r}\left|x_j\right| + (n-r)T + b + |y|\right]^{(r+a+1)}}, -\infty < y < \infty$$

And Predictive interval is:

$$\frac{\alpha}{2} = \frac{\left[\sum_{j=1}^{r}\left|x_j\right| + (n-r)T + b\right]^{(r+a)}}{2\left[\sum_{j=1}^{r}\left|x_j\right| + (n-r)T + b + L\right]^{(r+a)}}, \quad \frac{\alpha}{2} = \frac{\left[\sum_{j=1}^{r}\left|x_j\right| + (n-r)T + b\right]^{(r+a)}}{2\left[\sum_{j=1}^{r}\left|x_j\right| + (n-r)T + b + U\right]^{(r+a)}}$$

where a and b are hyperparameters determined by empirical Bayes method.

3.7 Predictive Distribution and Predictive Intervals Using Inverted-Chi Prior

The posterior predictive distribution for a random variable ‘y’ given that data $x = x_1, x_2, ...x_n$ is:

$$p(y|\mathbf{x}) = \frac{\left(n+\frac{a}{2}\right)\left[\frac{b}{2}+\sum_{i=1}^{n}|x_i|\right]^{\left(\frac{a}{2}+n\right)}}{2\left(\sum_{i=1}^{n}|x_i|+|y|+\frac{b}{2}\right)^{\left(\frac{a}{2}+n+1\right)}}, -\infty < y < \infty$$

Equations of Predictive Intervals are:

$$\frac{\alpha}{2} = \frac{\left[\frac{b}{2}+\sum_{i=1}^{n}|x_i|\right]^{\left(\frac{a}{2}+n\right)}}{2\left(\sum_{i=1}^{n}|x_i|+L+\frac{b}{2}\right)^{\left(\frac{a}{2}+n\right)}}, \quad \frac{\alpha}{2} = \frac{\left[\frac{b}{2}+\sum_{i=1}^{n}|x_i|\right]^{\left(\frac{a}{2}+n\right)}}{2\left(\sum_{i=1}^{n}|x_i|+U+\frac{b}{2}\right)^{\left(\frac{a}{2}+n\right)}}$$

For Censored Data

$$p(y|\mathbf{x}) = \left[\frac{\left(r+\frac{a}{2}\right)\left[\sum_{j=1}^{r}\left|x_j\right|+(n-r)T+\frac{b}{2}\right]^{\left(r+\frac{a}{2}\right)}}{2\left[\sum_{j=1}^{r}\left|x_j\right|+(n-r)T+\frac{b}{2}+|y|\right]^{\left(r+\frac{a}{2}+1\right)}}\right], -\infty < y < \infty$$

Equations of Predictive Intervals are:

$$\frac{\alpha}{2} = \frac{\left(\sum_{j=1}^{r}\left|x_j\right|+(n-r)T+\frac{b}{2}\right)^{\left(r+\frac{a}{2}\right)}}{2\left(\sum_{j=1}^{r}\left|x_j\right|+(n-r)T+\frac{b}{2}+L\right)^{\left(r+\frac{a}{2}\right)}}, \quad \frac{\alpha}{2} = \frac{\left(\sum_{j=1}^{r}\left|x_j\right|+(n-r)T+\frac{b}{2}\right)^{\left(r+\frac{a}{2}\right)}}{2\left(\sum_{j=1}^{r}\left|x_j\right|+(n-r)T+\frac{b}{2}+U\right)^{\left(r+\frac{a}{2}\right)}}$$

Where ‘a’ and ‘b’ are hyperparameters to be elicited by empirical Bayes procedure.

3.8 Predictive Distribution and Predictive Intervals Using Levy Prior

The posterior predictive distribution for a future random variable ‘y’ given that data $x = x_1, x_2, ...x_n$ is:

$$p(y|\mathbf{x}) = \frac{\left(n+\frac{1}{2}\right)\left[\frac{b}{2}+\sum_{i=1}^{n}|x_i|\right]^{\left(\frac{1}{2}+n\right)}}{2\left(\sum_{i=1}^{n}|x_i|+|y|+\frac{b}{2}\right)^{\left(\frac{3}{2}+n\right)}}, -\infty < y < \infty$$

And Predictive interval is:

$$\frac{\alpha}{2} = \frac{\left[\frac{b}{2}+\sum_{i=1}^{n}|x_i|\right]^{\left(\frac{1}{2}+n\right)}}{2\left(\sum_{i=1}^{n}|x_i|+L+\frac{b}{2}\right)^{\left(\frac{1}{2}+n\right)}}, \quad \frac{\alpha}{2} = \frac{\left[\frac{b}{2}+\sum_{i=1}^{n}|x_i|\right]^{\left(\frac{1}{2}+n\right)}}{2\left(\sum_{i=1}^{n}|x_i|+U+\frac{b}{2}\right)^{\left(\frac{1}{2}+n\right)}}$$

For Censored Data

$$p(y|\mathbf{x}) = \frac{\left[\sum_{j=1}^{r}\left|x_j\right|+(n-r)T+\frac{b}{2}\right]^{\left(r+\frac{1}{2}\right)}\left(r+\frac{1}{2}\right)}{2\left[\sum_{j=1}^{r}\left|x_j\right|+(n-r)T+\frac{b}{2}+|y|\right]^{\left(r+\frac{3}{2}\right)}}, -\infty < y < \infty$$

And Predictive interval is:

$$\frac{\alpha}{2} = \frac{\left[\sum_{j=1}^{r}\left|x_j\right|+(n-r)T+\frac{b}{2}\right]^{\left(r+\frac{1}{2}\right)}}{2\left[\sum_{j=1}^{r}\left|x_j\right|+(n-r)T+\frac{b}{2}+L\right]^{\left(r+\frac{1}{2}\right)}}, \quad \frac{\alpha}{2} = \frac{\left[\sum_{j=1}^{r}\left|x_j\right|+(n-r)T+\frac{b}{2}\right]^{\left(r+\frac{1}{2}\right)}}{2\left[\sum_{j=1}^{r}\left|x_j\right|+(n-r)T+\frac{b}{2}+U\right]^{\left(r+\frac{1}{2}\right)}}$$

<Table 1>

3.9 Predictive Distribution and Predictive Intervals Using Gumbel Type-II Prior

The posterior predictive distribution for a random variable ‘y’ given that data $x = x_1, x_2, ...x_n$ is:

$$p(y|\mathbf{x}) = \frac{(n+1)\left[b+\sum_{i=1}^{n}|x_i|\right]^{(n+1)}}{2\left(\sum_{i=1}^{n}|x_i|+|y|+b\right)^{(n+2)}}, -\infty < y < \infty$$

Equations of Predictive Intervals are:

$$\frac{\alpha}{2} = \frac{\left[b + \sum_{i=1}^{n} |x_i|\right]^{(n+1)}}{2\left(\sum_{i=1}^{n} |x_i| + L + b\right)^{(n+1)}}, \quad \frac{\alpha}{2} = \frac{\left[b + \sum_{i=1}^{n} |x_i|\right]^{(n+1)}}{2\left(\sum_{i=1}^{n} |x_i| + U + b\right)^{(n+1)}}$$

For Censored Data

$$p(y|\mathbf{x}) = \frac{(r+1)\left[\sum_{j=1}^{r} \left|x_j\right| + (n-r)T + b\right]^{(r+1)}}{2\left[\sum_{j=1}^{r} \left|x_j\right| + (n-r)T + b + |y|\right]^{(r+2)}}, -\infty < y < \infty$$

Equations of Predictive Intervals are:

$$\frac{\alpha}{2} = \frac{\left[\sum_{j=1}^{r} \left|x_j\right| + (n-r)T + b\right]^{(r+1)}}{2\left[\sum_{j=1}^{r} \left|x_j\right| + (n-r)T + b + L\right]^{(r+1)}}, \quad \frac{\alpha}{2} = \frac{\left[\sum_{j=1}^{r} \left|x_j\right| + (n-r)T + b\right]^{(r+1)}}{2\left[\sum_{j=1}^{r} \left|x_j\right| + (n-r)T + b + U\right]^{(r+1)}}$$

4. Simulation Study

For conducting simulation study, data from Laplace model is generated using inverse transformation. Simulation study of different properties discussed in above section are given in appendix. From Tables 1-8 in appendix, one can easily observe that mode is more close to the true supposed parameter values than the mean and variance value increase as we increase the parameter value and especially for censored data it is clear that variance is more than complete data because in censored data we are using less information as compared to complete data set. Similar conclusion for skewness and kurtosis tables can be drawn. One thing is common in Table 1-11 that with the increase of sample size our parameter values approach to true values of parameters and variances, skewness and kurtosis decreases. Also comparing the prior, one can easily observe that GTII prior has better results in terms of posterior risk, skewness and kurtosis than the other informative priors.

4.1 Real Data Findings

Childs and Balakrishnan (1996) consider the data of mean difference of 33 year in flood stage for two stations on the Fox River in Wisconsin. Following table shows the different posterior distribution properties using real data set assuming IG and ICS informative priors.

GT-II prior performance is better than IG, ICHI and LP prior based on variance, mode, skewness and kurtosis. Graphical presentation as follows:

<Figure 1-4>

Above figures confirm the same behaviors in terms of their shape but Gumbel Type-II prior is different from others and also from above figure one can easily observe that using censored data information lost but due to application this is unavoidable. Figure 1(i) is for complete data set using informative priors and Figure 1(ii) is for complete and censored data using IG prior. Here in figure 1(ii) IG (c) denotes the censored data. One can easily observe that using censored data we lost information.

5. Bayesian Interval Estimation

In Bayesian analysis, credible interval becomes the counterpart of the classical confidence interval when we put hyper-parameter (s) values equal to zero. Also credible interval may be unique for all models. The difference between credible interval and highest density region (HPDR) is that credible interval is comparable with classical interval but HPDR is unique.

5.1 Credible Interval Assuming Informative Priors

Let $A_c = \frac{2(B)}{\chi^2_{2(A),1-\frac{\alpha}{2}}}$, $B_c = \frac{2(B)}{\chi^2_{2(A),\frac{\alpha}{2}}}$, where A, B are the parameters of respective posterior distribution. Thus a $(1-\alpha)100\%$ credible interval is $\{A_c \le \lambda \le B_c\}$, (See Ahmed Abu-Taleb et al. (2007) and Saleem and Aslam (2009)).

5.2 Highest Posterior Density Region (HPDR)

The HPD interval is defined on the posterior density such that the posterior density at every point inside the HPD interval is greater than the posterior density at every point outside the HPD interval. An interval (λ_1, λ_2) would be a (1- α) 100%

HPD interval for λ if it satisfy the following two conditions simultaneously.

$$\left.\begin{array}{l}\int_{\lambda_1}^{\lambda_2} p(\lambda|\mathbf{x})d\lambda = 1-\alpha \\ p(\lambda_1|\mathbf{x}) = p(\lambda_2|\mathbf{x})\end{array}\right\} \quad (7)$$

5.2.1 HPDR for Inverted Gamma Prior

Using above two conditions given in equation (7) solving for IG prior we get $(an+1)\ln(\lambda_1/\lambda_2)-(b+\sum_{i=1}^n |x_i|)(\lambda_2^{-1}-\lambda_1^{-1}) = 0$ and $\Gamma(a+n,(b+\sum_{i=1}^n |x_i|)/\lambda_1) - \Gamma(a+n,(b+\sum_{i=1}^n |x_i|)/\lambda_2) - (\alpha-1)\Gamma(a+n) = 0$ where $\Gamma(n-1,\sum_{i=1}^n |x_i|/\lambda_1) = \int_{\sum_{i=1}^n |x_i|/\lambda_1}^{\infty} \lambda^{n-1}\exp(-\lambda)d\lambda$ is incomplete gamma function.

For Censored data

λ and $\Gamma(a+r,(b+(n-r)T+\sum_{i=1}^r |x_i|)/\lambda_1) - \Gamma(a+r,(b+(n-r)T+\sum_{i=1}^r |x_i|)/\lambda_2) - (\alpha-1)\Gamma(a+r) = 0.$

5.2.2 HPDR for Inverted Chi-squared Prior

HPDR equations for Inverted Chi-Squared prior as follows $(0.5a+n+1)\ln(\lambda_1/\lambda_2) - (0.5b+\sum_{i=1}^n |x_i|)(\lambda_2^{-1}-\lambda_1^{-1}) = 0$ and $\Gamma(n+0.5a,(0.5b+\sum_{i=1}^n |x_i|)/\lambda_1) - \Gamma(n+0.5a,(0.5b+\sum_{i=1}^n |x_i|)/\lambda_2) - (\alpha-1)\Gamma(n+0.5a) = 0.$

For censored data

$(0.5a+r+1)\ln(\lambda_1/\lambda_2) - (0.5b+(n-r)T+\sum_{i=1}^r |x_i|)(\lambda_2^{-1}-\lambda_1^{-1}) = 0$ and λ (43)

5.2.3 HPDR for Levy Prior

Above two conditions given in equation (6), solving for Levy prior we get HPDR equations for Levy prior as follows: $(n+1.5)\ln(\lambda_1/\lambda_2)-(0.5b+\sum_{i=1}^n |x_i|)(\lambda_2^{-1}-\lambda_1^{-1}) = 0$ and $\Gamma(n+0.5,(0.5b+\sum_{i=1}^n |x_i|)/\lambda_1)-\Gamma(n+0.5,(0.5b+\sum_{i=1}^n |x_i|)/\lambda_2)-(\alpha-1)\Gamma(n+0.5) = 0.$

For censored data

λ and $\Gamma(r+0.5,(0.5b+(n-r)T+\sum_{i=1}^r |x_i|)/\lambda_1) - \Gamma(r+0.5,(0.5b+(n-r)T+\sum_{i=1}^r |x_i|)/\lambda_2) - (\alpha-1)\Gamma(r+0.5) = 0.$

5.2.4 HPDR for Gumbel Type-II Prior

HPDR equations for Gumbel Type-II prior as follows $(n+2)\ln(\lambda_1/\lambda_2) - (b+\sum_{i=1}^n |x_i|)(\lambda_2^{-1}-\lambda_1^{-1}) = 0$ and $\Gamma(n+1,(b+\sum_{i=1}^n |x_i|)/\lambda_1) - \Gamma(n+1,(b+\sum_{i=1}^n |x_i|)/\lambda_2) - (\alpha-1)\Gamma(n+1) = 0$, where $\Gamma(n-1,\sum_{i=1}^n |x_i|/\lambda_1) = \int_{\sum_{i=1}^n |x_i|/\lambda_1}^{\infty} \lambda^{n-1}\exp(-\lambda)d\lambda$ is incomplete gamma function.

For Censored data

For real data set the HPD and Credible intervals for different level of significance evaluated in the following table considering different informative priors.

<Table 2>

The HPD has smaller band than credible interval. ICSP and GTII has wider CI and HPD than IGP and LP.

6. Bayesian Hypotheses Testing and Bayes Factor

Bayesian hypothesis testing is less formal than non-Bayesian varieties. In fact, Bayesian researchers typically summarize the posterior distribution without applying a rigid decision process. If one wanted to apply a formal process, Bayesian decision theory is the way to go because it is possible to get a probability distribution over the parameter space and one can make expected utility calculations based on the costs and benefits of different outcomes. Considerable energy has been given, however, in trying to map Bayesian statistical models into the null hypothesis testing framework, with mixed results at best. This section contains the Bayesian hypotheses for parameterλ under the posterior distribution, considering using noninformative priors. Posterior probabilities are defined for the $H_1 : \lambda \geq \lambda_1$ vs. $H_2 : \lambda < \lambda_1$hypotheses:

$$P(H_1) = P(\lambda \geq \lambda_1) = \int_{\lambda_1}^{\infty} p(\lambda|x)d\lambda$$

where $p(\lambda|x)$ is the posterior distribution of λ given x and $P(H_2) = 1 - P(H_1)$.

Bayes Factors are the dominant method of Bayesian model testing. These are Bayesian analogues of likelihood ratio tests. The basic intuition is that prior and posterior information are combined in a ratio that provides evidence in favor of one model specification verses another. By dividing posterior odds under null and alternative hypotheses we get Bayes Factor. Bayes Factors are very flexible, allowing multiple hypotheses to be compared simultaneously and nested models are not required in order to make comparisons – it goes without saying that compared models should obviously have the same dependent variable.

The Bayes factor can be interpreted as the 'odds for H_1 to H_2 that is given by the data'. While the Bayesian approach typically avoids arbitrary decision thresholds, Jeffreys (1961) gives the following typology for comparing H_1 vs H_2.

$B \geq 1$ H_1 is supported $10^{-\frac{1}{2}} \leq$ B ≤ 1 minimal evidence against H_1, $10^{-1} \leq$ B $\leq 10^{-\frac{1}{2}}$ substantial evidence against H_1 $10^{-2} \leq$ B $\leq 10^{-1}$ strong evidence against H_1 $B < 10^{-2}$ decisive evidence against H_1. Here B is used for Bayes factor.

<Table 3>

For the smaller parameter value than Bayes Estimator value we does not get strong support for H_1 but when we go further from it we get strong evidence support for H_1. For testing $\lambda \leq 6.0$ we observe that decisive evidence againstH_1 for all priors, $\lambda \leq 7.5$ we have strong evidence againstH_1 for informative priors. Minimal evidence against H_1 occur in case of $\lambda \leq 9.0$ and in case of $\lambda \leq 11.0$, H_1 is strongly supported especially for Inverted Gamma prior. Similarly for censored data testing $\lambda \leq 15.0$ we observe that minimal evidence against H_1 for all priors, $\lambda \leq 16.5$ we have minimal evidence against H_1 using informative priors. We have strong evidence against H_1 occur in case of $\lambda \leq 18.0$ and $\lambda \leq 19.5$, H_1 is strongly supported especially for Gumbel prior. The values taken in complete data cannot be taken here because using that values we have decisive evidences due to involvement of time, we are using less information about data.

7. Bayes Estimators Under Different Loss Functions

This section spotlight on the derivation of the Bayes Estimator (BE) under different loss functions and their respective Posterior Risk (PR). The results are also compared their results for informative prior. Bayes decision is a decision d which minimizes risk function then d is the best decision. If the decision is choice of an estimator then the Bayes decision is a Bayes estimator. The Bayes estimator for different loss function is given below: we use four loss functions name as squared error loss function (SELF), weighted squared error loss function (WSELF), precautionary loss function and modified (quadratic) squared error loss function (M/QSELF).

The squared error loss function (SELF) was proposed by Legendre (1805) and Gauss (1810) to develop least square theory. Later, it was used in estimation problems when unbiased estimators of parameter were evaluated in terms of the risk function which becomes nothing but the variance of the estimators. It was also observed that SELF is a convex loss function and, therefore, restricts the class of estimators by excluding randomized estimators. The difficulty with unbounded loss function, like SELF is that Bayes estimates may change enormously when the observation of the random variable changes infinitesimally. Therefore, the investigator has to be absolutely precise about his probability statements. Furthermore, in real life situations, it is usually being impossible to lose an infinite amount of money. The extensive form of analysis provides Bayes estimate under SELF, as $E(\lambda|x)$. Also note that squared error loss function is not the only loss function for which posterior mean is the Bayes estimate. The natural exponential family $f(\lambda|x) = a(\lambda)b(x)\exp(\lambda x)$, the Bayes estimates under entropy loss function is the posterior mean. The Bayes estimate under the weighted SELF may not exist if the weight function $w(\lambda)$ i ncreases too fast to infinity. Norstrom (1996) introduced an alternative asymmetric precautionary loss function, and also presented a general class of precautionary loss functions as a special case. These loss functions approach infinitely near the origin to prevent underestimation, thus giving conservative estimators, especially when underestimation may lead to serious consequence. Since in risk analysis, both the potentiality of an undesired event and its consequences are investigated. This potentiality is usually measured by either a probability or a failure rate. The Bayes approach is widely applied to estimate this failure rate. When dealing with disastrous consequences, it can be worse to underestimate the potentiality of an event than to overestimate it. This is important when risk-level is the basis of a risk-reducing initiative, either by reducing the potentiality or the consequences. An erroneously low estimated risk-level can lead to the absence of necessary initiative to reduce the risk level. It is unreasonable to use a loss function that allows one to estimate a failure probability of zero. A positive loss function at the origin allows estimating zero, and in a risk analysis, estimating zero failure probability simply means that no risk is anticipated. Hence, a precautionary loss function is used. Also optimal policy selection has traditionally been discussed in relation to symmetric and often quadratic loss functions. So by using non-symmetric loss functions one is able to deal with cases where it is more damaging to miss the target on one side than the other.

7.1 Bayes Estimate and Posterior Risk Under Different Priors

Following tables gives the comparison of Bayes estimates and Posterior risk under different priors.

<Table 4>

7.2 Simulation of Bayes Estimates (BE) and Posterior Risk (PR)

Simulation is a flexible methodology, we can use it to analyze the behavior of a pattern of proposed business activity, new product, manufacturing line or plant expansion, and so on (analysts call this the 'system' under study). In simulation one generates a sample of random data in such a way that mimics a real problem and summarizes that sample in the same way. It is one of the most widely used quantitative methods because it is so flexible and can yield so many useful results. There are different method such as Monte Carlo simulation and Bootstrap to simulate the data. Here we did simulation in Minitab for non-informative prior. Following tables contains the Bayes estimates and their posterior risks.

It is immediate from appendix Table 17-24, as we increase sample size posterior risk comes down and also with the increase of parameters values posterior risk also increase. For censored data posterior risk is greater than the complete data because of time contribution. The choice of loss function as concerned, one can easily observe that modified squared error loss function has smaller posterior risk than other three loss functions.

7.3 Real Data Bayes Estimates (BE) and Posterior Risk (PR)

Following tables shows the Bayes Estimates (BE) and their Posterior Risk (PR) in brackets for real data

<Table 5>

Modified squared error loss function has smaller posterior risk than other three loss functions using Gumbel Type-II prior.

8. Model Comparison

The comparison of model performances is proposed to be based on the generated posterior predictive distributions. The criterion used to compare them is based on the use of the logarithmic score as a utility function in a statistical decision framework. This was proposed by Bernardo (1979) and used, for example, by Walker and Gutiérrez-Peña (1999) and Martín and Pérez (2009) in a similar context. In situations where the uncertainty is contained in the value of a future observation $y = x_{n+1}$, the logarithmic score $\log(p_k(y|x))$ is used, where $p_k(y|x)$ denotes the posterior predictive density under model M_k. Then, the posterior predictive expected utility is given by:

$\bar{U}_k = \int \log(p_k(y|x))\, p_k(y|x)dy$. The optimal solution to the decision problem of choosing among the competing models M $_0$; M $_1$, ...; M$_l$ is given by the model M$_{k*}$, such that: $\bar{U}_{k*} = \max\limits_{k\in\{0,1,\ldots,l\}} \bar{U}_k$. From a practical viewpoint, $\bar{U}_k$ can be estimated as: $\hat{\bar{U}}_k = \frac{1}{m}\sum_{i=1}^{m} \log(p_k(y_i|x))$, where y$_1$; y$_2$, ... , y$_m$ are an independent and identically distributed random sample from $p_k(y|x)$.

In order to illustrate this and the applicability of the proposed approaches, a performance comparison between posterior distribution using Inverted Gamma and Inverted Chi-Squared priors, a random sample of size 10 generated from laplace distribution with mean 0 and scale parameter equal to 4 using Minitab v 12 (0.97289, 4.02645, 5.07323, 0.75942, 2.09396, 0.98549, 2.22772, 5.06832, 0.40607, 3.74112). We get U$_{IG}$=-2.734129, U$_{ICS}$=-2.732629 U$_{Levy}$=-2.721256 and U$_{GTII}$=-2.714222 so Gumbel Type-II prior is best.

9. Conclusion and Suggestions

We consider the Bayesian analysis of the Laplace Model using complete and censored data assuming informative prior. Based on different properties of posterior distribution, hypotheses testing, HPD and credible intervals, we conclude that Gumbel Type II prior results are more precised than other prior. The choice of loss function as concerned, one can easily observed based on evidence (different properties as discussed above) that modified squared error loss function has smaller posterior risk. Model comparison method also suggest that Gumbel Type II prior is best. One thing is common as we increase sample size posterior risk comes down. Also note that we cannot compare the result of complete data with censored data because in censored data we are using less information than the complete data set. In future this work can be extended using truncated Laplace model and considering location parameter.

References

Ahn, S. E., Park, C. S., & Kim, H. M. (2007). Hazard rate estimation of a mixture model with censored lifetimes. *Stoch Environ Res Risk Assess, 21*, 711-716. http://dx.doi.org/10.1007/s00477-006-0082-1

Abu-Taleb, A. A., Smadi, M. M., & Alawneh, J. A. (2007). Bayes Estimation of the Lifetime Parameters for the Exponential Distribution. *Journal of Mathematics and Statistics, 3*(3), 106-108. http://dx.doi.org/10.3844/jmssp.2007.106.108

Balakrishnan, N., & Chandramouleeswaran, P. M. (1996). Reliability estimation and tolerance limits for Laplace distribution based on censored samples. *Microelectron Reliab., 36*, N (3), 375-378.

Bansal, A. K. (2007). Bayesian Parametric Inference. Narosa Publishing House Pvt. Ltd.

Bernardo, J. M. (1979). Expected information as expected utility. *Annals of Statistics*, 686-690. http://dx.doi.org/10.1214/aos/1176344689

Childs, A., & Balakrishnan, N. (2000). Conditional inference procedures for the Laplace distribution when the observed samples are progressively censored. *Metrika, 52*, N (3), 253-265.

Childs, A., & Balakrishnan, N. (1997). Maximum likelihood estimation of place Laplace parameters based on general Type-II censored samples. *Statistical Papers, 38*, N (3), 343-349.

Gauss, C. F. (1810). *Method des Moindres Carres Memoire sur la Combination des Observations*. Translated by J. Bertrand. (1955). Mallet-Bachelier, City Paris.

Grabski, F., & Sarhan, A. (1996). Empirical Bayes estimation in the case of exponential reliability. *Reliability Engineering and System Safty, 53*, 105-113. http://dx.doi.org/10.1016/0951-8320(96)00046-4

Jeffreys, H. (1961). *Theory of Probability* (3*rd* Edt.). Oxford University Press.

Kass, R. E., & Steffey, D. (1989). Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models). *Journal of the American Statistical Association, 84*(407), 717-726. http://dx.doi.org/10.2307/2289653

Kappenman, Russell F. (1975). Conditional confidence intervals for Double Exponential distribution parameters. *Technometrics, 17*, N (2), 233-235.

Kappenman, Russell F. (1977). Tolerance intervals for the Double Exponential distribution. *Journal of the American Statistical Association (JASA), 72*, N (360), 908-909.

Kozubowski, J. T., & Nadarajah, S. (2010). Multitude of Laplace distributions. *Stat Papers, 51*, 127-148. http://dx.doi.org/10.1007/s00362-008-0127-2

Legendre, A. (1805). Nouvelles Methodes Pour la Determination des Orbites des Cometes, Courcier, City Paris.

Martín, J., & Pérez, C. J. (2009). Bayesian analysis of a generalized lognormal distribution. *Computational Statistics and Data Analysis, 53*, 1377-1387. http://dx.doi.org/10.1016/j.csda.2008.11.029

Mendenhall, W., & Hader, A. R. (1958). Estimation of parameters of mixed exponentially distributed failure time distribution from censored life test data. *Biometrika, 45*, 1207-1212.

Nadarajah, S. (2009). Laplace random variables with application to price indices. *AStA Adv Stat Anal*, (2009) 93, 345-369. http://dx.doi.org/10.1007/s10182-009-0108-3

Nadarajah, S. (2010). Some Posterior Distributions for the Laplace Mean. *Acta Mathematica Scientia, 30B*, N(1), 330-340.

Norstrom, J. G. (1996). The use of precautionary loss functions in risk analysis. *IEEE Trans. Reliab., 45*, N (1), 400-403.

Sarhan, A. M. (2003). Empirical bayes estimates in exponential reliability model. *Applied Mathematics and Computation, 135*, 319-332. http://dx.doi.org/10.1016/S0096-3003(01)00334-4

Robbins, H. (1955). An Empirical Bayes Approach to Statistics. *Proceedings of third Berkeley Symposium on Mathematical Statistics and Probability, 1*, 157-164.

Saleem, M., & Aslam, M. (2009). On Bayesian analysis of the Rayleigh Survival Time assuming the Random censor Time. *Pak. J. Statist., 25*(2), 71-82.

Walker, S. G., Gutiérrez-Peña, E. (1999). Robustifying Bayesian procedures. In: Bernardo, J. M., Berger, J. O., Dawid, A. P., Smith, A. F. M. (Eds.), *Bayesian Statistics, 6*. Oxford University Press, 685-710.

Table 1. Properties of posterior distribution using real data set

Prior	Mean	Mode	Variance	Skewness	Excess Kurtosis
IGP	9.642985	9.080899	3.0369759	0.738423	1.050505
ICSP	9.664608	9.097899	3.002612	0.740996	1.058037
Levy	9.498996	8.948329	2.864474	0.736064	1.043623
GTII	9.360482	8.825597	2.738082	0.729917	1.025806
For Censored data, $r = 16$, $T = 10$					
IGP	17.420240	15.407660	21.904380	1.136782	2.583209
ICSP	17.571010	15.516860	21.884360	1.146198	2.628963
Levy	17.023700	15.078140	19.986650	1.128263	2.542222
GTII	16.502870	14.669220	18.156310	1.106567	2.439560

Table 2. HDR and CI using complete real data set; Parentheses contain credible intervals, Curly braces HPDR and brackets classical intervals

Prior	90%	95%	99%
IGP	(7.1894,12.7594)	(6.8420,13.5622)	(6.2258,15.3328)
	[7.1787,12.7754]	[6.8304,13.5836]	[6.2128,15.3672]
	{7.1909,12.7499}	{6.8421,13.5619}	{6.2299,15.3227}
ICSP	(7.1989,12.7989)	(6.8500,13.6071)	(6.2315,15.3902)
	[7.1787,12.7754]	[6.8304,13.5836]	[6.2128,15.3672]
	{7.1998,12.7898}	{6.8501,13.6070}	{6.2320,15.3891}
Levy	(7.0882,12.5592)	(6.7465,13.3468)	(6.1403,15.0833)
	[7.0845,12.5528]	[6.7431,13.3400]	[6.1372,15.0756]
	{7.0883,12.5522}	{6.7469,13.3386}	{6.1409,15.0732}
GT2	(7.0005,12.3509)	(6.6652,13.1188)	(6.0701,14.8105)
	[6.9929,12.3375]	[6.6580,13.1046]	[6.0635,14.7944]
	{7.0012,12.3305}	{6.6668,13.1038}	{6.0730,14.7901}
	For Censored Data		
IGP	(11.3633,25.9382)	(10.6148,28.4342)	(9.3342,34.2887)
	[11.4218,26.2865]	[10.6632,28.8462]	[9.3669,34.8632]
	{11.3695,25.9228}	{10.6170,28.4336}	{9.3350,34.2880}
ICSP	(11.4288,26.2273)	(10.6719,28.7708)	(9.3780,34.7448)
	[11.4218,26.2865]	[10.6632,28.8462]	[0.3669,34.8632]
	{11.4299,26.1867}	{10.6722,28.7700}	{9.3850,34.7431}
Levy	(11.1327,25.2910)	(10.4038,27.7074)	(9.1544,33.3686)
	[11.1313,25.2855]	[10.4016,27.7014]	[9.1524,33.3614]
	{11.1330,25.2800}	{10.4039,27.7010}	{9.1590,33.3618}
GT2	(10.8656,24.3761)	(10.1623,26.6628)	(8.9562,32.0030)
	[10.8558,24.3543]	[10.1532,26.6390]	[8.9482,31.9744]
	{10.8657,24.3559}	{10.1631,26.6321}	{8.9559,32.9609}

Table 3. Posterior probabilities under null and alternative hypotheses, Bayes factors using IG and ICS for real data

Null Hypothesis H_1	Alt. Hypothesis H_2	Prior Distribution	Posterior Probabilities $P(H_1)$	$P(H_2)$	Bayes Factor B
$\lambda \leq 6.0$	$\lambda > 6.0$	IG	0.0023779	0.997622	0.002384
		ICS	0.0023379	0.997662	0.002384
		Gumbel	0.0039733	0.996027	0.003989
		Levy	0.0031591	0.996841	0.003169
$\lambda \leq 7.5$	$\lambda > 7.5$	IG	0.083724	0.916276	0.091374
		ICS	0.082374	0.917626	0.089769
		Gumbel	0.112830	0.887170	0.127180
		Levy	0.098252	0.901748	0.108957
$\lambda \leq 9.0$	$\lambda > 9.0$	IG	0.389080	0.610920	0.636876
		ICS	0.384700	0.615300	0.625223
		Gumbel	0.455220	0.544780	0.835603
		Levy	0.422580	0.577420	0.731842
$\lambda \leq 11.0$	$\lambda > 11.0$	IG	0.803010	0.196990	4.076400
		ICS	0.799060	0.200940	3.976610
		Gumbel	0.846710	0.153290	5.523583
		Levy	0.825490	0.174510	4.730331
For Censored Data					
$\lambda \leq 15.0$	$\lambda > 15.0$	IG	0.331460	0.668540	0.4957970
		ICS	0.320480	0.679520	0.471627
		Gumbel	0.410820	0.589180	0.697274
		Levy	0.365120	0.634880	0.575101
$\lambda \leq 16.5$	$\lambda > 16.5$	IG	0.481430	0.518570	0.928380
		ICS	0.468720	0.531280	0.882247
		Gumbel	0.565640	0.434360	1.302238
		Levy	0.517500	0.482500	1.072539
$\lambda \leq 18.0$	$\lambda > 18.0$	IG	0.617240	0.382760	1.612603
		ICS	0.604510	0.395490	1.528509
		Gumbel	0.695460	0.304540	2.283641
		Levy	0.651040	0.348960	1.865658
$\lambda \leq 19.5$	$\lambda > 19.5$	IG	0.728020	0.271980	2.676741
		ICS	0.716500	0.283500	2.527337
		Gumbel	0.794250	0.205750	3.860267
		Levy	0.756870	0.243130	3.113087

Table 4. Bayes estimators and respective posterior risk under different loss function for complete data set

Posterior Distribution Using Prior	Bayes Estimator	Posterior Risk
	Squared Error Loss Function$=L_1 = L(\lambda, d) = (\lambda - d)^2$	
	Bayes Estimator $d = E(\lambda\|x)$	Posterior Risk $E_{\lambda\|x} L_1(\lambda, d) = Var(\lambda\|x)$
Inverse Gamma	$\frac{b+\sum_{i=1}^{n}\|x_i\|}{(a+n-1)}$	$\frac{\left(b+\sum_{i=1}^{n}\|x_i\|\right)^2}{(a+n-1)^2(a+n-2)}$
Inverse Chi-Squared	$\frac{\frac{b}{2}+\sum_{i=1}^{n}\|x_i\|}{\left(\frac{a}{2}+n-1\right)}$	$\frac{\left(\frac{b}{2}+\sum_{i=1}^{n}\|x_i\|\right)^2}{\left(\frac{a}{2}+n-1\right)^2\left(\frac{a}{2}+n-2\right)}$
Gumbel Type-II	$\frac{b+\sum_{i=1}^{n}\|x_i\|}{n}$	$\frac{\left(b+\sum_{i=1}^{n}\|x_i\|\right)^2}{(n)^2(n-1)}$
Levy	$\frac{\frac{b}{2}+\sum_{i=1}^{n}\|x_i\|}{\left(n-\frac{1}{2}\right)}$	$\frac{\left(\frac{b}{2}+\sum_{i=1}^{n}\|x_i\|\right)^2}{\left(n-\frac{1}{2}\right)^2\left(n-\frac{3}{2}\right)}$
	Modified /Quadratic Square Error Loss Function$=L_2 = L(\lambda, d) = \left(1 - \frac{d}{\lambda}\right)^2$	
	Bayes Estimator $d = \frac{E(\lambda^{-1}\|x)}{E(\lambda^{-2}\|x)}$	Posterior Risk $E_{\lambda\|x} L_2(\lambda, d) = 1 - \frac{\left[E(\lambda^{-1}\|x)\right]^2}{E(\lambda^{-2}\|x)}$
Inverse Gamma	$\frac{b+\sum_{i=1}^{n}\|x_i\|}{a+n+1}$	$\frac{1}{a+n+1}$
Inverse Chi-Squared	$\frac{\frac{b}{2}+\sum_{i=1}^{n}\|x_i\|}{\left(\frac{a}{2}+n+1\right)}$	$\frac{1}{\left(\frac{a}{2}+n+1\right)}$
Gumbel Type-II	$\frac{b+\sum_{i=1}^{n}\|x_i\|}{(2+n)}$	$\frac{1}{(2+n)}$
Levy	$\frac{\frac{b}{2}+\sum_{i=1}^{n}\|x_i\|}{\left(\frac{3}{2}+n\right)}$	$\frac{1}{\left(\frac{3}{2}+n\right)}$
	Weighted Loss Function$=L_3 = L(\lambda, d) = \frac{(\lambda-d)^2}{\lambda}$	
	Bayes Estimator $d = \frac{1}{E(\lambda^{-1}\|x)}$	Posterior Risk $E_{\lambda\|x} L_3(\lambda, d) = E(\lambda\|x) - \frac{1}{E(\lambda^{-1}\|x)}$
Inverse Gamma	$\frac{b+\sum_{i=1}^{n}\|x_i\|}{(a+n)}$	$\frac{b+\sum_{i=1}^{n}\|x_i\|}{(a+n)(a+n-1)}$
Inverse Chi-Squared	$\frac{\frac{b}{2}+\sum_{i=1}^{n}\|x_i\|}{\left(\frac{a}{2}+n\right)}$	$\frac{\frac{b}{2}+\sum_{i=1}^{n}\|x_i\|}{\left(\frac{a}{2}+n\right)\left(\frac{a}{2}+n-1\right)}$
Gumbel Type-II	$\frac{b+\sum_{i=1}^{n}\|x_i\|}{(n+1)}$	$\frac{b+\sum_{i=1}^{n}\|x_i\|}{n(n+1)}$
Levy	$\frac{\frac{b}{2}+\sum_{i=1}^{n}\|x_i\|}{\left(\frac{1}{2}+n\right)}$	$\frac{\frac{b}{2}+\sum_{i=1}^{n}\|x_i\|}{\left(\frac{1}{2}+n\right)\left(n-\frac{1}{2}\right)}$
	Precautionary Loss Function$=L_4 = L(\lambda, d) = \frac{(\lambda-d)^2}{d}$	
	Bayes Estimator $d = \sqrt{E(\lambda^2\|x)}$	Posterior Risk $E_{\lambda\|x} L_4(\lambda, d) = 2\left(\sqrt{E(\lambda^2\|x)} - E(\lambda\|x)\right)$
Inverse Gamma	$\frac{b+\sum_{i=1}^{n}\|x_i\|}{\sqrt{(a+n-1)(a+n-2)}}$	$2\left(\frac{b+\sum_{i=1}^{n}\|x_i\|}{\sqrt{(a+n-1)(a+n-2)}} - \frac{b+\sum_{i=1}^{n}\|x_i\|}{(a+n-1)}\right)$
Inverse Chi-Squared	$\frac{\frac{b}{2}+\sum_{i=1}^{n}\|x_i\|}{\sqrt{\left(\frac{a}{2}+n-1\right)\left(\frac{a}{2}+n-2\right)}}$	$2\left(\frac{\left(\frac{b}{2}+\sum_{i=1}^{n}\|x_i\|\right)}{\sqrt{\left(\frac{a}{2}+n-1\right)\left(\frac{a}{2}+n-2\right)}} - \frac{\left(\frac{b}{2}+\sum_{i=1}^{n}\|x_i\|\right)}{\left(\frac{a}{2}+n-1\right)}\right)$
Gumbel Type-II	$\frac{b+\sum_{i=1}^{n}\|x_i\|}{\sqrt{(n)(n-1)}}$	$2\left(\frac{b+\sum_{i=1}^{n}\|x_i\|}{\sqrt{(n)(n-1)}} - \frac{b+\sum_{i=1}^{n}\|x_i\|}{n}\right)$
Levy	$\frac{\frac{b}{2}+\sum_{i=1}^{n}\|x_i\|}{\sqrt{\left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)}}$	$2\left(\frac{\left(\frac{b}{2}+\sum_{i=1}^{n}\|x_i\|\right)}{\sqrt{\left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)}} - \frac{\frac{b}{2}+\sum_{i=1}^{n}\|x_i\|}{\left(n-\frac{1}{2}\right)}\right)$

Table 5. BEs and PRs using IC and ICS priors under different LFs

	Complete Data		Censored Data		Complete Data		Censored Data	
LF	IG	ICS	IG	ICS	Levy	GTII	Levy	GTII
L_1	9.6430	9.6502	17.420237	17.571007	9.5009	9.3568	17.023701	16.502870
	(2.969753)	(3.002612)	(21.9044)	(21.8844)	(2.865598)	(2.735934)	(19.9866)	(18.1563)
L_2	9.0809	9.0979	15.407665	15.516856	8.9483	8.8256	15.078135	14.669217
	(0.029345)	(0.029319)	(0.058765)	(0.058453)	(0.028986)	(0.028571)	(0.057143)	(0.055556)
L_3	9.3535	9.3727	16.352259	16.531939	9.2154	9.0852	15.991961	15.532112
	(0.299480)	(0.291913)	(1.097978)	(1.097931)	(0.283552)	(0.275308)	(1.031739)	(0.970757)
L_4	9.7958	9.804538	18.018575	18.183088	9.650528	9.501876	17.600937	17.044090
	(0.308550)	(0.308677)	(1.196677)	(1.224162)	(0.299257)	(0.290151)	(1.154473)	(1.082441)

Table 6. Posterior Distribution Properties via IGP for complete data

n	λ=0.5			λ=1.0		
	Mean	Mode	Variance	Mean	Mode	Variance
25	0.634249	0.586038	0.014256	1.142528	1.055682	0.055997
50	0.482757	0.464850	0.004577	0.959504	0.923913	0.018081
100	0.505960	0.495941	0.002612	1.008485	0.988515	0.010378
500	0.500441	0.498443	0.000503	1.002525	0.998523	0.002018
1000	0.499920	0.498921	0.000250	1.000841	0.998841	0.001004
n	λ=3.0			λ=4.0		
	Mean	Mode	Variance	Mean	Mode	Variance
25	3.206086	2.962383	0.440942	4.249382	3.926375	0.774610
50	2.893189	2.785872	0.164391	3.870105	3.726552	0.294151
100	3.027071	2.967129	0.093502	4.025859	3.946139	0.165383
500	3.003206	2.991218	0.018111	4.004609	3.988623	0.032203
1000	3.004404	2.998402	0.009045	4.000901	3.992907	0.016039

Table 7. Posterior distribution properties via ICSP for complete data

n	λ=0.5			λ=1.0		
	Mean	Mode	Variance	Mean	Mode	Variance
25	0.586962	0.541998	0.014909	1.099535	1.015304	0.052319
50	0.494853	0.476055	0.004932	0.983546	0.946184	0.019484
100	0.505960	0.495941	0.002612	1.008485	0.988515	0.010378
500	0.500441	0.498443	0.000503	1.002525	0.998523	0.002018
1000	0.499920	0.498921	0.000250	1.000841	0.998841	0.001004
n	λ=3.0			λ=4.0		
	Mean	Mode	Variance	Mean	Mode	Variance
25	3.180520	2.936875	0.437762	4.232628	3.908385	0.775285
50	2.965684	2.853026	0.177148	3.967078	3.816380	0.316977
100	3.027071	2.967129	0.093502	4.025859	3.946139	0.165383
500	3.003206	2.991218	0.018111	4.004609	3.988623	0.032203
1000	3.004404	2.998402	0.009045	4.000901	3.992907	0.016039

Table 8. Posterior distribution properties via IGP for censored data

n	λ=0.5			λ=1.0		
	Mean	Mode	Variance	Mean	Mode	Variance
25	0.625303	0.571655	0.019250	1.210194	1.070379	0.102336
50	0.584629	0.558242	0.008274	1.089618	1.024198	0.039169
100	0.545618	0.533120	0.003531	1.064532	1.031426	0.018483
500	0.508678	0.506330	0.000601	1.012137	1.005758	0.003259
1000	0.503393	0.502231	0.000294	1.006078	1.002901	0.001606
n	λ=3.0			λ=4.0		
	Mean	Mode	Variance	Mean	Mode	Variance
25	3.079968	2.815723	0.467038	4.529387	4.083393	0.211624
50	3.168282	3.025281	0.242984	4.106361	3.902629	0.086218
100	3.101677	3.030629	0.114106	4.159384	4.053157	0.043284
200	3.025270	2.990558	0.053425	4.063137	4.011150	0.020533
300	3.032845	3.009543	0.035747	4.028638	3.994250	0.013466
500	3.019045	3.005110	0.021181	4.024768	4.004091	0.008063
1000	3.003221	2.996288	0.010447	4.006019	3.995725	0.003989

Table 9. Posterior distribution properties via ICSP for censored data

n	λ=0.5			λ=1.0		
	Mean	Mode	Variance	Mean	Mode	Variance
25	0.571209	0.519347	0.016227	1.142500	1.002617	0.092524
50	0.557316	0.530748	0.007556	1.055953	0.988944	0.037035
100	0.532012	0.519154	0.003365	1.047588	1.013202	0.017959
500	0.505974	0.503512	0.000595	1.008764	1.002061	0.003240
1000	0.502043	0.500821	0.000292	1.004392	1.001048	0.001601
n	λ=3.0			λ=4.0		
	Mean	Mode	Variance	Mean	Mode	Variance
25	3.049553	2.772676	0.462496	4.510230	4.039965	0.209604
50	3.153462	3.003133	0.241908	4.094997	3.880368	0.085695
100	3.094186	3.019406	0.113830	4.153836	4.041896	0.043155
500	3.017527	3.002845	0.021170	4.023606	4.001816	0.008058
1000	3.002460	2.995154	0.010445	4.005435	3.994586	0.003988

Table 10. Posterior distribution properties via LP for complete data

n	λ=0.5			λ=1.0		
	Mean	Mode	Variance	Mean	Mode	Variance
25	0.512586	0.473901	0.011181	1.016954	0.940202	0.044008
50	0.506343	0.486680	0.005286	1.006384	0.967301	0.020883
100	0.503417	0.493498	0.002573	1.003417	0.983645	0.010222
500	0.499940	0.497946	0.000501	1.001522	0.997527	0.002012
1000	0.499670	0.498672	0.000250	1.000340	0.998342	0.001002
n	λ=3.0			λ=4.0		
	Mean	Mode	Variance	Mean	Mode	Variance
25	3.064627	2.833334	0.399657	4.099892	3.790467	0.715282
50	3.034545	2.916699	0.189865	4.059192	3.901553	0.339733
100	3.011859	2.952512	0.092094	4.005628	3.926700	0.162894
500	3.000200	2.988235	0.018057	4.000601	3.984646	0.032106
1000	3.002901	2.996905	0.009031	3.998899	3.990914	0.016015

Table 11. Posterior distribution properties via GTIIP for complete data

n	λ=0.5			λ=1.0		
	Mean	Mode	Variance	Mean	Mode	Variance
25	0.509476	0.471737	0.010815	1.003756	0.929404	0.041980
50	0.501280	0.482000	0.005128	0.996320	0.958000	0.020258
100	0.500900	0.491078	0.002534	0.998400	0.978824	0.010069
500	0.499440	0.497450	0.000500	1.000520	0.996534	0.002006
1000	0.499420	0.498423	0.000250	0.999840	0.997844	0.001001
n	λ=3.0			λ=4.0		
	Mean	Mode	Variance	Mean	Mode	Variance
25	3.010476	2.787478	0.377624	4.025036	3.726886	0.675038
50	3.004200	2.888654	0.184188	4.018600	3.864038	0.329574
100	2.996800	2.938039	0.090715	3.985600	3.907451	0.160455
500	2.997200	2.985259	0.018002	3.996600	3.980677	0.032010
1000	3.001400	2.995409	0.009017	3.996900	3.988922	0.015991

Table 12. Posterior distribution properties via LP for censored data

n	λ=0.5			λ=1.0		
	Mean	Mode	Variance	Mean	Mode	Variance
25	0.486743	0.445318	0.011557	1.010882	0.895352	0.070475
50	0.514714	0.491581	0.006384	0.992266	0.933026	0.032282
100	0.510952	0.499273	0.003090	1.015542	0.984052	0.016770
500	0.501825	0.499509	0.000585	1.002464	0.996149	0.003195
1000	0.499974	0.498820	0.000289	1.001247	0.998086	0.001590
n	λ=3.0			λ=4.0		
	Mean	Mode	Variance	Mean	Mode	Variance
25	2.919877	2.671377	0.415887	4.328560	3.906262	1.070653
50	3.086903	2.948165	0.229614	4.011932	3.813811	0.429216
100	3.061373	2.991398	0.110911	4.111730	4.006973	0.223925
500	3.011095	2.997203	0.021061	4.015425	3.994807	0.041717
1000	2.999257	2.992334	0.010418	4.001362	3.991082	0.020646

Table 13. Posterior distribution properties via GTIIP for censored data

n	λ=0.5			λ=1.0		
	Mean	Mode	Variance	Mean	Mode	Variance
25	0.483796	0.443480	0.011146	0.990451	0.880401	0.065400
50	0.512882	0.490087	0.006263	0.982341	0.924556	0.031129
100	0.510057	0.498465	0.003061	1.010316	0.979229	0.016464
500	0.501657	0.499345	0.000584	1.001443	0.995144	0.003184
1000	0.499891	0.498738	0.000289	1.000737	0.997580	0.001587
n	λ=3.0			λ=4.0		
	Mean	Mode	Variance	Mean	Mode	Variance
25	2.861632	2.623163	0.389950	4.224048	3.821758	0.991255
50	3.055161	2.919376	0.222238	3.965075	3.771656	0.413732
100	3.045650	2.976431	0.109129	4.087350	3.983872	0.219821
500	3.008023	2.994161	0.020994	4.010711	3.990143	0.041565
1000	2.997729	2.990814	0.010401	3.999017	3.988750	0.020608

Table 14. Skewness and excess kurtosis of posterior distribution assuming different prior for complete data

Prior	IGP		ICSP		LP		GTIIP	
n	Skewness	Excess Kurtosis	Skewness	Excess Kurtosis	Skewness	Excess Kurtosis	Skewness	Excess Kurtosis
25	0.912720	1.622659	0.898291	1.571810	0.853655	1.417511	0.838324	1.365766
50	0.555161	0.587239	0.569644	0.618607	0.583428	0.649236	0.577350	0.635638
100	0.408227	0.315077	0.408227	0.315077	0.406133	0.311837	0.404061	0.308647
200	0.179605	0.060581	0.179605	0.060581	0.284987	0.152904	0.284268	0.152130
300	0.126745	0.030145	0.126745	0.030145	0.232103	0.101282	0.231714	0.100942
500	0.912720	1.622659	0.898291	1.571810	0.179424	0.060459	0.179244	0.060338
1000	0.555161	0.587239	0.569644	0.618607	0.126681	0.030114	0.126618	0.030084

Table 15. Skewness and excess kurtosis of posterior distribution assuming IG for censored data

Prior	IG				ICSP			
n	λ=0.5		λ=1.0		λ=0.5		λ=1.0	
	Skewness	Excess Kurtosis	Skewness	Excess Kurtosis	Skewness	Excess Kurtosis	Skewness	Excess Kurtosis
25	0.933504	1.706192	1.136782	2.583209	0.938710	1.726108	1.146198	2.628963
50	0.637774	0.778282	0.751322	1.088564	0.639430	0.782411	0.754033	1.096653
100	0.440858	0.367969	0.519315	0.512514	0.441404	0.368890	0.520209	0.514302
500	0.193276	0.070173	0.226341	0.096303	0.193322	0.070206	0.226415	0.096366
1000	0.136295	0.034863	0.159578	0.047808	0.136311	0.034871	0.159603	0.047823
n	λ=3.0		λ=4.0		λ=3.0		λ=4.0	
25	0.933504	1.706192	1.020316	2.055418	0.938710	1.726108	1.027119	2.084350
50	0.637774	0.778282	0.672880	0.868301	0.639430	0.782411	0.674826	0.873443
100	0.440858	0.367969	0.467127	0.413619	0.441404	0.368890	0.467777	0.414783
500	0.193276	0.070173	0.204041	0.078224	0.193322	0.070206	0.204095	0.078265
1000	0.136295	0.034863	0.143841	0.038834	0.136311	0.034871	0.143860	0.038845

Table 16. Skewness and excess kurtosis of posterior distribution assuming LP for censored data

Prior	LP				GTII			
n	λ=0.5		λ=1.0		λ=0.5		λ=1.0	
	Skewness	Excess Kurtosis	Skewness	Excess Kurtosis	Skewness	Excess Kurtosis	Skewness	Excess Kurtosis
25	0.928757	1.688150	1.128263	2.542222	0.916515	1.642105	1.106567	2.439560
50	0.636252	0.774496	0.748838	1.081178	0.632267	0.764634	0.742369	1.062069
100	0.440354	0.367120	0.518492	0.510869	0.439026	0.364888	0.516328	0.506557
500	0.193234	0.070142	0.226273	0.096245	0.193121	0.070060	0.226092	0.096091
1000	0.136280	0.034855	0.159554	0.047793	0.136240	0.034835	0.159490	0.047755
n	λ=3.0		λ=4.0		λ=3.0	λ=4.0	λ=3.0	λ=4.0
25	0.928757	1.688150	1.014133	2.029326	0.916515	1.642105	0.998268	1.963235
50	0.636252	0.774496	0.671093	0.863593	0.632267	0.764634	0.666423	0.851351
100	0.440354	0.367120	0.466527	0.412546	0.439026	0.364888	0.464949	0.409730
500	0.193234	0.070142	0.203991	0.078185	0.193121	0.070060	0.203858	0.078084
1000	0.136280	0.034855	0.143823	0.038825	0.136240	0.034835	0.143777	0.038800

Table 17. BEs and PRs using IGP under L_1

Prior	IGP				ICSP			
n	λ=0.5	λ=1	λ=3	λ=4	λ=0.5	λ=1	λ=3	λ=4
25	0.634249	1.142528	3.206086	4.249382	0.586962	1.099535	3.180520	4.232628
	(0.017490)	(0.056755)	(0.446912)	(0.785098)	(0.014909)	(0.052319)	(0.437762)	(0.775285)
50	0.567465	1.069417	3.105335	4.133900	0.543974	1.048008	3.092367	4.125197
	(0.006709)	(0.023826)	(0.200898)	(0.356024)	(0.006151)	(0.022830)	(0.198777)	(0.353732)
100	0.533760	1.034710	3.046966	4.042622	0.522051	1.024030	3.040420	4.038122
	(0.002907)	(0.010925)	(0.094735)	(0.166763)	(0.002778)	(0.010689)	(0.094224)	(0.166209)
500	0.505974	1.007745	3.007178	4.007957	0.503637	1.005613	3.005862	4.007049
	(0.000514)	(0.002039)	(0.018159)	(0.032256)	(0.000509)	(0.002030)	(0.018139)	(0.032235)
1000	0.502685	1.003449	3.006389	4.002575	0.501517	1.002384	3.005731	4.002120
	(0.000253)	(0.001009)	(0.009056)	(0.016053)	(0.000252)	(0.001007)	(0.009052)	(0.016047)
For Censored Data								
25	0.625303	1.210194	3.079968	4.529387	0.571209	1.142500	3.049553	4.510230
	(0.019250)	(0.102336)	(0.277969)	(0.482817)	(0.016227)	(0.092524)	(0.272060)	(0.475945)
50	0.584629	1.089618	3.168282	4.106361	0.557316	1.055953	3.153462	4.094997
	(0.008274)	(0.039169)	(0.132667)	(0.191173)	(0.007556)	(0.037035)	(0.131189)	(0.189492)
100	0.545618	1.064532	3.101677	4.159384	0.532012	1.047588	3.094186	4.153836
	(0.003531)	(0.018483)	(0.061720)	(0.093415)	(0.003365)	(0.017959)	(0.061357)	(0.092993)
500	0.508678	1.012137	3.019045	4.024768	0.505974	1.008764	3.017527	4.023606
	(0.000601)	(0.003259)	(0.011564)	(0.017216)	(0.000595)	(0.003240)	(0.011550)	(0.017199)
1000	0.503393	1.006078	3.003221	4.006019	0.502043	1.004392	3.002460	4.005435
	(0.000294)	(0.001606)	(0.005720)	(0.008517)	(0.000292)	(0.001601)	(0.005717)	(0.008513)

Table 18. BEs and PRs using IGP under L_2

Prior	IGP				ICSP			
n	λ=0.5	λ=1	λ=3	λ=4	λ=0.5	λ=1	λ=3	λ=4
25	0.586038	1.055682	2.962383	3.926375	0.541998	1.015304	2.936875	3.908385
	(0.038006)	(0.038006)	(0.038006)	(0.038006)	(0.038303)	(0.038303)	(0.038303)	(0.038303)
50	0.545346	1.027734	2.984297	3.972771	0.522687	1.006996	2.971354	3.963766
	(0.019489)	(0.019489)	(0.019489)	(0.019489)	(0.019566)	(0.019566)	(0.019566)	(0.019566)
100	0.523223	1.014283	2.986815	3.962816	0.511725	1.003774	2.980278	3.958244
	(0.009871)	(0.009871)	(0.009871)	(0.009871)	(0.009890)	(0.009890)	(0.009890)	(0.009890)
500	0.503955	1.003724	2.995181	3.991967	0.501627	1.001600	2.993865	3.991056
	(0.001995)	(0.001995)	(0.001995)	(0.001995)	(0.001996)	(0.001996)	(0.001996)	(0.001996)
1000	0.501681	1.001445	3.000384	3.994580	0.500515	1.000381	2.999726	3.994125
	(0.000999)	(0.000999)	(0.000999)	(0.000999)	(0.000999)	(0.000999)	(0.000999)	(0.000999)
For Censored Data								
25	0.571655	1.070379	2.815723	4.083393	2.785611	4.061625	2.785611	4.061625
	(0.042897)	(0.057765)	(0.042897)	(0.049233)	(0.043275)	(0.049732)	(0.043275)	(0.049732)
50	0.558242	1.024198	3.025281	3.902629	3.010473	3.890798	3.010473	3.890798
	(0.022568)	(0.030020)	(0.022568)	(0.024807)	(0.022672)	(0.024933)	(0.022672)	(0.024933)
100	0.533120	1.031426	3.030629	4.053157	3.023143	4.047475	3.023143	4.047475
	(0.011453)	(0.015549)	(0.011453)	(0.012770)	(0.011480)	(0.012803)	(0.011480)	(0.012803)
500	0.506330	1.005758	3.005110	4.004091	3.003593	4.002925	3.003593	4.002925
	(0.002308)	(0.003151)	(0.002308)	(0.002569)	(0.002309)	(0.002570)	(0.002309)	(0.002570)
1000	0.502231	1.002901	2.996288	3.995725	2.995527	3.995140	2.995527	3.995140
	(0.001154)	(0.001579)	(0.001154)	(0.001285)	(0.001155)	(0.001285)	(0.001155)	(0.001285)

Table 19. BEs and PRs using IGP under L_3

Prior	IGP				ICSP			
n	λ=0.5	λ=1	λ=3	λ=4	λ=0.5	λ=1	λ=3	λ=4
25	0.609191	1.097390	3.079420	4.081498	0.563584	1.055742	3.053846	4.064050
	(0.025058)	(0.045139)	(0.126666)	(0.167884)	(0.023378)	(0.043793)	(0.126675)	(0.168578)
50	0.556186	1.048161	3.043613	4.051734	0.533118	1.027093	3.030653	4.042870
	(0.011279)	(0.021256)	(0.061722)	(0.082166)	(0.010856)	(0.020915)	(0.061714)	(0.082326)
100	0.528439	1.024395	3.016591	4.002321	0.516836	1.013801	3.010049	3.997784
	(0.005321)	(0.010315)	(0.030375)	(0.040301)	(0.005215)	(0.010229)	(0.030371)	(0.040338)
500	0.504962	1.005731	3.001168	3.999946	0.502630	1.003602	2.999852	3.999036
	(0.001011)	(0.002014)	(0.006011)	(0.008011)	(0.001007)	(0.002011)	(0.006010)	(0.008012)
1000	0.502182	1.002446	3.003383	3.998573	0.501015	1.001381	3.002726	3.998118
	(0.000503)	(0.001003)	(0.003005)	(0.004001)	(0.000501)	(0.001002)	(0.003005)	(0.004002)
For Censored Data								
25	0.597276	1.136001	2.941924	4.294842	0.545371	1.071572	2.911613	4.274189
	(0.028026)	(0.074193)	(0.106498)	(0.149708)	(0.025837)	(0.070928)	(0.105796)	(0.149336)
50	0.571131	1.055896	3.095131	4.001903	0.544388	1.023065	3.080309	3.990287
	(0.013498)	(0.033722)	(0.054052)	(0.067938)	(0.012928)	(0.032888)	(0.053871)	(0.067805)
100	0.539297	1.047718	3.065741	4.105584	0.525833	1.030988	3.058252	4.099966
	(0.006322)	(0.016814)	(0.026429)	(0.034308)	(0.006178)	(0.016600)	(0.026382)	(0.034274)
500	0.507501	1.008938	3.012061	4.014403	0.504803	1.005573	3.010544	4.013239
	(0.001177)	(0.003200)	(0.005160)	(0.006641)	(0.001171)	(0.003191)	(0.005158)	(0.006640)
1000	0.502811	1.004487	2.999751	4.000866	0.501463	1.002803	2.998990	4.000281
	(0.000582)	(0.001591)	(0.002568)	(0.003306)	(0.000580)	(0.001589)	(0.002568)	(0.003306)
25	0.597276	1.136001	2.941924	4.294842	0.545371	1.071572	2.911613	4.274189
	(0.028026)	(0.074193)	(0.106498)	(0.149708)	(0.025837)	(0.070928)	(0.105796)	(0.149336)

Table 20. BEs and PRs using IGP under L_4

Prior	IGP				ICSP			
n	λ=0.5	λ=1	λ=3	λ=4	0.599528	1.123074	3.248610	4.323242
					(0.025131)	(0.047079)	(0.136181)	(0.181229)
25	0.647890	1.167101	3.275042	4.340777	0.549599	1.058844	3.124341	4.167851
	(0.027282)	(0.049146)	(0.137912)	(0.182790)	(0.011249)	(0.021672)	(0.063949)	(0.085308)
50	0.573346	1.080499	3.137515	4.176739	0.524705	1.029236	3.055876	4.058650
	(0.011762)	(0.022164)	(0.064361)	(0.085679)	(0.005308)	(0.010412)	(0.030912)	(0.041056)
100	0.536476	1.039976	3.062472	4.063195	0.504142	1.006622	3.008878	4.011069
	(0.005432)	(0.010532)	(0.031013)	(0.041146)	(0.001010)	(0.002018)	(0.006031)	(0.008040)
500	0.506482	1.008756	3.010196	4.011979	0.501768	1.002886	3.007236	4.004124
	(0.001016)	(0.002022)	(0.006035)	(0.008044)	(0.000502)	(0.001004)	(0.003011)	(0.004009)
1000	0.502936	1.003952	3.007895	4.004580	0.599528	1.123074	3.248610	4.323242
	(0.000503)	(0.001005)	(0.003011)	(0.004009)	(0.025131)	(0.047079)	(0.136181)	(0.181229)
For Censored Data								
25	0.640511	1.251761	3.124767	4.582375	0.585241	1.182299	3.093838	4.562688
	(0.030415)	(0.083134)	(0.089599)	(0.105977)	(0.028064)	(0.079597)	(0.088569)	(0.104915)
50	0.591663	1.107446	3.189150	4.129573	0.564054	1.073346	3.174195	4.118069
	(0.014068)	(0.035656)	(0.041736)	(0.046424)	(0.013476)	(0.034786)	(0.041465)	(0.046144)
100	0.548844	1.073178	3.111610	4.170598	0.535165	1.056125	3.104085	4.165014
	(0.006452)	(0.017292)	(0.019867)	(0.022429)	(0.006306)	(0.017074)	(0.019798)	(0.022357)
500	0.509268	1.013746	3.020959	4.026906	0.506562	1.010369	3.019440	4.025743
	(0.001181)	(0.003217)	(0.003829)	(0.004276)	(0.001175)	(0.003209)	(0.003826)	(0.004273)
1000	0.503685	1.006876	3.004173	4.007082	0.502334	1.005189	3.003412	4.006497
	(0.000584)	(0.001596)	(0.001904)	(0.002126)	(0.000581)	(0.001593)	(0.001904)	(0.002125)

Table 21. BEs and PRs using LP under L_1

Prior	LP				GTII			
n	λ=0.5	λ=1	λ=3	λ=4	λ=0.5	λ=1	λ=3	λ=4
25	0.512586	1.016954	3.064627	4.099892	0.509476	1.003756	3.010476	4.025036
	(0.011181)	(0.044008)	(0.399657)	(0.715282)	(0.010616)	(0.041587)	(0.376441)	(0.673457)
50	0.507502	1.007543	3.035704	4.060351	0.505998	1.001038	3.008918	4.023318
	(0.005310)	(0.020931)	(0.190010)	(0.339927)	(0.005177)	(0.020354)	(0.184478)	(0.329961)
100	0.503994	1.003994	3.012436	4.006205	0.503259	1.000759	2.999159	3.987959
	(0.002579)	(0.010234)	(0.092130)	(0.162941)	(0.002546)	(0.010093)	(0.090787)	(0.160550)
500	0.500055	1.001636	3.000315	4.000715	0.499912	1.000992	2.997672	3.997072
	(0.000502)	(0.002013)	(0.018058)	(0.032108)	(0.000500)	(0.002007)	(0.018005)	(0.031013)
1000	0.499727	1.000398	3.002959	3.998957	0.499656	1.000076	3.001636	3.997136
	(0.000250)	(0.001002)	(0.009031)	(0.016016)	(0.000250)	(0.001001)	(0.009018)	(0.015992)
For Censored Data								
25	0.486743	1.010882	2.919877	4.328560	0.483796	0.990451	2.861632	4.224048
	(0.011577)	(0.070475)	(0.240887)	(0.418715)	(0.011146)	(0.065400)	(0.226266)	(0.388698)
50	0.514714	0.992266	3.086903	4.011932	0.512882	0.982341	3.055161	3.965075
	(0.006384)	(0.032282)	(0.123442)	(0.177967)	(0.006263)	(0.031129)	(0.119595)	(0.171766)
100	0.510952	1.015542	3.061373	4.111730	0.510057	1.010316	3.045650	4.087350
	(0.003090)	(0.016770)	(0.059521)	(0.090120)	(0.003061)	(0.016464)	(0.058594)	(0.088527)
500	0.501825	1.002464	3.011095	4.015425	0.501657	1.001443	3.008023	4.010711
	(0.000585)	(0.003195)	(0.011481)	(0.017093)	(0.000584)	(0.003184)	(0.011445)	(0.017033)
1000	0.499974	1.001247	2.999257	4.001362	0.499891	1.000737	2.997729	3.999017
	(0.000289)	(0.001590)	(0.005700)	(0.008487)	(0.000289)	(0.001587)	(0.005691)	(0.008472)

Table 22. BEs and PRs using LP under L_2

Prior	LP				GTII			
n	λ=0.5	λ=1	λ=3	λ=4	λ=0.5	λ=1	λ=3	λ=4
25	0.473901	0.940202	2.833334	3.790467	0.471737	0.929404	2.787478	3.726886
	(0.037736)	(0.037736)	(0.037736)	(0.037736)	(0.037037)	(0.037037)	(0.037037)	(0.037037)
50	0.487793	0.968415	2.917813	3.902667	0.486537	0.962537	2.893191	3.868575
	(0.019417)	(0.019417)	(0.019417)	(0.019417)	(0.019231)	(0.019231)	(0.019231)	(0.019231)
100	0.494063	0.984210	2.953077	3.927265	0.493391	0.981136	2.940352	3.909764
	(0.009852)	(0.009852)	(0.009852)	(0.009852)	(0.009804)	(0.009804)	(0.009804)	(0.009804)
500	0.498061	0.997642	2.988350	3.984760	0.497920	0.997004	2.985729	3.981147
	(0.001994)	(0.001994)	(0.001994)	(0.001994)	(0.001992)	(0.001992)	(0.001992)	(0.001992)
1000	0.498729	0.998400	2.996962	3.990971	0.498659	0.99808	2.995645	3.989158
	(0.000999)	(0.000999)	(0.000999)	(0.000999)	(0.000998)	(0.000998)	(0.000998)	(0.000998)
For Censored Data								
25	0.445318	0.895352	2.671377	3.906262	0.443480	0.880401	2.623163	3.821758
	(0.042553)	(0.057143)	(0.042553)	(0.048780)	(0.041667)	(0.055556)	(0.041667)	(0.047619)
50	0.491581	0.933026	2.948165	3.813811	0.490087	0.924556	2.919376	3.771656
	(0.022472)	(0.029851)	(0.022472)	(0.024691)	(0.022222)	(0.029412)	(0.022222)	(0.024390)
100	0.499273	0.984052	2.991398	4.006973	0.498465	0.979229	2.976431	3.983872
	(0.011429)	(0.015504)	(0.011429)	(0.012739)	(0.011364)	(0.015385)	(0.011364)	(0.012658)
500	0.499509	0.996149	2.997203	3.994807	0.499345	0.995144	2.994161	3.990143
	(0.002307)	(0.003150)	(0.002307)	(0.002567)	(0.002304)	(0.003145)	(0.002304)	(0.002564)
1000	0.498820	0.998086	2.992334	3.991082	0.498738	0.997580	2.990814	3.988750
	(0.002307)	(0.003150)	(0.002307)	(0.002567)	(0.001153)	(0.001577)	(0.001153)	(0.001284)

Table 23. BEs and PRs using LP under L_3

Prior	LP				GTII			
n	λ=0.5	λ=1	λ=3	λ=4	λ=0.5	λ=1	λ=3	λ=4
25	0.492485	0.977073	2.944446	3.939112	0.489881	0.965150	2.894689	3.870227
	(0.020101)	(0.039881)	(0.120181)	(0.160780)	(0.019595)	(0.038606)	(0.115788)	(0.154809)
50	0.497453	0.987591	2.975591	3.979948	0.496077	0.981410	2.949920	3.944430
	(0.010050)	(0.019951)	(0.060113)	(0.080403)	(0.009922)	(0.019628)	(0.058998)	(0.078889)
100	0.498979	0.994004	2.982461	3.966342	0.498276	0.990851	2.969464	3.948474
	(0.005015)	(0.009990)	(0.029974)	(0.039863)	(0.004983)	(0.009909)	(0.029695)	(0.039485)
500	0.499056	0.999635	2.994320	3.992722	0.498914	0.998994	2.991688	3.989094
	(0.000999)	(0.002001)	(0.005995)	(0.007993)	(0.000998)	(0.001998)	(0.005983)	(0.007978)
1000	0.499228	0.999398	2.999957	3.994960	0.499157	0.999077	2.998637	3.993143
	(0.000499)	(0.001000)	(0.003001)	(0.003997)	(0.000499)	(0.000999)	0.002999)	(0.003993)
For Censored Data								
25	0.465110	0.949616	2.790105	4.106583	2.737214	4.012846	2.737214	4.012846
	(0.021633)	(0.061266)	(0.098765)	(0.138817)	(0.094775)	(0.132255)	(0.094775)	(0.132255)
50	0.502882	0.961734	3.015939	3.910364	2.985725	3.865948	2.985725	3.865948
	(0.011833)	(0.030531)	(0.052032)	(0.065402)	(0.050937)	(0.063870)	(0.050937)	(0.063870)
100	0.505045	0.999549	3.025981	4.058676	3.010643	4.034948	3.010643	4.034948
	(0.005907)	(0.015993)	(0.025927)	(0.033658)	(0.025652)	(0.033254)	(0.025652)	(0.033254)
500	0.500664	0.999297	3.004133	4.005090	3.001076	4.000401	3.001076	4.000401
	(0.001160)	(0.003167)	(0.005140)	(0.006616)	(0.005129)	(0.006600)	(0.005129)	(0.006600)
1000	0.499396	0.999664	2.995791	3.996215	2.994268	3.993876	2.994268	3.993876
	(0.000578)	(0.001583)	(0.002563)	(0.003300)	(0.002560)	(0.003296)	(0.002560)	(0.003296)
25	0.465110	0.949616	2.790105	4.106583	2.737214	4.012846	2.737214	4.012846
	(0.021633)	(0.061266)	(0.098765)	(0.138817)	(0.094775)	(0.132255)	(0.094775)	(0.132255)

Table 24. BEs and PRs using LP under L_4

Prior	LP				GTII			
n	λ=0.5	λ=1	λ=3	λ=4	λ=0.5	λ=1	λ=3	λ=4
25	0.523379	1.038366	3.129152	4.186215	0.519790	1.024262	3.072362	4.107843
	(0.021586)	(0.042823)	(0.129051)	(0.172646)	(0.020628)	(0.041012)	(0.123771)	(0.165613)
50	0.512707	1.017877	3.066840	4.101997	0.511088	1.011153	3.039419	4.064117
	(0.017426)	(0.020668)	(0.062272)	(0.083291)	(0.010180)	(0.020231)	(0.061001)	(0.081598)
100	0.506546	1.009078	3.027689	4.026490	0.505782	1.005789	3.014256	4.008038
	(0.005104)	(0.010167)	(0.030506)	(0.040569)	(0.005046)	(0.010060)	(0.030195)	(0.040157)
500	0.500557	1.002640	3.003323	4.004726	0.500412	1.001994	3.000674	4.000949
	(0.001003)	(0.002009)	(0.006016)	(0.008021)	(0.000999)	(0.002004)	(0.006003)	(0.007755)
1000	0.499977	1.000899	3.004462	4.000954	0.499906	1.000576	3.003138	3.999136
	(0.000500)	(1.001351)	(0.003007)	(0.003994)	(0.000500)	(0.001001)	(0.003004)	(0.003999)
For Censored Data								
25	0.498493	1.045159	2.960839	4.376659	0.495181	1.022934	2.900897	4.269810
	(0.003501)	(0.068554)	(0.081924)	(0.096199)	(0.022771)	(0.064965)	(0.078530)	(0.091524)
50	0.520878	1.008402	3.106833	4.034051	0.518952	0.998059	3.074671	3.972429
	(0.012329)	(0.032271)	(0.039860)	(0.044237)	(0.012139)	(0.031437)	(0.039021)	(0.014708)
100	0.513967	1.023765	3.071079	4.122674	0.513049	1.018431	3.055254	4.098165
	(0.006030)	(0.016447)	(0.019412)	(0.214499)	(0.005984)	(0.016231)	(0.019208)	(0.021630)
500	0.502407	1.004056	3.013001	4.017553	0.502239	1.003031	3.009925	4.012834
	(0.001165)	(0.003185)	(0.003812)	(0.004256)	(0.001163)	(0.003177)	(0.003804)	(0.004246)
1000	0.500263	1.002041	3.000207	4.002422	0.500180	1.001530	2.998678	4.000076
	(0.000578)	(0.001587)	(1.900170)	(0.002121)	(0.000578)	(0.001585)	(0.001898)	(0.002118)

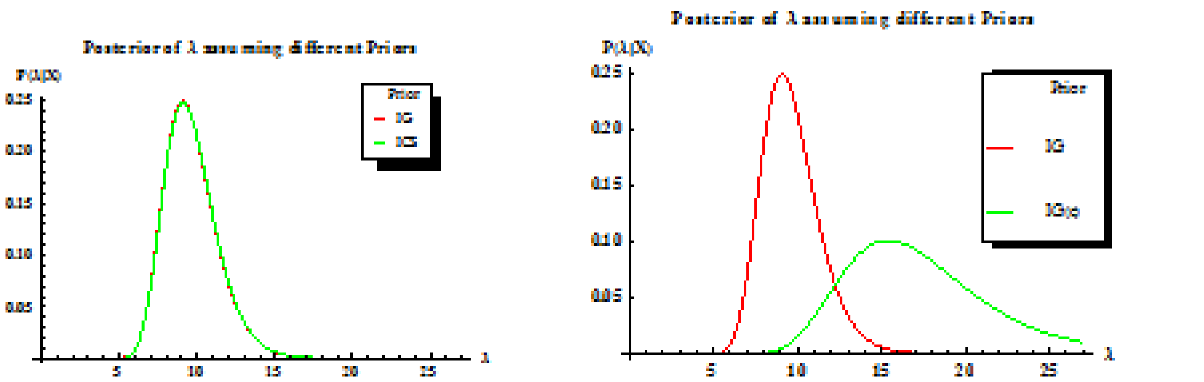


Figure 1. (i) For complete data set (ii) Using IG prior where IG(c) denote censored data

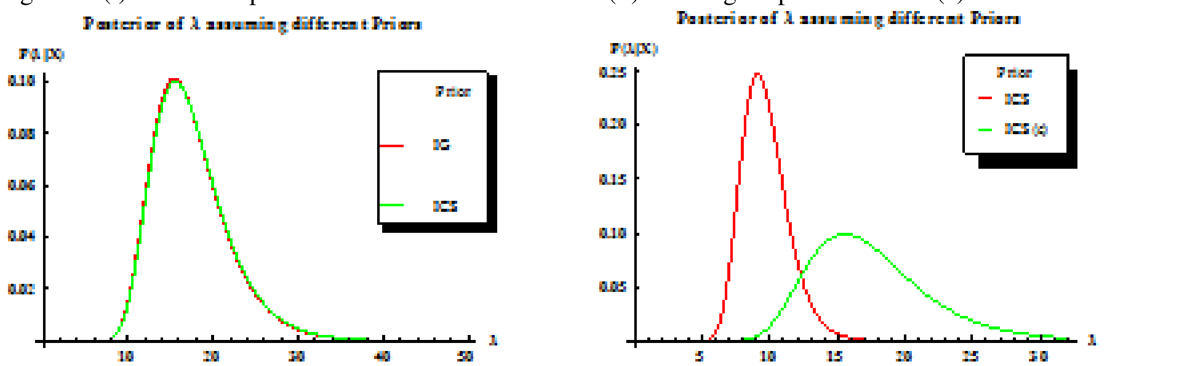


Figure 2. (i) For censored data set (ii) Using IG prior where IG(c) denote censored data

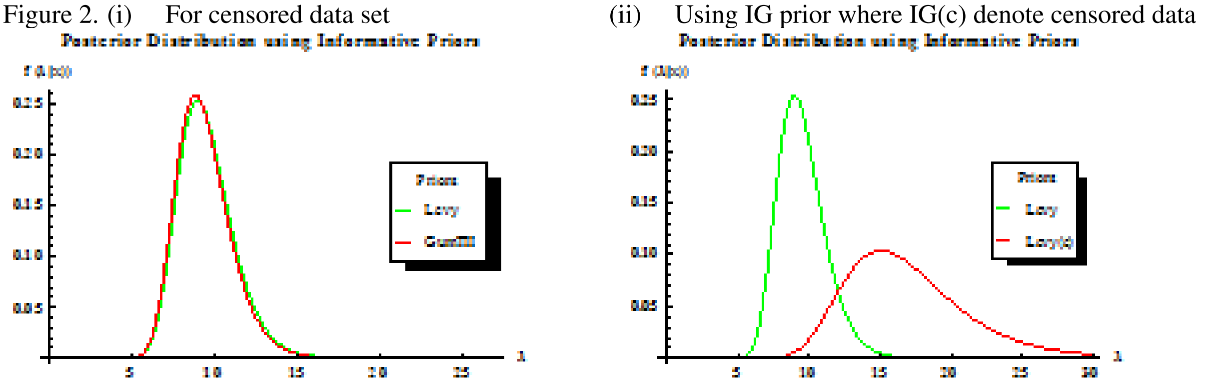


Figure 3. (i) For censored data set (ii) Using IG prior where IG(c) denote censored data

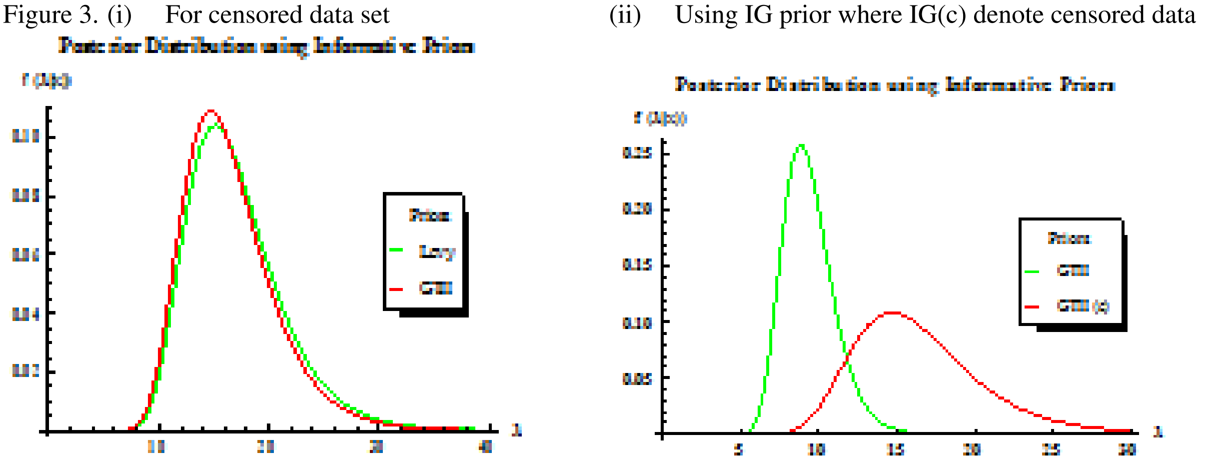


Figure 4. (i) For censored data set (ii) Using IG prior where IG(c) denote censored data

A Parametric Approach to Estimate Survival Time of Diabetic Nephropathy with Left Truncated and Right Censored Data

Gurprit Grover

Department of Statistics, University of Delhi, Delhi, India

E-mail: gurpritgrover@yahoo.com

Alka Sabharwal (Corresponding author)

Kirori Mal College, University of Delhi, Delhi, India

E-mail: sabharwal_alka@hotmail.com

Received: January 29, 2012 Accepted: March 1, 2012 Published: May 1, 2012

doi:10.5539/ijsp.v1n1p128 URL: http://dx.doi.org/10.5539/ijsp.v1n1p128

Abstract

Diabetic Nephropathy (DN) is the deleterious effect of diabetes mellitus on renal structure, or a function caused by it. The rate of rise in, serum creatinine is observed for the progression of DN. Retrospective data is collected from 132 patients. As 60 patients form uncensored cases, remaining 72 patients are censored cases. In this paper we have developed three models. The first model is based only on uncensored cases with left truncation. The second model also includes censored data. Fisher's information matrix has been applied to find variances and asymptotic confidence interval for the estimated parameters. The method of maximum likelihood is used to estimate unknown parameters in both the models. In the last model, onset times are arranged as to estimate mean onset time of DN for all the patients by using application of order statistic. We aim to predict onset time of DN for new diabetic patients.

Keywords: Diabetic Nephropathy, Fisher's information matrix, Left truncated and right censored data, Order statistic, Serum creatinine, Survival time, Weibull distribution

1. Introduction

Diabetic nephropathy (DN) develops in 20 - 40% of the patients within 10 to 15 years after the onset of diabetes. About one third of those affected eventually have progressive deterioration of renal function (Remmuzz, Schipath & Rugenetiti, 2002). Data from the UKPDS demonstrated that 25% of patients approximately with type-2 diabetes develop worse diabetic nephropathy by 10 years (UKPDS, 1998a, 1998b). Also, it is estimated that almost 50% of patients develop DN within 19 years from diagnosis of diabetes. According to the American Diabetic Association in year 2004, about 20-30% of patients with type-1 or type-2 diabetes develop nephropathy (ADA, 2004). Thus, DN is a serious problem in terms of financial load, morbidity and mortality in the developed world (Banerjee, Ghosh & Saha, 2005).

In many prospective and retrospective studies, survival data are subject to left truncation in addition to the usual right censoring. There is a widespread application and use of left truncation and right censored data (LTRC) in survival analysis. Alioum and Commenges have given proportional hazard models for arbitrarily censored and truncated data to estimate the distribution of time from diabetes onset to development of DN (Alioum & Commenges, 1996). Enzo, Vittorio, and others have applied conditional probability to find that a diabetic patient will develop a second complication, given that they had already developed the first complication (Enzo, Vittorio et al., 2003). They also propose the Bayes' formula for the same problem. Chappell, Fine and Jiang have considered semi-parametric analysis of survival data with LTRC to quantify the association between time to DN and diabetes related death and to estimate the probability of developing DN after being diagnosed with insulin dependent diabetes (Chappell, Fine & Jiang, 2005). Sparling, Younes, and Lachin have applied a parametric family of regression models for interval-censored event-time data accommodating both fixed (e.g. baseline) and time-dependent covariates. Their model employs a three-parameter family of survival distributions including special cases of Weibull, negative binomial, and log-logistic distributions , and can be applied to data with left, right, interval, or non-censored event times. The model is applied by them on diabetic patients' data to describe the effects of longitudinal measures of glycemia over time on the risk of progression of diabetic retinopathy (Sparling, Younes & Lachin, 2006).

Medical and epidemiological studies are mostly conducted with an interest in measuring the occurrence of an outcome event. In this paper we are estimating the time from onset of diabetes to development of DN subject to censoring and truncation. Weibull distribution is fitted on both uncensored and LTRC data. Weibull distribution fitting is tested through

(i) chi-square test for goodness of fit and (ii) log-cumulative hazard plot for uncensored cases. Fitting of the Weibull distribution for censored cases is tested by applying (i) Hollander and Proschan test and (ii) modified Cox-Snell residuals. The unknown parameters shape , and scale, of the distribution are estimated by the method of maximum likelihood. Fisher information matrix has been used to construct variances and asymptotic confidence intervals of the unknown parameters for LTRC data. The DN onset times of uncensored out of total cases are arranged in ascending order for the application of order statistic as: $t_{(1)}, t_{(2)}, ..., t_{(i)}, ..., t_{(n_1)}$. The probability density functions and means of minimum and maximum DN onset time have been estimated using the application of order statistic. Also, we have estimated mean DN onset times for all the patients who are under advanced nephropathy. Observed DN onset times have been compared with DN onset times obtained by using the application of order statistic. Besides introduction, the course of this paper is as follows: In section 2 developments of the models is discussed. Section 3 applies the models to type-2 diabetic patient data and section 4 contains discussion.

2. Methodology

2.1 Model for Uncensored but Left Truncated Data

Let $t_{(1)}, t_{(2)}, ..., t_{(i)}, ..., t_{(n_1)}$ be the exact DN onset times of n_1 individuals out of n patients which are known under this study. Left truncation arises when a subject is not included in the study because its diabetic nephropathy onset times (event of interest) originated prior to the starting time of the study. The DN onset times are assumed to follow Weibull distribution. The probability density and survival function of a $W(\lambda, \gamma)$ for the $i^{th}(i = 1, 2, ..., n_1)$ patient are given by:

$$f(t_i) = \frac{\gamma(\lambda)^{\gamma} t_i^{\gamma-1} exp(-(\lambda t_i)^{\gamma})}{exp(-(5\lambda)^{\gamma})} \tag{1}$$

$$S(t_i) = \frac{exp(-(\lambda t_i)^{\gamma})}{exp(-(5\lambda)^{\gamma})};\ t \geq 5,\ \lambda > 0;\ \gamma > 0 \tag{2}$$

2.1.1 Estimation of Unknown Parameters Using Maximum Likelihood Estimation

The log likelihood function is given as follows:

$$logL = n_1 log\gamma + n_1\gamma log\lambda - n_1(5\lambda)^{\lambda} + (\gamma - 1)\sum_{i=1}^{n_1} logt_i - \sum_{i=1}^{n_1}(\lambda t_i)^{\gamma} \tag{3}$$

The maximum likelihood estimates of λ and γ is found by differentiating above function with respect to λ and γ, and equating the derivatives to zero. The resulting equations are:

$$n_1 - (\lambda)^{\gamma}[n_1(5)^{\lambda} + \sum_{i=1}^{n_1}(t_i)^{\lambda}] = 0 \tag{4}$$

$$\frac{n_1}{\gamma} + n_1 log\lambda + \sum_{i=1}^{n_1} logt_i - n_1(5\lambda)^{\gamma} log5\lambda - \sum_{i=1}^{n_1}(\lambda t_i)^{\gamma}(log\lambda t_i) = 0 \tag{5}$$

The maximum likelihood estimates (MLE) of λ and γ obtained as $\tilde{\lambda}$, $\tilde{\gamma}$ by solving above two equations (4) and (5) simultaneously.

2.1.2 Estimated Mean, Variance and 95% Confidence Interval of DN Onset Time from the Diagnosis of Diabetes

Using MLE of λ and γ as $\tilde{\lambda}$, $\tilde{\gamma}$ we obtained estimated mean, variance and 95% confidence interval of DN onset time/survival time, which are given by:

$$E(onset\ time) = \int_5^{\infty} t\tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}}(t)^{\tilde{\gamma}-1}\frac{exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{exp(-(5\tilde{\lambda})^{\tilde{\gamma}})}dt \tag{6}$$

$$Var(onset\ time) = \int_5^{\infty} t^2\tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}}(t)^{\tilde{\gamma}-1}\frac{exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{exp(-(5\tilde{\lambda})^{\tilde{\gamma}})}dt - E^2(onset\ time) \tag{7}$$

$$95\%\ C.\ I\ of\ onset\ time = E(onset\ time) \pm 1.96\sqrt{Var(onset\ time)/n_1} \tag{8}$$

2.2 Model for Left Truncated and Right Censored Data

All subjects under study experienced an initial event E1 (diagnosed as diabetes as per ADA standards), but not all of them experienced a second event E2 i.e. the event of interest (onset of renal disease/diabetic nephropathy) till the study was terminated on November 2007. Right censoring happens when DN onset time of a subject is not completely observed as a second event, the end time of the study is fixed and patients enter the study at different times i.e. duration of diabetes is different for different patients. The form of censoring is generalized Type I censoring (Lee & Wang, 2003). This can be explained as: suppose that there are four patients with different duration of time and study is terminated at fixed time, where each individual's minimum diabetes duration is 5 years (left truncated). Then the starting time is backed up to 5 years for four individuals. DN onset time for first patient is observed before the end of the study, it is an uncensored case with ($\delta_1 = 1$); DN onset time for second patient cannot be observed before the end of the study, it is a censored case with ($\delta_2 = 0$); DN onset time for third patient is known and an uncensored case with ($\delta_3 = 1$) and DN onset time for fourth patient is unknown and is a censored case with ($\delta_4 = 0$). The DN onset time of first and third patient may not be same, also, censoring time of second and fourth may not be same, because patients under study are with different duration of diabetes.

There are n patients under study, out of which diabetic nephropathy onset times for n_1 patients are uncensored and DN onset time for remaining $n - n_1$ patients are censored. The DN onset time from the diagnosis of diabetes is a non-negative random variable which is assumed to follow Weibull distribution .The probability density function of a $W(\lambda, \gamma)$ for the i^{th} $(i = 1, 2, ..., n)$ patient is given by (Lee & Wang, 2003)

$$f(t_i) = \begin{cases} \frac{\gamma(\lambda)^\gamma t_i^{\gamma-1} exp(-(\lambda t_i)^\gamma)}{exp(-(5\lambda)^\gamma)}, & 5 \le t_i \le T_i \\ \frac{exp(-(\lambda T_i)^\gamma)}{exp(-(5\lambda)^\gamma)}, & t_i > T_i \end{cases} \tag{9}$$

2.2.1 MLE for Unknown Parameters for Right Censored and Fixed Left Truncated Weibull Distribution

When the censoring times T_i are different for different patients, the likelihood function for the survival times is defined as:

$$L = \prod_1^n (f(t_i))^{\delta_i} (S(T_i))^{1-\delta_i}$$

where δ_i is zero, if for the i^{th} patient DN onset time is censored and δ_i is unity when DN onset time is known. The corresponding log likelihood function is given by:

$$logL = \sum_1^n \delta_i log(\frac{\gamma(\lambda)^\gamma t_i^{\gamma-1} exp(-(\lambda t_i)^\gamma)}{exp(-(5\lambda)^\gamma)}) + \sum_1^n (1 - \delta_i) log \frac{exp(-(\lambda T_i)^\gamma)}{exp(-(5\lambda)^\gamma)}$$

$$= \sum_1^n \delta_i (log\gamma + \gamma log\lambda + (\gamma - 1) logt_i - (\lambda t_i)^\gamma + (5\lambda)^\gamma) + \sum_1^n (1 - \delta_i)((5\lambda)^\gamma - (\lambda T_i)^\gamma) \tag{10}$$

The maximum likelihood estimates of λ and γ is found by partially differentiating the above function with respect to λ and γ, and equating the derivatives to zero. The resulting equations are:

$$\lambda^\gamma (\sum_1^n \delta_i t_i^\gamma + \sum_1^n (1 - \delta_i) T_i^\gamma - n5^\gamma) - \sum_1^n \delta_i = 0 \tag{11}$$

$$\lambda^\gamma (\sum_1^n \delta_i t_i^\gamma log(\lambda t_i) + \sum_1^n (1 - \delta_i) log(\lambda T_i) T_i^\gamma - n5^\gamma log(5\lambda)) - \sum_{i=1}^n \delta_i (\frac{1}{\gamma} + log(\lambda t_i)) = 0 \tag{12}$$

The MLE of λ and γ are obtained as $\tilde{\lambda}$, $\tilde{\gamma}$ by solving above (11) and (12) equations by trial and error method.

2.2.2 Approximate Fisher Information Matrix to Construct Asymptotic Confidence Intervals of the Unknown Parameters

In this sub-section, the approximate Fisher information matrix of maximum likelihood estimators of parameters from Weibull distribution when the data are left truncated and right censored has been obtained, which can be used to construct asymptotic variance and confidence intervals of the parameters (Gross & Clark, 1975). The Fisher information matrix can be written as:

$$\begin{pmatrix} [-\frac{\partial^2}{\partial\lambda^2} logL]_0 & [-\frac{\partial^2}{\partial\lambda\partial\gamma} logL]_0 \\ [-\frac{\partial^2}{\partial\gamma\partial\lambda} logL]_0 & [-\frac{\partial^2}{\partial\gamma^2} logL]_0 \end{pmatrix}^{-1} = \begin{pmatrix} Var(\tilde{\lambda}) & Cov(\tilde{\lambda}, \tilde{\gamma}) \\ Cov(\tilde{\lambda}, \tilde{\gamma}) & Var(\tilde{\gamma}) \end{pmatrix}$$

$$\frac{\partial^2}{\partial\lambda^2} logL = \gamma\lambda^{\gamma-1} (\sum_1^n \delta_i t_i + \sum_1^n (1 - \delta_i) T_i^\gamma - n5^\gamma) \tag{13}$$

$$\frac{\partial^2}{\partial\lambda\partial\gamma}logL = \frac{\partial^2}{\partial\gamma\partial\lambda}logL = [(\sum_1^n \delta_i + n5^\gamma\lambda^\gamma(1+\gamma log(5\lambda)))/\lambda] - \lambda^{\gamma-1}[\sum_1^n \delta_i t_i^\lambda(\gamma log(\lambda t_i)+1) + \sum_1^n (1-\delta_i)T_i(\gamma log(\lambda T_i)+1)] \quad (14)$$

$$\frac{\partial^2}{\partial\gamma^2}logL = \frac{\sum_1^n \delta_i}{\gamma^2} + \lambda^\gamma[\sum_1^n \delta_i log^2(\lambda t_i)t_i^\gamma + \sum_1^n (1-\delta_i)log^2(\lambda T_i)T_i^\gamma - nlog^2(5\lambda)5^\gamma] \quad (15)$$

The approximate 95% confidence intervals for λ and γ are $\tilde{\lambda} \pm 1.96\sqrt{Var(\tilde{\lambda})}$ and $\tilde{\gamma} \pm 1.96\sqrt{Var(\tilde{\gamma})}$, respectively; where $Var(\tilde{\lambda})$ and $Var(\tilde{\gamma})$ are the diagonal elements of the variance-covariance matrix.

2.2.3 Estimated Mean, Variance and 95% Confidence Interval of DN Onset Time

Using maximum likelihood estimates of λ and γ for LTRC data as $\tilde{\lambda}$, $\tilde{\gamma}$ we obtained the estimated mean, variance and 95% confidence interval of onset time /survival time which are given by:

$$E(DN\ onset\ time|5 \le time \le 26.6) = \frac{\int_5^T t\tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}}(t)^{\tilde{\gamma}-1}\frac{exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{exp(-(5\tilde{\lambda})^{\tilde{\gamma}})}dt}{\int_5^T \tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}}(t)^{\tilde{\gamma}-1}\frac{exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{exp(-(5\tilde{\lambda})^{\tilde{\gamma}})}dt} \quad (16)$$

$$E((DN\ onset\ time)^2|5 \le time \le 26.6) = \frac{\int_5^T t^2\tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}}(t)^{\tilde{\gamma}-1}\frac{exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{exp(-(5\tilde{\lambda})^{\tilde{\gamma}})}dt}{\int_5^T \tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}}(t)^{\tilde{\gamma}-1}\frac{exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{exp(-(5\tilde{\lambda})^{\tilde{\gamma}})}dt} \quad (17)$$

$$Var(DN\ onset\ time) = E((DN\ onset\ time)^2|5 \le time \le 26.6) - E^2(DN\ onset\ time|5 \le time \le 26.6) \quad (18)$$

$$95\%\ C.I\ = E(DN\ onset\ time) \pm 1.96\sqrt{Var(DN\ onset\ time)/n} \quad (19)$$

Where, T denotes the maximum duration of diabetes among n_1 (uncensored) patients under study. The above results are compared with those obtained for uncensored cases and observed cases.

2.3 The Probability Density Function and Mean Time of First, Last and Order Statistic for the Patients Who Proceed Towards Nephropath

Let $\tau_1 \le \tau_2 \le ... \le \tau_i \le ... \le \tau_n$ be the ordered duration of n diabetic patients and n_1 be the number of uncensored diabetic nephropathy of patients with ordered onset times as, $t_{(1)} \le t_{(2)} \le ... \le t_{(i)} \le ... \le t_{(n_1)}$ where $t_{(1)}$ and $t_{(n_1)}$ denote the minimum and maximum time taken by type-2 diabetic patient proceeding towards nephropathy, respectively, among n_1 patients. The DN onset times $t_{(1)}, t_{(2)}, ..., t_{(i)}, ..., t_{(n_1)}$ are following left truncated Weibull distribution with parameters λ and γ as given in equation (1).

The probability density function and mean of first order statistic or the patient who has taken minimum time to proceed towards nephropathy is given as:

$$f_{t_{(1)}}(t) = n_1[\frac{exp(-(\lambda t)^\gamma)}{exp(-(5\lambda)^{\tilde{\gamma}})}]^{n_1-1}\gamma(\lambda)^\gamma(t)^{\gamma-1}\frac{exp(-(\lambda t)^\gamma)}{exp(-(5\lambda)^\gamma)} \quad (20)$$

$$E(t_{(1)}) = \int_5^T t_{(1)}n_1[\frac{exp(-(\lambda t)^\gamma)}{exp(-(5\lambda)^{\tilde{\gamma}})}]^{n_1-1}\tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}}(t_{(1)})^{\tilde{\gamma}-1}\frac{exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{exp(-(5\tilde{\lambda})^{\tilde{\gamma}})}dt_{(1)} \quad (21)$$

Here we are dealing with right censored data and T denotes the maximum duration of diabetes among n_1 patients under study.

The probability density function and mean of r^{th} order statistic is given as:

$$f_{t_{(r)}}(t) = rB(n_1,r)[1 - \frac{exp(-(\lambda t)^\gamma)}{exp(-(5\lambda)^\gamma)}]^{r_1-1}\gamma(\lambda)^\gamma(t_{(r)})^{\gamma-1}\frac{exp(-(\lambda t)^\gamma)}{exp(-(5\lambda)^\gamma)}[\frac{exp(-(\lambda t)^\gamma)}{exp(-(5\lambda)^\gamma)}]^{n_1-r} \quad (22)$$

$$E(t_{(r)}) = rB(n_1,r)\int_5^T t_{(r)}^{\tilde{\gamma}}[1 - \frac{exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{exp(-(5\tilde{\lambda})^{\tilde{\gamma}})}]^{r-1}\tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}}\frac{exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{exp(-(5\tilde{\lambda})^{\tilde{\gamma}})}[\frac{exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{exp(-(5\tilde{\lambda})^{\tilde{\gamma}})}]^{n_1-r}dt_{(r)}$$

The probability density function and mean of n_1^{th} order statistic or the patient who has taken maximum time to proceed towards nephropathy is given as:

$$f_{t_{(n_1)}}(t) = n_1t_{(n_1)}^{\gamma-1}[1 - \frac{exp(-(\lambda t)^\gamma)}{exp(-(5\lambda)^\gamma)}]^{n_1-1}\gamma(\lambda)^\gamma\frac{exp(-(\lambda t)^\gamma)}{exp(-(5\lambda)^{\gamma)}} \quad (23)$$

$$E_{t_{(n_1)}}(t) = n_1\tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}}[\int_5^T t_{(n_1)}^{\tilde{\gamma}} \frac{exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{exp(-(5\tilde{\lambda})^{\tilde{\gamma}})}dt_{(n_1)} + (-1)(n_1-1)\int_5^T t_{(n_1)}^{\tilde{\gamma}} \frac{exp(-2(\tilde{\lambda}t)^{\tilde{\gamma}})}{exp(-2(5\tilde{\lambda})^{\tilde{\gamma}})}dt_{(n_1)}+$$

$$(-1)^2((n_1-1)(n_1-2)/2)\int_5^T t_{(n_1)}^{\tilde{\gamma}} \frac{exp(-3(\tilde{\lambda}t)^{\tilde{\gamma}})}{exp(-3(5\tilde{\lambda})^{\tilde{\gamma}})}dt_{(n_1)} + ... + (-1)^{n_1-1}\int_5^T t_{(n_1)}^{\tilde{\gamma}} \frac{exp(-n_1(\tilde{\lambda}t)^{\tilde{\gamma}})}{exp(-n_1(5\tilde{\lambda})^{\tilde{\gamma}})}dt_{(n_1)}]$$

Thus, the mean time of first or minimum, last or maximum and r^{th} order statistic for the patients who proceed towards nephropathy are computed.

2.4 Model Checking

There are a number of techniques for evaluating the fit of parametric survival models, including analogues of residuals and influence diagnostics (Collett, 2003). To assess the adequacy of the present model, two methods have been applied for the testing of the distribution assumption for both the uncensored and censored data. Firstly for uncensored data, log-cumulative hazard plot is used, if the single sample of the survival data is available. For obtaining log-cumulative hazard plot, first compute the Kaplan Meier estimate of the survivor function, $\hat{S}(t)$ for each uncensored observation and then find the survivor function, $\hat{S}(t)$ of the fitted distribution. If the plot of $log(-logS(t))$ against $log(t)$ is an approximate straight line, and close to the line obtained from plotting $log(-log\hat{S}(t))$ against $log(t)$ then the fitted distribution assumption is tenable. Secondly, the fitting has been tested by applying Karl-Pearson's chi-square goodness of fit test. In case of censored data, firstly, the modified Cox-Snell residual plot is used for fitting distribution, proposed by Crowley and Hu. To include censored observation, Cox-Snell residual is modified by the addition of a positive constant Δ, the excess residual. Suppose that the i^{th} survival time is a censored observation, t_i^*, and let t_i be the actual, but unknown, survival time, so that $t_i > t_i^*$. Then modified Cox-Snell residual can be used for i^{th} censored observation as: $r'_{C_i} = r_{C_i}$ for uncensored and $r'_{C_i} = r_{C_i} + \Delta$ for censored observation, where $r_{C_i} = -log\hat{S}_i(t_i^*) = \hat{H}_i(t_i^*)$. The, $\hat{S}_i(t_i^*)$ and $\hat{H}_i(t_i^*)$ are the estimated KM survival and cumulative hazard functions, respectively, for the i^{th} individual at the censored survival time. After computing r'_{C_i}, it is plotted against the $-log(S(t))$ where $S(t)$ is the survival function of the fitted distribution. If this plot is close to a straight line passing through the origin, this suggests that model fitted to the data is satisfactory. Secondly, Hollander and Proschan test has been applied to test the fitted distribution survival function with the survival function obtained by applying Kaplan Meier method (Hollander & Proschan, 1979).

3. Applications

The methods discussed above are applied to the following data concerning nephropathy of type-2 diabetic patients. Up-to-date pathological reports/records of 132 diabetic patients, using a common path lab, are collected through a house to house Survey. Retrospective study is conducted on the collected data. Patients with less than 5 years diabetic history are not included or data considered is left truncated which arises when a patient is not included in the study because its DN onset time (event of interest) originated prior to the starting time of the study. Under truncation these individuals are never considered for inclusion into the study.

On the basis of serum creatinine the event of interest nephropathy, using reference as the rate of rise in SrCr, a well-accepted marker for the progression of DN, (creatinine value 1.4 to 3.0 mg/dl) is the indicator for impaired renal function (Adler, Stevens, Manley et al., 2003). Also, the renal health/dysfunction can be estimated from serum creatinine level for 132 patients under study. Thus, at the end of the study, out of 132 patients, there are only 60 diabetic nephropathy cases and remaining 72 are treated as censored. Data is simultaneously left truncated and right censored. Table-1 represents the duration-distribution of type-2 diabetic patients who developed diabetic nephropathy. It shows that out of 60 DN cases, 16.67% are with less than 10 years of diabetes, 25% with 10 to 15 years, 38.33% with 15 to 20 years and 20% are of more than or equal to 20 years of duration of diabetes. In all there are 45.45% cases with diabetic nephropathy.

The minimum age at which diabetes is diagnosed is found to be 29 yrs and maximum age as 58 years, as available from data. Table 2 represents descriptive statistics of 132 patients giving minimum, maximum and mean ± S.D of the variables; age at diagnosis, duration of disease and SrCr for two groups; censored or non diabetic nephropathy (NDN) and uncensored or diabetic nephropathy group.

<Table 1-2>

3.1 Estimated Mean, Variance and 95% Confidence Interval of DN Onset Time from the Diagnosis of Diabetes for Uncensored Cases

For modelling the survival time (from the diabetes diagnosis till the onset of nephropathy) for the uncensored cases only from the data, the minimum onset time is observed as 6 years and the maximum onset time is 26.6 years. Left truncated Weibull distribution is assumed for this case. Fitting is tested, firstly, by applying Karl-Pearson's chi-square goodness of fit test and calculated chi-square value came out to be 2.732 with p=0.435. Therefore, we conclude that there is insufficient evidence to say that data are not from left truncated Weibull distribution with $\tilde{\lambda}$=0.079515 and $\tilde{\gamma}$=2.811274. Secondly, graphical method, log-cumulative hazard plot is also used to test the fitting of distribution. For obtaining this plot, the values $log(-log(\hat{S}(t)))$, where, $\hat{S}(t)$ is Kaplan-Meier estimate of survival, and $log(-log(S(t)))$, where $S(t)$ is fitted distribution estimates of survival are computed. Then plotting, $log(-log(\hat{S}(t)))$ against log(survival time) and $log(-log(S(t)))$ against log(survival time), the resulting log-cumulative hazard plot is shown in Figure 1. In this figure the line $log(-log(S(t)))$ against log(survival time) is reasonably straight and close to the other line. This means that the assumption of left truncated Weibull distribution for the uncensored data of DN onset times is quite plausible (Collett, 2003).

Using maximum likelihood estimates as $\tilde{\lambda}$=.079515 and $\tilde{\gamma}$ =2.811274, estimated mean DN onset time is obtained as 15.19228, variance of DN onset time as 17.09349 and 95% confidence interval as (14.14615, 16.23842). These results are also compared with mean, variance and 95% confidence interval of DN onset time obtained from the observed data, which are displayed in Table 4.

<Figure 1>

3.2 Estimated Mean, Variance and 95% Confidence Interval of DN Onset Time from the Diagnosis of Diabetes by Including Censored Cases

For modelling the DN onset times including censored cases, the minimum duration of diabetes is found to be 5.6 years and maximum as 27 years as available from the data. From the data we have only partial information about these 72 cases, as they are censored cases. For this model, data is left truncated and right censored simultaneously, and Weibull distribution is assumed for this case also. Fitting is tested, firstly, on applying Hollander and Proschan goodness of fit test for censored data. The observed value of the test statistic is 1.256186 with p= 0.182, which does not fall in the rejection area. Therefore, we conclude that there is insufficient evidence to say that data are not from Weibull distribution with $\tilde{\lambda}$ =0.066834 and $\tilde{\gamma}$=1.78211 estimated by applying maximum likelihood method. Secondly, graphical method, i.e., modified Cox-Snell residual plot (Gross & Clark, 1975) is also used to test the fitting of distribution. For obtaining this plot, $r = -(log(S(t)))$ is plotted against $-log(\hat{S}(t))$, where $S(t)$ and $\hat{S}(t)$ are same as defined in section 3.1. The resultant line is modified Cox-Snell residual plot and line is close to straight line as shown in Figure 2.

Therefore, from both the methods Weibull distribution is found to be an appropriate parametric model for DN onset times.

3.2.1 Construct Asymptotic Confidence Intervals of the Estimated Parameter by Using Fisher Information Matrix

Again using maximum likelihood estimators as $\tilde{\lambda}$=0.066834 and $\tilde{\gamma}$=1.78211 of unknown parameters for censored case, the Fisher information matrix is used to obtain the asymptotic confidence intervals, $Var(\tilde{\lambda})$ and $Var(\tilde{\gamma})$, the results are depicted in Table 3. The estimated mean DN onset time is found to be 13.9058552 with corresponding variance and 95% confidence interval are obtained as 13.953207 and (13.4213197, 14.695785), respectively. The results are displayed in Table 4.

<Figure 2>

<Table 3-4>

3.3 Estimated Mean Time of First, Last and r^{th} Order Statistic for the Patients Who are under Advanced Nephropathy Group

For modelling the nephropathy onset time from diagnosis of diabetes for estimating mean DN onset time for every patient, the application of order Statistic has been used. We have used this model to estimate mean DN onset time for 27 patients with mean SrCr = 1.91mg/dl, who are under advanced nephropathy group out of 60 uncensored cases. Estimated minimum and maximum mean DN onset time are obtained and the results came out to be 5.646346 years and 24.58083 years respectively. Also, estimated mean DN onset times for 27 patients are compared with observed onset times and results are displayed in Table 5. Comparison is also illustrated graphically in Figure 3. It is observed from graph that the mean values of onset of DN for each patient up to 10^{th} order are close to observed onset time, but as the order increases the difference between observed and estimated mean value increases as higher ordered values depend on all the prior ordered values. Also, it can be observed from Table 5. The mean and variance of means of mean DN onset time obtained through application of order statistic are 15.71540 and 37.949, respectively. SPSS for Windows, Version 15 and MATLAB, Version 6.5 statistical packages were used for the calculation and analysis.

<Table 5>

<Figure 3>

4. Discussion

The aim of this paper is to estimate nephropathy onset time arising out of type-2 diabetic patients. Previous studies reveal that type-2 DM is the most common form of diabetes constituting 90% of the diabetic population and nephropathy is a life-threatening complication of diabetes mellitus. The DN onset times calculations throughout this paper are based on the values of serum creatinine, since, Bio-Stratum at their 64th Scientific Sessions ADA meet accepted rate of rise in serum creatinine as a marker for the progression of diabetic nephropathy.

There is a widespread application and use of left truncation and right censored data (LTRC) in survival analysis. Data used in this paper are simultaneously left truncated and right censored. Thus, no information was available for the patients whose DN onset time appear prior to five years and partial information was available for the patient whose DN onset times could not be observed till the end of study.

The main objectives of this paper are to estimate mean, variance and 95% confidence interval of DN onset time by fitting (i) model for uncensored cases, and (ii) model including censored cases. Weibull distribution is assumed for both the cases as it is very flexible and because of flexible shape and ability to model a wide range of failure rates, the Weibull has been successfully used in many applications (Collett, 2003; Lee & Wang, 2003). The distribution fitting is tested through graphical method, to access whether a particular theoretical distribution provides an adequate fit to the data (Lee & Wang, 2003). Also, most common Chi-square goodness of fit test is used for testing the fitting for uncensored data and Hollander and Proschan goodness of fit test is used for testing the fitting of data that includes censored observations. Weibull distribution is found to be appropriate in both the cases. Previous studies suggest that when survival time has a specific statistical distribution; the statistical power of parametric distribution is higher than nonparametric and semi-parametric models.

Firstly, for uncensored cases using Weibull model the mean of DN onset time is found to be 15.19228 years whereas the the observed mean of DN onset time is 15.21333 years. Thus, the estimated and the observed mean DN onset times are found to be remarkably similar. Secondly, mean of DN onset time, 13.905855 years, is obtained by including censored cases. The mean of DN onset time based on uncensored observations is higher than the case which includes censored observations and this matches with Li and Lagakos finding suggesting that considering only uncensored cases will increase the mean of the survival time (Li & Lagakos, 1997). The mean DN onset times obtained from both the cases are almost consistent with the previous findings which suggest that diabetic nephropathy develop within 10 to 15 years after the onset of diabetes (ADA, 2004).

In part three, the model is applied with the application of order statistic to estimate minimum and maximum mean DN onset time and also, mean DN onset time for all the 27 patients out of 60, who are under the advanced nephropathy group. It is found from Table 5, which is visible through graph, that estimating mean DN onset time using the application of order statistic is a useful exercise if the sample size is small. This is further, verified by applying paired t-test. We have divided 27 patients in accordance with the ascending order of DN onset time into two groups of sizes 14 and 13. By applying paired t-test for comparing the observed DN onset times with those obtained from the application of order statistic on first group(comparing first 14 ordered values) the test statistic value is found to be 1.659 with p=0.121, (p >0.05) on 13 degrees of freedom. Thus, we conclude that there is not enough evidence to reject the hypothesis as there is no significant difference between the estimated and the observed DN onset times. Then by applying paired t-test on second group (ordered values from 15 to 27) the test statistic value is found to be 14.645 with p <0.001 on 12 degrees of freedom. From this we conclude that difference between observed and estimated mean DN onset time increases with an increase in the order of the observation. The minimum and maximum mean of DN onset times are found to be 5.646346 years and 24.58083 years respectively.

Weibull model can empirically fit a wide range of data. Thus, Weibull distribution may be used to model the survival distribution of a population with increasing, decreasing, or constant risk and can be taken as a baseline model for future studies.The model can be applied for estimating survival times of other complication arising out of type-2 diabetes such as retinopathy, CVD and others. The application of order statistic can be further explored with other survival models and also for other biomedical complications. The major use of estimating the nephropathy onset time of diabetic patient is that current as well as future DN onset time of new diabetic patients can be predicted.

References

Adler, A. I., Stevens, R. J., Manley, S. E. et al. (2003). Development and progression of nephropathy in type 2 diabetes. United Kingdom Prospective Diabetes Study, 64. Kidney International, 63, 225-232. http://dx.doi.org/10.1046/j.1523-1755.2003.00712.x

Alioum, A., & Commenges, D. (1996). A Proportional Hazard models for Arbitrarily Censored and Truncated Data. *Biometrics, 52*, 512-524. http://dx.doi.org/10.2307/2532891

American Diabetes Association. (2004). Nephropathy in Diabetes. *Diabetes Care, 27*, S79-S83. http://dx.doi.org/10.2337/diacare.27.2007.S79

Banerjee, S., Ghosh, U. S., & Saha, S. J. (2005). Role of GFR Estimation in Assessment of the Status of Nephropathy in Type 2 Diabetes Mellitus. *JAPI, 53*, 181-184.

Chappell, R., Fine , J., & Jiang, H. (2005). Semiparametric Analysis of Survival Data with Left Truncation and Right Censoring. *Biometrics, 61*, 567-575. http://dx.doi.org/10.1111/j.1541-0420.2005.00335.x

Collett, D. (2003). Modeling Survival data in Medical Research. Second edition. London: Chapman and Hall/CRC.

Enzo, B., Vittorio, C., Nicola, M. Di, et al. (2003). Probabilistic approach to developing nephropathy in diabetic patient with retinopathy. *Statistics in Medicine, 20*(24), 3889-3897.

Gross, J. A., & Clark, A.V. (1975). Survival Distributions: Reliability Applications in the Biomedical Sciences. New York: John Wiley and Sons.

Hollander, M., & Proschan, F. (1979). Testing to determine the Underlying Distribution Using Randomly Censored Data. *Biometrics, 35*, 393-401. http://dx.doi.org/10.2307/2530342

Lee, Elisa. T., & Wang, J. W. (2003). Statistical Methods for Survival Data Analysis, third edition. New York: John Wiley and Sons.

Li, Q. H., & Lagakos, S. W. (1997). Use of the Wei-Lin-Weissfield Method for the Analysis of a Recurring and a Terminating Event. *Statistics in Medicine, 16*, 925-940. http://dx.doi.org/10.1002/(SICI)1097-0258(19970430)16:8<925::AID-SIM545>3.0.CO;2-2

Remmuzz, G., Schipath, A., & Rugenetiti, P. (2002). Nephropathy in patients with type 2 diabetes mellitus. *New Eng. J. Med., 346*, 1145-51. http://dx.doi.org/10.1056/NEJMcp011773

Sparling, Y. H., Younes, N., & Lachin, J. M. (2006). Parametric survival models for interval-censored data with time-dependent covariates. *Biostatistics, 7*(4), 599-614. http://dx.doi.org/10.1093/biostatistics/kxj028

U. K. Prospective Diabetes Study Group. (1998). Intensive blood glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type-2 diabetes. UKPDS 33. Lancet, 352, 837-853. http://dx.doi.org/10.1016/S0140-6736(98)07019-6

U. K. Prospective Diabetes Study Group. (1998). Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes, UKPDS 38, BMJ, 317, 703-713. http://dx.doi.org/10.1136/bmj.317.7160.703

Villar, E., Chang, S. H., & McDonald, S. P. (2007). Incidences, Treatments, Outcomes, and Sex Effect on Survival in Patients with End-Stage Renal Disease by Diabetes Status in Australia and New Zealand (1991-2005). *Diabetes Care, 30*(12), 3070-76. http://dx.doi.org/10.2337/dc07-0895

Table 1. Duration-distribution of Diabetic patients till Nov'2007, who developed Diabetic Nephropathy

Duration of Diabetes (years)	Number of diabetic patients	Diabetic Nephropathy (%)
<10	56	16.67
10-15	20	25
15-20	31	38.33
≥20	25	20
Total	132	45.45

Table 2. Descriptive statistics of 132 patients giving minimum, maximum and mean ± standard deviation of age at diagnosis, duration of diabetes, and serum creatinine for two groups i.e. NDN (censored) and DN (uncensored) Group

Variable	Statistic	Censored (N=72)	Uncensored (N=60)
Age at Diagnosis (years)	Minimum	35	29
	Maximum	58	56
	Mean±S.D	44.0108±4.36	45.0034±5.2824
Duration of Disease (yrs)	Minimum	5.6	6
	Maximum	27	26.6
	Mean±S.D	10.2784±5.5	14.0931±5.0528
Serum Creatinine (mg/dl)	Minimum	0.71	1.29
	Maximum	1.39	2.21
	Mean±S.D	0.9780±0.12616	1.6686±0.28233

Table 3. Variance and asymptotic confidence intervals for the estimates of the parameters obtained by using Fisher information matrix

Parameter	Maximum likelihood estimator	Variance	95% Confidence interval
$\tilde{\lambda}$	0.066834	0.001078	(0.002487, 0.131181)
$\tilde{\gamma}$	1.78211	0.000391	(1.820866, 1.743354)

Table 4. Estimated mean, variance and 95% confidence interval of DN onset time obtained from observed and fitting (i) left truncated Weibull distribution on uncensored cases (ii) left truncated Weibull distribution by including censored cases

Cases	Mean	Variance	95% C.I
Observed uncensored cases	15.21333	22.87686	(14.003195,16.42357)
LT Weibull uncensored cases	15.19228	17.09349	(14.14615,16.23842)
Weibull LTRC data	13.9058552	13.953207	(13.4213197,14.6957854)

Table 5. Estimated and observed DN onset time for 27 patients who are under advance diabetic nephropathy group

Ordered Observation	Diabetes Duration(yrs)	Serum Creatinine (mg/dl)	Diabetic Nephropathy Onset time Observed (yrs)	Estimated Diabetic Nephropathy Onset time (yrs)
1	7.8	1.98	6	5.646346
2	12	1.7	6.6	6.215418
3	13	1.54	7.6	6.297053
4	13	1.66	7.9	7.321488
5	13.6	1.97	8.6	7.522939
6	15	2.14	9.2	10.32963
7	15	2.11	9.6	10.80505
8	15.6	2.21	11	12.42911
9	16	2.18	11.6	12.6954
10	16	1.99	12	14.13889
11	16	1.62	13	14.72397
12	16	1.66	13.6	15.22681
13	16.6	1.89	13.6	15.86091
14	17	1.57	14	16.52023
15	17	1.97	14.6	17.30451
16	17.2	1.6	15	17.76943
17	17.6	2.11	15.6	18.49336
18	18	1.79	15.6	19.55992
19	18	2.18	16	20.68734
20	19	2.19	16.6	20.79337
21	20	1.79	18	21.0774
22	20	1.74	18	21.9298
23	23	1.79	19.6	23.04411
24	24.4	2.05	19.6	23.28542
25	26	2.01	21	23.66707
26	26	1.94	22	24.08705
27	26.6	2.14	22.6	24.58083

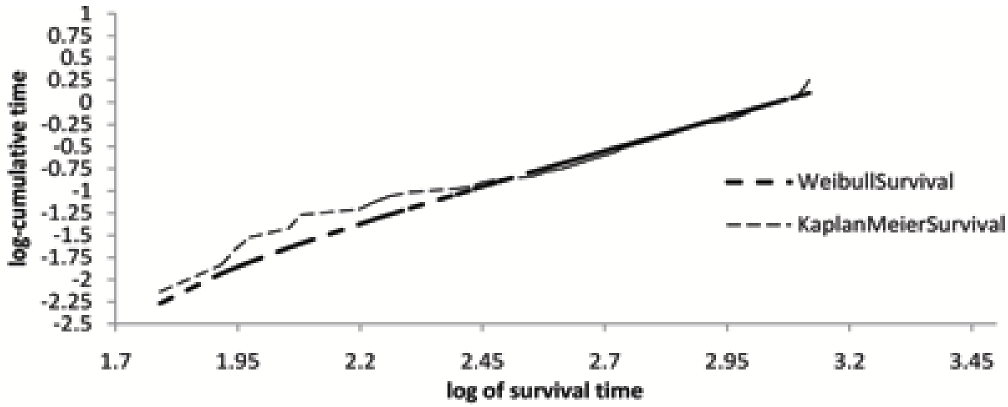


Figure 1. Fitting of Weibull distribution for the uncensored data by plotting (i) $log(-log(\hat{S}(t)))$ against log(survival time) (ii) $log(-log(S(t)))$ against log(survival time)

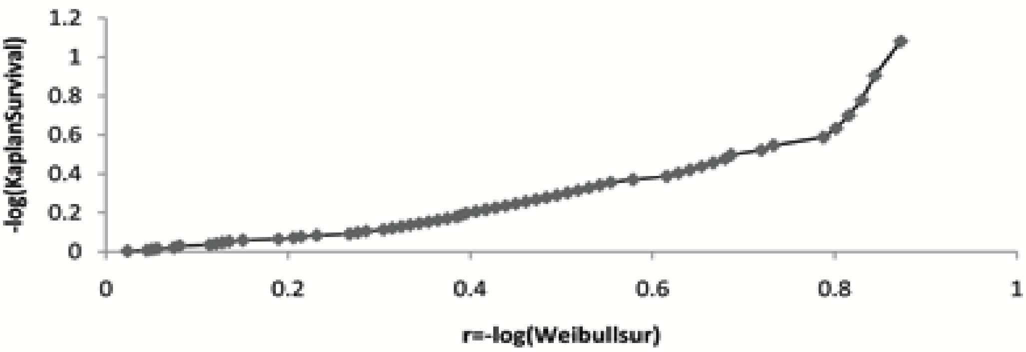


Figure 2. Fitting of Weibull distribution for censored data by modified Cox-Snell residual method by plotting r=-log(Weibull Survival) against (-log(Kaplan-Meier Survival))

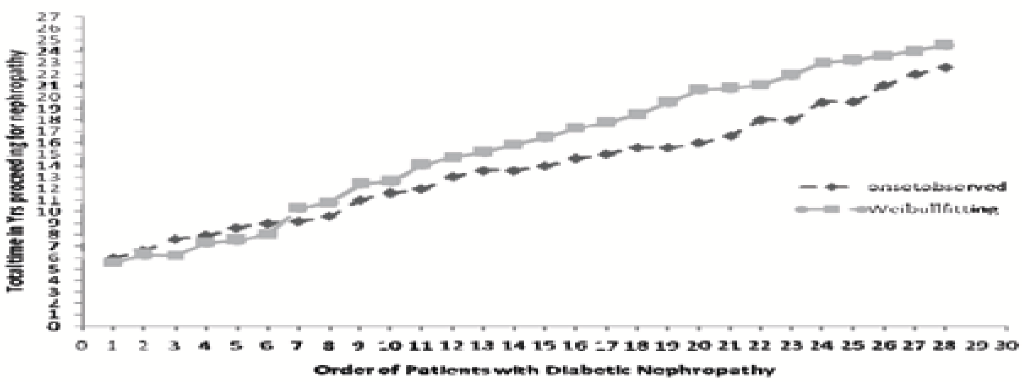


Figure 3. Estimated and observed survival time for 27 patients who are under advance diabetic nephropathy group

The Importance of Sustained Professional Development for Teaching Statistics – An Example Involving the Mode and Range

Tim Jacobbe (Corresponding Author)
University of Florida, 2403 Norman Hall
Gainesville, FL 32611, USA
Tel: 1-352-273-4232 E-mail: jacobbe@coe.ufl.edu

Bob Horton
Clemson University, 409B Tillman Hall
Clemson, SC 29634-0705, USA
Tel: 1-864-656-5127 E-mail: bhorton@clemson.edu

Received: February 3, 2012 Accepted: February 15, 2012 Published: May 1, 2012
doi:10.5539/ijsp.v1n1p138 URL: http://dx.doi.org/10.5539/ijsp.v1n1p138

Abstract

This paper presents a progression of events from a case study that explored elementary school teachers' understanding of essential topics in statistics. One particular teacher revealed misconceptions regarding the mode and range during professional development. These misconceptions were corrected, the concepts taught appropriately in the classroom, and then the original misconceptions resurfaced. This example sheds light on the importance of sustained professional development.

Keywords: Teachers knowledge, Statistics education, Professional development

1. Introduction

Statistics is a relatively new component of the K-12 curriculum. In fact, the efforts of the Quantitative Literacy Project (Schaeffer, 1986) led the National Council of Teachers of Mathematics (NCTM) to first include data analysis as part of its Curriculum and Evaluation Standards (1989). The emphasis has gradually increased in the Principles and Standards for School Mathematics and most notably in American Statistical Association's (ASA) Guidelines for Assessment and Instruction in Statistics Education (GAISE) (2007). These efforts to increase the quantity and quality of statistics content are not isolated to the United States. For example, statistics educators in Australia have been attempting to increase the amount of statistical content in their schools since the release of A National Statement on Mathematics for Australian Schools (Australian Education Council, 1991). All of these efforts built upon the momentum that began with the Quantitative Literacy Project and have eventually led to the inclusion of statistics in the College and Career Readiness Standards that are currently being finalized at a national level in the United States.

The inclusion of statistics in the K-12 curriculum has been an exhausting effort for many statisticians and statistics educators. Now that it is in the curriculum, there is another concern that must be addressed - the preparation of teachers to teach statistical content at the level of sophistication described in the GAISE framework. As Shaughnessy stated almost twenty years ago, "Teachers' backgrounds are weak or nonexistent in stochastics and in problem solving. This is not their fault, as historically our teacher preparation programs have not systematically included either stochastics or problem solving for prospective mathematics teachers" (1992, p. 467). This concern by Shaughnessy was still evident 15 years later, as supported by his statement, "Most K-12 teachers in the United States have very little background in statistics" (p. 995). It likely continues to be a concern.

Over the past several years, statistics educators have called for courses focused on the teaching and learning of statistics to be included in teacher preparation programs (e.g., Jacobbe, 2007; Kader & Perry, 2002). As important as these are for the preparation of preservice teachers, they cannot address the needs of inservice teachers who have already completed their preservice training. For these teachers, many researchers have recommended and some have implemented professional development programs (e.g., Friel & Bright, 1998; Watson, 1998).

Professional development is important for changing what transpires in the classroom. For inservice teachers, these opportunities represent one of the only avenues for exposing them to, and helping them master, the advanced content included in frequent reform efforts or additions to the curriculum. Research on key characteristics in professional development (e.g., Abell, S., & Lee, M., 2008; Chval, K., Abell, S., Pareja, E., Musikul, K., & Ritzka, G., 2008; Loucks-Horsley, Hewson, Love, & Stiles, 1998) clearly emphasizes the importance of sustained professional development. A report from the Institute of Educational Sciences (2007) confirms this, stating:

> Studies that had more than 14 hours of professional development showed a positive and significant effect on student achievement from professional development. The three studies that involved the least amount of professional development (5 - 14 hours total) showed no statistically significant effects on student achievement. (2007, p. iv)

Of the studies cited by IES (2007), only six met this standard of a minimum of 14 hours. On average, the professional development involved in these six studies contained 49 hours of contact hours, which translated to a net increase in students' achievement by 21 percentile points, though it is not always clear on what measures these improvements were noted.

Guskey (2002) identified five levels of professional development evaluation. These include participants' reactions (Level 1), participants' learning (Level 2), organization support & change (Level 3), participants' use of new knowledge and skills (Level 4), and student learning outcomes (Level 5).

In another study, Guskey (2003) analyzed 13 lists of characteristics of effective professional development, in part to determine if there was overlap between some of the characteristics various organizations recommended. Based on Guskey's analysis, the most common characteristic among the lists was that effective professional development "enhances teachers' content and pedagogical knowledge," which appeared on 12 of the 13 lists. The second most common was that effective professional development "provides sufficient time and other resources," which appeared on 10 of the 13 lists.

In addition to the issue of time, and of more concern to the statistics education community, is the issue of whether or not teachers actually receive training that promotes statistical rather than mathematical thinking. As delMas (2004) eloquently described by pulling on the work of Cobb & Moore (1997), there is a difference between mathematical and statistical reasoning. According to delMas,

> ...it is likely that many aspects of statistical and mathematical reasoning are highly similar. The task demands of each discipline, however, may produce different sources of reasoning error. While instruction can be driven and facilitated by contextualization in both disciplines, statistical practice is highly dependent on real world context whereas mathematical practice tends to be removed from real world context (Cobb & Moore, 1997). (delMas, 2004, p. 91)

This difference is also highlighted in the GAISE framework and is important to consider when developing and implementing professional development programs to improve teachers' preparation to teach statistics. Mathematics educators - who themselves were most likely trained in a manner which did not promote statistical thinking focused on dealing with the prevalence of variability - are the ones who are asked to primarily deliver professional development to schools. The result of this may leave teachers being trained in a manner that promotes various measures of central tendency or variation without engaging in the four components of statistical thinking as described in the GAISE framework - Formulating Questions, Collecting Data, Analyzing Data, and Interpreting Results. The remainder of this paper describes a professional development experience that one teacher experienced and then highlights a progression of events that emphasizes the need for sustained professional development situated within statistics.

The progression of experiences described in this paper provides a strong example of why professional development must be a sustained effort. Further, it explores one participant's learning (Level 2) as well as her use of new knowledge and skills (Level 4) as described by Guskey. In particular, it sheds light on the impact a particular professional development opportunity had on one participant's content knowledge as measured on three occasions.

2. Method

A case study involving elementary school teachers was conducted in a school district in the USA. The example presented in this paper is from a larger study that explored elementary school teachers' understanding of basic ideas in statistics (for further details see Jacobbe, 2007). Ms. Brown (a pseudonym), who had taught Grade 3 (age 9) for all four years of her teaching career at the time of the study, is the participant discussed in this paper.

Initial contact was made with Ms. Brown through a gatekeeper - the district supervisor for mathematics and science education. Ms. Brown was highly recommended by her principal and district supervisor as she was viewed to be an exemplary teacher of mathematics. Although generalizations cannot be made based on this example, the sequence of events provides insight into the importance of sustained development, particularly since the participant had been identified as an exemplary teacher.

Over the course of 18 months, the teachers in the study participated in interviews, completed questionnaires and assessments, and allowed the researcher to observe their classroom at least 12 times. The researcher was also present for professional development training the teachers received. The results reported in this paper are from the researcher's observation of the professional development training as well as accounts from two of the observations involving Ms. Brown. The professional development experience was designed to introduce the teachers to the investigations they would be asked to lead while implementing the curriculum materials. Based on when the professional development was scheduled, it worked out that Ms. Brown was actually scheduled to implement the very same lessons only two days later. The lessons all came from the teachers' manual of a newly developed curriculum by Clemson University entitled Math Out of the Box (Moss, Diaz, Lashley, Moss, & Sanders 2005), which contained 10 lessons focused on statistics at each grade level. The developers of the Math Out of the Box materials provided the professional development involved in this study. It is important to point out that the professional development providers were not specifically trained in the area of statistics.

3. Results

The discussion presented here centers on Ms. Brown's experiences with two fundamental topics in statistics, the mode and the range. The first event stems from the professional development training, which is designated as Day 0. During this experience the first author sat with Ms. Brown as she was participating in the professional development. Two days later, Ms. Brown was observed teaching a lesson. Six days after the first observation of Ms. Brown, she was observed teaching a lesson once again. All three of these experiences (researcher's presence at professional development and two observations of teaching) involved the concepts of the mode and range. Each of these events is described in succession.

3.1 Professional Development Training (Day 0)

The professional development training was provided by the authors of the Math Out of the Box materials. The purpose of the training was to introduce some of the statistics lessons to the teachers who would be implementing them during the academic year, as well as to review some of the statistical content that might arise during the lessons.

The context for the lesson introduced during the professional development was the circumference of individuals' wrists. During this training, the researcher sat at the table with Ms. Brown as she was participating in the professional development. The following lineplot was constructed based on data from the participants in the training.

<Figure 1>

During Ms. Brown's exploration with the teachers, she looked over at the researcher and indicated, "I know this stuff." She proceeded to say that the mode of the data was 9, because 8 inches occurred as a data value 9 times. However, the mode is the data value or values that occur most often, not the frequency of that value. For this particular data set, the mode was 8 inches since that data point occurred most often (9 times). Ms. Brown reported the range as 9 - 2 or 7. This was obtained by taking the highest frequency (9) minus the lowest frequency (2). For this particular data set, the range should have been reported as 9 (the maximum value represented in the data set) minus 6 (the minimum value represented in the data set, or 3 inches, so, as with mode, Ms. Brown confused the value of the variable in question with the frequency of that value.

The authors believe that it is more valuable to report the range as a range of values (e.g. 6 to 9 inches) in addition to a singular value (3 inches) that represents the difference between these values. By reporting the range of values, one knows where the data are situated without having to view the original data set. In contrast, by reporting only a singular value one only knows that the maximum and the minimum are 3 inches apart. In other words, if one knows that the range is 3 inches, one cannot tell if the data range from 6 inches to 9 inches, from 12 to 15 inches, or between any other values 3 inches apart. We further believe that this clarification may help reinforce the idea that the mode and range concern the values of the variables rather than their frequencies.

The error exhibited my Ms. Brown is not uncommon (e.g., Groth & Bergner, 2006). It was most likely due to confusion regarding the variable in question. The variable in this context was the wrist circumference; however Ms. Brown considered the frequency as the variable. The point made regarding the reporting of the mode above (reporting the mode as 3 inches, from 6 to 9 inches, vs. reporting it only as 3 inches) may have helped Ms. Brown as the range would have been situated within the context of the investigation.

During the professional development training, the correct method to find the mode and the range were discussed. Ms. Brown realized her error (i.e. using the frequencies rather than the values of the variables) and by the end of the training indicated that the mode was 8 inches and the range was 3 inches. Given other contexts as part of the professional development training, Ms. Brown was able to correctly determine the mode and the range.

Two days after this training, Ms. Brown introduced the same lesson involving wrist circumference to her students and included a discussion of the mode and range. Her success with these ideas is described below.

3.2 Wrist Circumference Lesson Two Days Later (Day 2)

During this lesson, students measured the circumference of each other's wrists. The data were collected and organized into a tally chart. Based on the data, the following lineplot was constructed.

<Figure 2>

During a discussion one of the students in Ms. Brown's class introduced the terminology of the range. Comments 1 through 44 are from a discussion that occurred between Ms. Brown and her students regarding the mode and range.

<Table 1>

During this sequence of events, Ms. Brown corrected the misconceptions introduced by students. These were the same misconceptions (comments number 4 and 8 for the range; comment number 34 for the mode) that Ms. Brown exhibited during the professional development training just two days earlier. Based on this lesson, it would seem as though the professional development training was successful.

Six days after the lesson on wrist circumference outlined above, Ms. Brown had another opportunity to work with this same group of students on mode and range. This encounter is described in the following section.

3.3 Birth Length Lesson (Day 8)

The context of the lesson observed on this day was students' birth lengths. Students had been asked to bring in their birth lengths (in inches) for the class to investigate. During the morning of this lesson, Ms. Brown also re-introduced the concept of the mode and range for a data set represented once again by a lineplot. However, in her discussion, there was no context, so no variables were identified; further, Ms. Brown did not mention that the values of the variable were identified on the horizontal axis, nor did she indicate that the vertical values, represented by the X's in the chart, showed the frequency, with each X representing one occurrence. The lineplot was displayed on the board as follows:

<Figure 3>

During this time, Ms. Brown told the students that the mode of this data set was 7 since this was the highest frequency (note that the correct mode was 6) and the range for this data set was 7 (the highest frequency) minus 2 (the lowest frequency) or 5 (note that the correct range was 3 to 7 or 4). This misconception matches the one Ms. Brown exhibited during the professional development training and the one Ms. Brown had corrected during the class six days earlier (comments 4, 8, and 34 above). This same misconception became an issue yet again during the session on birth lengths, which involved the following lineplot.

<Figure 4>

Comments 45 through 52 represent the conversation between Ms. Brown and a student (the same student who made comments 35 through 44 above) concerning the mode and range after Ms. Brown had explained (incorrectly) how to determine these values.

<Table 2>

It should be noted that the student who went up to the board and correctly showed that the mode of the wrist sizes in the first lesson was 6 inches (comment number 38) was the same student who questioned finding the mode and range in the birth lengths lesson (comment number 45).

3.4 Summary of Observations

Table 3 provides a timeline for what transpired. In the table, Day 0 corresponds to the professional development training, Day 2 corresponds to the wrist sizes lesson, and Day 8 corresponds to the lesson on birth lengths.

<Table 3>

This sequence of events causes a concern over the power of one-day professional development workshops. Even in instances where they appear to have a clear and positive impact on teachers' understanding, it is possible, perhaps even likely, for teachers to revert back to their original misconceptions. The importance of sustained professional development is widely documented (Chval, Abell, Pareja, Musikul, & Ritzka, 2008: Loucks-Horsley, Hewson, Love, & Stiles, 1998; Abell & Lee, 2008; Fraser, Reid, & McKinney, 2007). Much of this research, however, has focused on sustained PD in an effort to support and reinforce transformation in pedagogy. However the sequence of events described in this paper provides an example of the need to sustain PD in regard to content, particularly in the area of correcting misconceptions.

4. Discussion and Implications

The sequence of events outlined in this paper provides an example of the importance for sustained professional development. It not only calls into concern the quantity of time, but also the quality of training teachers receive in regard to statistics. That isn't to say that the training Ms. Brown received was poor, but that it was not sufficient to overcome her misconceptions. Perhaps if the PD had been more often and had addressed all four components of the statistical process that the GAISE identifies - formulating questions, collecting data, analyzing data, and interpreting results - Ms. Brown may have been able to overcome her misconceptions. Though the training the lead Author witnessed was of high quality, it focused more on collecting and analyzing data, but did not engage her with formulating questions and interpreting results. In the former case, the questions were pre-determined for her, and in the latter case, Ms. Brown did not have to reflect upon and discuss what the results suggested. This may have been a result of the training not being provided by individuals specifically trained in statistics. However, mathematics educators or exemplary teachers are often the ones asked to provide professional development to districts. If the process of statistics is to be emphasized, it is important for statisticians and statistics educators to become more engaged in providing professional development to teachers.

Although the professional development opportunity appeared to affect Ms. Brown's knowledge (Level 2 as described by Guskey), in that she was able to identify the mode and range for data sets after the professional development, and upon her ability to implement this knowledge successfully in the classroom (Level 4 as described by Guskey), in that she was able to teach the concepts appropriately two days after the professional development, this one-day training did not make a lasting impact on Ms. Brown. This is evident by the fact that she reverted to her original misconceptions only eight days after receiving the professional development training.

This situation seems to support the report from the IES and other professional development research (e.g., Chval, Abell, Pareja, Musikul, & Ritzka, 2008; Loucks-Horsley, Hewson, Love, & Stiles, 1998; Abell & Lee, 2008; Fraser, Reid, & McKinney, 2007), which calls not only for on-going professional development, but the creation of communities of teachers. Despite initial indications that showed the professional development had an impact on Ms. Brown, follow-up observations revealed that the professional development did not last. Perhaps even more troublesome is the fact that she told her students one thing on Day 2 and another on Day 8, and in fact contradicted a student who had learned the ideas correctly during the initial lesson and questioned the teacher when she reversed course on the later day.

Despite the setbacks, the study also revealed reasons for optimism. By the end of the study, Ms. Brown (and indeed the other teachers in the study) had acknowledged an awareness of her lack of content knowledge in the area of statistics and a desire to receive professional development focused on this particular content strand. From the larger study it was clear that although the teachers did not experience lasting gains in their understanding of essential topics in statistics, it caused them to reconsider the suitability of their own content knowledge.

As Gal (2004) points out, one of the biggest obstacles toward the inclusion of statistics in the curriculum is teachers' (and society's in general) disposition toward the discipline of statistics. If teachers have only been exposed to statistical content as a discipline entrenched in procedures rather than the conceptual underpinnings of the processes, then they will not realize that there is a gap in their knowledge. For all the teachers in the case study, the interaction caused the teachers to question their dispositions toward statistics. In other words, this interaction problematized the awareness of their knowledge.

Once teachers recognize new viewpoints or what may be lacking in their own understanding, problematization occurs. Problematizing teachers' knowledge is essential for professional development to be successful in changing teachers' preparedness for teaching (Cobb & Bauersfled, 2005). Teachers who realize they have a lack of understanding in a particular area are more likely to benefit from professional development focused on content. The teachers involved in this study provide an example that illustrates the importance of such a realization. At the beginning of the study, the teachers preferred activities they could use in their classrooms over developing their own content knowledge. At the end of the study, they preferred professional development focused on content. With the problematization they experienced, the teachers would be more likely to absorb the content introduced during sustained professional development that is fundamental to true understanding and the processes of statistical inquiry. Though the initial professional development had not had the desired effect in terms of knowledge and pedagogy, the teachers' recognition of their gaps suggested that, in the long term, the goals for understanding would be met.

5. Conclusion

This study confirmed that a single effort at correcting teachers' misconceptions may not be successful. Professional development in the area of statistics might be more successful if it were offered on a more sustained basis and focused on the statistical processes as described in the GAISE framework. Further studies are necessary to investigate the effect of various professional development opportunities on the statistical content knowledge of elementary school teachers;

however it is clear that one-shot professional development will not overcome some teachers' misconceptions.

References

Abell, S., & Lee, M. (2008). Making the most of professional development. *Science and Children, 45*, 63-64.

Australian Education Council. (1991). *A national statement on mathematics for Australian schools*. Carlton, Australia: Author.

Chval, K., Abell, S., Pareja, E., Musikul, K., & Ritzka, G. (2008). Science and mathematics teachers' experiences, needs, and expectations regarding professional development. *Eurasia Journal of Mathematics, Science & Technology Education, 4*(1), 31-43.

Cobb, G. W., & Moore, D. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly, 104*, 801-823. http://dx.doi.org/10.2307/2975286

DelMas, R. C. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenges of developing statistical literacy, reasoning, and thinking*, pp. 79-96. Dordrecht, The Netherlands: Kluwer.

Fraser, C., Kennedy, A., Reid, L., & McKinney, S. (2007). Teachers' continuing professional development: Contested concepts, understandings and models. *Journal of In-Service Education, 33*(2), 153-169. http://dx.doi.org/10.1080/13674580701292913

Groth, R. E., & Bergner, J. A. (2006). Preservice elementary teachers' conceptual and procedural knowledge of mean, median, and mode. *Mathematical Thinking and Learning, 8*, 37-63. http://dx.doi.org/10.1207/s15327833mtl0801_3

Guskey, T. R. (2002). Does it make a difference? Evaluating professional development. *Educational Leadership, 59*(6), 45-51.

Guskey, T. R. (2003). Analyzing lists of the characteristics of effective professional development to promote visionary leadership. *NASSP Bulletin, 87*(637), 4-17. http://dx.doi.org/10.1177/019263650308763702

Jacobbe. (2007). Elementary school teachers' understanding of essential topics in statistics and the influence of assessment instruments and a reform curriculum upon their understanding. [Online] Available: http://www.stat.auckland.ac.nz/ iase/publications/dissertations/dissertations.php

Kader, G., & Perry, M. (2002). A statistics course for elementary and middle school teachers. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*, Cape Town, South Africa: International Statistical Institute and International Association for Statistical Education. [Online] Available: http://www.stat.auckland.ac.nz/ iase/publications.

Loucks-Horsley, S., Hewson, P., Love, N., & Stiles, K. (1998). *Designing professional development for teachers of science and mathematics*. Thousand Oaks, CA: Corwin Press, Inc.

Moss, D., Diaz, D., Lashley, E. Moss, W., & Sanders, S. (2005). *Developing Algebraic Thinking: Plotting and Growing, Teacher's Manual, Third Grade*. Burlington, NC: Carolina Biological Supply.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Jacobbe.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007-No. 033). Washington, DC: U. S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs.

Table 1. Wrist circumference lesson transcript

Number	Person	Comment
1	Ms. Brown	[To a particular student] You brought up a good word earlier. What was that word?
2	Student	Range
3	Ms. Brown	What is the range?
4	Student	Range is the difference between the biggest number of X's and the smallest number of X's.
5	Ms. Brown	Anyone else?
6	Student	Difference between the most popular and least popular.
7	Ms. Brown	Can you show me?
The student showed that there were 10 X's at 6 and 2 X's at 5.		
8	Student	So you take the 10 minus 2 to get a range of 8.
9	Ms. Brown	Anyone else? How many number choices did we have when we recorded the information?
10	Student	15
11	Ms Brown	How many measurements, in inches, did we actually have?
12	Student	3
13	Ms. Brown	What measurement is the highest we used?
14	Student	6 inches
15	Ms. Brown	Was that the highest number of inches that we used?
16	Student	No. We used 7 as the highest.
17	Ms. Brown	What was the smallest length?
18	Student	5
Ms. Brown wrote the word "Range" on the board with the numbers 7 and 5 beneath the word.		
19	Ms. Brown	I want to know the difference between these two numbers. What do I do?
20	Student	Add them
21	Ms. Brown	Close

Number	Person	Comment
22	Student	Subtract them
23	Ms. Brown	So, 7 - 5 = 2. Guess what that is.
24	Student	The range.
25	Ms. Brown	I made a mistake. Two what?
26	Student	Inches
27	Ms. Brown	So it is not the number of X's. You just look down here (pointing to the numbers below the line in the lineplot).
28	Student	I do not get how it is 7 and 5.
29	Ms. Brown	We took the greatest number we had and the least.
30	Student	We took the most before.
31	Student	What about the mode?
32	Ms. Brown	Let's talk about the mode.
33	Student	I think I know the mode. I think the mode is the opposite of the range, so it is 7 + 5 = 12.
34	Student	I think when you do the number of X's, the highest amount of X's is 10, so the mode would be 10.
35	Ms. Brown	(Calling on a particular student) What did you say about the mode?
36	Student	It would be the measurement that has the most X's above it.
37	Ms. Brown	What would that be?
38	Student	6 (The student went up to the board and pointed to it.)
39	Ms. Brown	So what is the mode?
40	Student	6
41	Ms. Brown	Did I have to add anything?
42	Student	No
43	Ms. Brown	I just have to say 6 inches. Did I have to count X's?
44	Student	You could just look at it.

Table 2. Birth length lesson transcript

Number	Person	Comment
1	Student	We [a group of students working on analyzing a data set together] disagree how to find the mode and range.
2	Ms. Brown	How are you trying to find the mode?
3	Student	I think you should look at the bottom of the line plot rather than the number of X's. Well I guess you look at the number of X's to see which birth length happened most often. So I think the mode should be 21.
4	Ms. Brown	How are other people in your group finding the mode?
5	Student	They are saying that the mode is the most number of X's, so they say it is 8.
6	Ms. Brown	The mode is the most number of X's, so it is 8. I like your thinking in trying to get 21, but the mode is 8 here. What were you thinking about with the range?
7	Student	Well, I am probably wrong. I was thinking that the range was the difference between the smallest birth length and the biggest birth length. I thought it was 6 (this was from 25 - 19), but they are saying that it is 7 (this was from 8 - 1).
8	Ms. Brown	Yes, it is 7.

Table 3. Timeline for Ms. Brown's misconception regarding the range

Day	Event
0	Professional Development Training introduced the concepts of mode and range. During this training, Ms. Brown originally exhibited the misconceptions regarding the concepts of mode and range.
0	Ms. Brown's misconceptions were corrected by her experiences with the lesson in the professional development training.
2	Ms. Brown was able to correct the same misconceptions she had exhibited during the professional development training when they were exhibited by her students.
8	Ms. Brown reverted to her original misconceptions as more time had passed between the professional development training and her coverage of the mode and range. Early in the day, Ms. Brown taught the students how to find the mode and range of a data set incorrectly, using an example without context. This teaching contradicted the way she originally introduced the concepts to the students on Day 2.
8	Ms. Brown, in error, informed a student that he was incorrect in the manner he found the mode and range. This contradicted the responses she praised 6 days earlier.

		X	
		X	
	X	X	
	X	X	
	X	X	
X	X	X	
X	X	X	
X	X	X	X
X	X	X	X
6	7	8	9

Figure 1. Wrist circumference (in inches)

	X	
	X	
	X	
	X	
	X	
	X	
	X	
	X	X
X	X	X
X	X	X
5	6	7

Figure 2. Wrist circumference (in inches)

			X	
	X		X	
	X		X	
	X		X	
X	X		X	X
X	X	X	X	X
X	X	X	X	X
3	4	5	6	7

Figure 3. Lineplot in Ms. Brown's class

	X				
	X				
	X				
X	X				
X	X				
X	X				
X	X	X			
X	X	X	X		X
20	21	22	23	24	25

Figure 4. Birth length (inches)

Call for Manuscripts

International Journal of Statistics and Probability is a peer-reviewed journal, published by Canadian Center of Science and Education. The scopes of the journal include, but not limited to, the following topic areas: statistics, probability, its education and application which cover economics, management, biology, medicine, ecology, sociology, demography, statistical mechanics, computing, optimization and control, stochastic geometry, and stochastic algorithms. The journal is published in both printed and online versions, and the online version is free access and download.

We are seeking submissions for forthcoming issues. The paper should be written in professional English. The length of 3000-8000 words is preferred. All manuscripts should be prepared in LaTeX format, and submitted online, or sent to: ijsp@ccsenet.org

Paper Selection and Publication Process

a). Upon receipt of paper submission, the Editor sends an E-mail of confirmation to the corresponding author within 1-3 working days. If you fail to receive this confirmation, your submission/e-mail may be missed. Please contact the Editor in time for that.

b). Peer review. We use double-blind system for peer-review; both reviewers and authors' identities remain anonymous. The paper will be peer-reviewed by three experts; two reviewers from outside and one editor from the journal typically involve in reviewing a submission. The review process may take 2-3 weeks.

c). Notification of the result of review by E-mail.

d). The authors revise paper and pay publication fee.

e). After publication, the corresponding author will receive two copies of printed journals, free of charge.

f). E-journal in PDF is available on the journal's webpage, free of charge for download.

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 responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other languages, 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 published in CCSE journals 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.html

The journal is peer-reviewed
The journal is open-access to the full text
The journal is included in:

AMICUS
Canadiana
Google Scholar
Library and Archives Canada
LOCKSS
PKP Open Archives Harvester
Standard Periodical Directory

International Journal of Statistics and Probability

Semiannual

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