PARAMETER ESTIMATIONS AND COPULA METHODS FOR BURR TYPE III AND TYPE XII DISTRIBUTIONS

NOR HIDAYAH BINTI ISMAIL

UNIVERSITI TEKNOLOGI MALAYSIA

PARAMETER ESTIMATIONS AND COPULA METHODS FOR BURR TYPE III AND TYPE XII DISTRIBUTIONS

NOR HIDAYAH BINTI ISMAIL

A report submitted in fulfilment of the

requirements for the award of the degree of

Master of Science (Mathematics)

Faculty of Science

Universiti Teknologi Malaysia

DECEMBER 2014

Specially dedicated to my dear husband

Mohd Yustajudin Bin Mokhtar

my beloved parents,

En Ismail Mohamood

Pn Rusiah Ab Manaf

my siblings, my family

and to all my friends.

ACKNOWLEDGEMENT

First of all, I am very thankful to Almighty Allah for strength granted to complete and finish my research. Many individuals had contributed towards my understanding and thoughts in preparing this thesis for my master. In particular, I would like to give a special thank to my supervisor, Dr. Zarina Mohd Khalid for her encouragement, guidance, critics and advices that she had given to me to complete this research work. Without her continued support and interest, this research would not have been the same as presented here.

I wish to express my gratitude to my husband and family who always give their support to me. I thank all my friends and others who had directly or indirectly help me in this study.

Thank you for the support and encouragement all this while for completing this research for the degree of Master.

ABSTRACT

Continuous Burr distributions have gained popularity recently due to their potential use in practical situations. In particular, Burr Type III and Type XII distributions are suitable to describe lifetime data since these distributions, not only have flexible shape but also controllable scale and location parameters which are needed in characterizing lifetime distributions. In this study, 2-parameter and 3-parameter Burr Type III and XII distributions are employed to fit a set of simulated lifetime data. These lifetime data are assumed to be either complete, that is uncensored, or censored at varying levels of censoring, and are simulated from the specified Burr distributions using their inverse cumulative distribution functions. The distribution parameters are then estimated by using the classical maximum likelihood estimation (MLE) and expectation-maximization (EM) algorithm approaches. The performance of parameter estimates are then compared in terms of their accuracy and efficiency by comparing its bias and mean square errors. The study finds that as the censoring level varies, the EM estimates perform better than the MLE estimates for 2-parameter and 3-parameter Burr Type III and XII distributions with complete and censored lifetime data at certain censoring levels. In addition, the study also investigates a number of copula methods to join specific Burr Type III and XII distributions. The result reveals that Ali-Mikhail-Haq, Clayton and Gumbel methods fit well with Burr distributions for uncensored lifetime data since the values of copula lie within (0,1) interval.

ABSTRAK

Kebelakangan ini, taburan selanjar Burr telah mendapat populariti disebabkan oleh keupayaan menggunakannya dalam situasi praktikal. Secara khususnya, taburan Burr Jenis III dan Jenis XII sesuai untuk menerangkan tentang data hayat masa disebabkan taburan ini bukan sahaja mempunyai parameter bentuk yang fleksibel tetapi juga parameter skala dan lokasi yang boleh dikawal yang diperlukan untuk mencirikan taburan hayat masa. Dalam kajian ini, taburan Burr Jenis III dan XII berparameter 2 dan 3 digunakan untuk disuaikan dengan satu set data simulasi hayat masa. Data hayat masa ini diandaikan sama ada lengkap, iaitu tidak disensor, atau disensor pada pelbagai peringkat sensor dan disimulasi daripada taburan Burr tertentu menggunakan fungsi taburan kumulatif songsang. Parameter taburan kemudiannya dianggarkan dengan menggunakan kaedah anggaran kebolehjadian maksimum (MLE) dan algoritma pemaksimuman jangkaan (EM). Prestasi anggaran parameter kemudiannya dibandingkan dari segi ketepatan dan kecekapan dengan membandingkan kepincangan dan ralat min kuasa dua. Kajian ini mendapati bahawa anggaran parameter daripada kaedah algoritma EM memberikan prestasi yang lebih baik daripada anggaran MLE untuk taburan Burr Jenis III dan XII berparameter 2 dan 3 untuk data hayat masa yang lengkap dan disensor pada tahap sensor tertentu. Di samping itu, kajian ini juga menyiasat beberapa kaedah kopula untuk menghubungkan taburan Burr Jenis III and XII tertentu. Keputusan menunjukkan bahawa kaedah Ali-Mikhail-Haq, Clayton dan Gumbel menghasilkan penyuaian yang baik dengan taburan Burr untuk data hayat masa yang tidak disensor disebabkan nilai-nilai copula berada dalam selang (0,1).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS

$\phi(t)$	Generator of copula
θ	Parameter of copula
λ	Shape parameter
β	Scale parameter
c	Shape parameter
k	Shape parameter
s	Scale parameter
t	Simulated data
U	Uniformly distributed random variable
q(u)	Quantile function
q'(u)	Derivative of quantile function
f(t)	Probability distribution function
F(t)	Cumulative distribution function
S(t)	Survival function
r	Number of failure samples
n	Random sample
χ^2	Chi-square
τ	Kendall's tau
$\phi'(t)$	Generator's first derivative
$\phi^{-1}(t)$	Generator's inverse

u	Random number
v	Random number
$K_C(t)$	Distribution function of copula
$K_C^{-1}(t)$	Distribution function inverse
H_0	Null hypothesis
H_1	Alternative hypothesis
C	Copula

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Burr distribution is a continuous probability distribution for a non-negative random variable. A few forms of cumulative distribution function are suggested to fit the lifetime data when the Burr distribution was first introduced by Burr (1942) in the properties and theory of cumulative function. Indeed, in the past, Burr distribution got more attention from researchers because of its potential in practical situations. The distribution has been practiced in various fields for the purpose of statistical modeling and fitting such as in studies of household income, insurance risk, forestry, fracture roughness, meteorology and reliability. A wide range of skewness and kurtosis which has been generated from the Burr distribution can be used to fit almost any given set of unimodal data (Tadikamalla, 1980).

Burr distribution becomes more popular among the researchers and comes out with many types of distributions such as Burr Type II, Burr Type III, Burr Type X and Burr Type XII distributions with the same or different parameterizations. Burr distribution also appears in the extended parameter of the distributions such as extended three-parameter Burr Type III distribution for low-flow frequency analysis. However in the last few decades, many researchers focused on the Burr Type III and XII distributions due to its pragmatic nature. Burr Type III and XII distributions are used under other names in some cases. Christian Kleiber (2007) in his guide to the Dagum distribution, states that in economics, the Burr Type III distribution is named as Dagum distribution and the Burr Type XII distribution is known as the Singh-Maddala distribution. Meanwhile, in the actuarial literature and meteorological literature, Burr Type III is called an inverse Burr distribution and kappa distribution respectively.

The Burr Type III and XII distributions are more widely known because they include the characteristics of various well known distributions. Tadikamalla in 1980 summarized the relationship between the Burr distribution and the other distributions which include the Lomax, logistic, Weibull and Kappa family of distributions. Then, Lewis (1981) stated that the Burr Type XII distribution had many standard theoretical distributions as special cases or limiting cases, including the Weibull and the exponential distributions which are known as common failure time distributions. The log-logistic distribution is a special case of the Burr Type XII to point out the flexibility of the Burr distribution in modeling failure times (Zimmer et al., 1998).

Burr Type III distribution is widely known and has been applied in studies of income, wage and wealth distributions. The Burr Type III distribution has at least two unknown parameters which include shape, scale and location parameters. The Burr Type XII distribution is also employed in various applications that are business, chemical engineering and medical research. It also has at least two unknown parameters which include shape, scale and location parameters.

Wang and Cheng (2010) stated that several methods have been used to estimate the parameters of the Burr Type III and XII distributions such as Maximum Likelihood Estimation (MLE), Least Squares Estimation (LSE) and Bayesian Estimation. The parameter estimation procedure involves the presence of complete and censored data. MLE is commonly used in reliability analysis. It can be applied to every sort of censored or multi-censored data and provides an ordered approach to parameter estimation problems. The weakness of MLE is that it is very sensitive to initial values. Besides that, according to Wang and Cheng (2010), expectation-maximization (EM) algorithm can overcome this problem because of its robustness against the initial value. In other words, EM algorithm has sturdiness to oppose the initial value and gives the best result of estimated parameter which is approximately closed to the true parameter value.

Nowadays, researchers are interested in using the censoring model to estimate the parameter of the distribution. Censoring has been applied successfully in the fields of biology, genetics, astronomy, economics and psychological experiment for survival analysis models. Different mechanisms can lead to different type of censored data since the censoring have many types such as right-censored, left-censored and randomly censored.

Besides that, copula method is discovered to become a useful method to joint two distributions and is known as dependence functions. It was introduced by A. Sklar (1959) and can connect multivariate distribution function to its univariate marginal distribution. Copula method has many families such as *Gaussian*, *Archimedean* and *Student's t*. Each family of copula has many types of classes with different known generator, $\phi(t)$. It becomes a useful tool to interpret the relationships among multivariate variables and all multivariate distributions can be identified easily in terms of suitable copulas. Copula is an important method to describe the dependency between random variables, which means that different copulas representing different dependency since it has many families. Copula depends on the parameter θ. In fact, copula function does not focus on correlation coefficient but on scale invariant measures of association, Kendall's rank correlation measures which are functions of the parameter θ (Flores, 2008). In some ways, copula is independent of the univariate marginal and represents the dependence structure separately.

1.2 Statement of the Problem

The characteristics and limitations of the 2-parameter and 3-parameter Burr Type III and Burr Type XII distributions are investigated which involves the process of simulating and estimating the unknown parameters. Due to the absence of real data for the current study, this research alternatively uses simulated lifetime data for complete and censored at varying levels of censoring. Since the cumulative distribution function (cdf) is strictly monotonic increasing function, the lifetime data is simulated using inverse cumulative distribution function at different censoring levels.

Shape parameter is used in lifetime data to characterize the shape of the distribution. However, scale parameter is needed since it is the bulk of the distribution. Generally, scale parameter stretches or squeezes the graph of distribution. The 2-and 3-parameter (consist of shape and scale parameters) of Burr Type III and Burr Type XII distributions are estimated using Maximum likelihood estimation (MLE) approach. The estimated scale parameter tends to larger than the true parameter value in the presence of censoring. Therefore, expectation-maximization (EM) algorithm is considered to overcome the problem. The two estimation methods have different procedures which are MLE is maximized the likelihood function whereas EM algorithm has two steps, E and M steps, to estimate the parameter. Therefore, the comparison is important to assess the performance of two estimation methods.

Burr Type III and Type XII are two continuous distributions. Copula method is considered to assess the relation between the two distributions since copula is a convenient statistical method of measuring the correlation between variables by just considering the marginal distributions of these variables. Therefore, copula will be used to join these two distribution functions. It can be developed in the analysis since it does not matter whether the variable are independent or dependent with association parameter, θ. Three types of copula which are Ali-Mikhail-Haq (AMH), Clayton and Gumbel with their own known generator $\phi(t)$ are identified in this research for joining specific Burr Type III and Type XII distributions.

The research questions that are addressed in this study are summarized as follows:

a) How to simulate the complete and censored data based on the statistical model of 2- and 3-parameter Burr Type III and XII distributions?

b) How to estimate the unknown parameters in 2- and 3-parameter Burr Type III and XII distributions using Maximum Likelihood Estimation (MLE) and Expectation Maximization (EM) algorithm approaches in the presence of various level of censoring?

c) What are the performance of MLE and EM algorithm estimates in terms of their bias and mean square error?

d) What are the best copula methods for joining specific Burr Type III and Type XII distributions?

1.3 Objectives of the Study

The objectives of this study are as follows:

1.3.1 To simulates the complete and censored data based on the statistical model of 2- and 3-parameter Burr Type III and XII distributions.

1.3.2 To estimate the unknown parameters in 2- and 3-parameter Burr Type III and XII distributions using Maximum Likelihood Estimation (MLE) and Expectation Maximization (EM) algorithm approaches in the presence of various level of censoring.

1.3.3 To compare the results of estimated parameter that is obtained from Maximum Likelihood Estimation (MLE) and Expectation Maximization (EM) algorithm approaches for complete and censored cases.

1.3.4 To identify three types of copula which are Ali-Mikhail-Haq (AMH), Clayton and Gumbel for joining specific Burr Type III and Type XII distributions.

1.4 Scope of the Study

Twelve different forms of cumulative distribution functions are introduced by Burr (1942) for modeling data but among those distributions, Burr Type XII distribution received more popularity because of its advantages. One of the advantages is Burr Type XII distribution covers all larger regions of kurtosis towards the lower end of the skew and kurtosis boundary. Thus, the distribution gained interest from researchers to use in many fields for practice or application processes. However, this study will focus on the characteristics of 2- and 3-parameter Burr Type III and XII distributions including simulation and estimation processes based on a statistical model of the distributions.

From the various methods that have been used to estimate the parameter, this study will focus on the application of Maximum Likelihood Estimation (MLE) and Expectation-Maximization (EM) algorithm approaches in estimating the 2- and 3-parameter Burr Type III and XII distributions using complete and censored data. The technique or method of estimating the parameter is compared for complete and censored cases between two approaches which are MLE and EM algorithm. This study focuses on determining the closest estimated parameter value to the true value using MLE and EM algorithm methods for complete and censored data.

Besides that, Archimedean copula has been employed in this research to join the Burr Type III and Type XII distributions since it is easy to construct among the family of copula due to many parametric families that belong to this class. Archimedean copula is also popular because it allows modeling dependence in arbitrarily high dimensions with only one parameter, governing the strength of dependence. Archimedean has six classes which are Ali-Mikhail-Haq (AMH), Clayton, Frank, Gumbel, Independence and Joe copula. However, this work will identify Ali-Mikhail-Haq (AMH), Clayton and Gumbel copula for joining specific Burr Type III and Type XII distributions and verify whether the specific Burr Type III and Type XII distributions fit well with copula distribution.

1.5 Significance of the Study

The process of analysis in this study involves the characteristics of Burr Type III and XII distributions, Maximum Likelihood Estimation (MLE) and Expectation-Maximization (EM) algorithm approaches and copula method. The research work leads us to develop the methodology for simulating complete and censored data of 2- and 3-parameter Burr Type III and XII distributions based on the characteristics and the derivation of parametric forms of probability density function *(pdf)* and cumulative density function *(cdf)* of the distributions. The research work develops the simulation process for censored data and estimates the 3-parameter Burr Type XII distribution for censored data and 2- and 3-parameter Burr Type III distribution since other researcher always focused on 2-parameter Burr Type XII distribution.

On the other hand, the results of this study can give statistical evidence in choosing the estimation approaches that are more accurate to estimate the parameters of 2- and 3-parameter Burr Type III and XII distributions using complete and censored data. This is very important because the estimators will influence the performance of the distribution and will give the benefit to the fields that apply Burr Type III and Type XII distributions in their research project.

Besides the simulation and estimation processes, this study is an example for joining distribution between Burr Type III and Type XII distributions to obtain the copula distribution since there do not have any previous study that join this two distribution using specific parameter value. The process can be applied in real data to investigate the relationship between the distributions such as competing risk situation.

1.6 Research Layout

Chapter 1 introduces the terminology and framework of the study. In this chapter, the distribution and method that will be used in this research are introduced.

Chapter 2 presented about the brief review on the distributions, method of estimation and copula. This chapter also gives a review of censoring and survival function.

Chapter 3 described about the distribution that will be used in this research and the methodology for simulating processes of the distribution which involved an inversion method to simulate the data. The process of estimating the parameters of the distributions using two approaches that are Maximum Likelihood Estimation and EM algorithm for complete and censored data are also showed in this chapter.

All the results and finding of estimated parameter of the Burr Type III and XII distributions using different censoring level for Maximum Likelihood Estimation and EM algorithm will be discussed in Chapter 4. Besides that, the goodness of fit and the result of copula are also included in this chapter. The result of copula is also described in this chapter. Finally, the conclusion of the research project and suggestion for further research are described in chapter 5.

REFERENCES

Abd-Elfattah, A. M. and Al-Harbey, A. H. (2012). Bayesian Estimation for Burr Distribution Type III Based on Trimmed Samples. *ISRN Applied Mathematics*, 2012:1-18

Ahmad, K. E., Fakhry, M. E. and Jaheen, Z. F. (1997). Empirical Bayes estimation of P(Y < X) and characterization of Burr-type X model. *Journal of Statistical Planning and Inference*, 64:297-308.

Ait-Sahalia, Y. (2002). Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: a Closed-Form Approximation Approach. *Econometrica*, 70:223-262.

Allison, P. D. (2012) Handling Missing Data by Maximum Likelihood. K*eynote presentation at the SAS Global Forum*. April 23, Orlando, Florida

Alsina, C., Nelsen, R. B., and Schweizer, B. (1993). On the characterization of a class of binary operations on distribution functions. *Statist. Probab. Lett.*,17:85-89.

Al-Hussaini, E. K. (1991). A Characterization of Burr XII Distribution. *Applied Mathematics Letters*. 1(4):59-61.

Blimes, J. A. (1998). Maximum Likelihood from Incomplete Data via the EM Algorithm. International Computer Science Institute. Computer Science Division, Department of Electrical Engineering and Computer Science.

Burr, I. W. (1942). Cumulative frequency functions. *Annals of Mathematical Statistics*. 13:215-232.

Chang, S. C. and Kim, H. J. (2007). EM Algorithm. Department of Statistics and Actuarial Science. University of Iowa.

Dellaert, F. (2002) The Expectation Maximization Algorithm. GVU Center. College of Computing; Georgia Tech, GIT-GVU-02-20, 2002

Dempster, A. P., Laird, N. M. And Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*. Series B (Methodological). 1(39):1-38.

Domma, F. (2010). Some Properties of the Bivariate Burr Type III Distribution. *Statistics: A Journal of Theoretical and Applied Statistics*, 44 (13):203-215.

Eddie, K. H. Ng (2010). Kernel-Based Copula Processes. Doctor Philosophy, University of Toronto

El-Bassiouny, A. H. and Abdo, N. F. (2010). Reliability Properties of Seven Parameters Burr XII Distribution. *Computational Methods in Science and Technology*. 16(2):127-133.

Fisher, N. I. (1997). Copulas. In:*Encyclopedia of Statistical Sciences*.1:159-163. John Wiley Sons, New York.

Galambos, J. and Kotz, S. (1978). Characterizations of probability distribution. *Springer-Verlag*. 1(2):51-53.

Ghitany, M. E. and Al-Awadhi, S. (2002). Maximum likelihood estimation of Burr XII distribution parameters under random censoring. *Journal of Applied Statistics.* 29(7): 955-965.

Hao, Z. and Singh, V. P. (2009). Entropy-Based Parameter Estimation For Extended Three-Parameter Burr III Distribution For Low-Flow Frequency Analysis. *American Society of Agricultural and Biological Engineers*. 52(4): 1193-1202.

Hatke, M. A. (1949). A Certain Cumulative Probability Function. *Annals of Mathematical Statistics*. 20:461-63.

Headrick, T. D., Mohan, D. P. And Sheng, Y. (2010). On Simulating Univariate and Multivariate Burr Type III and Type XII Distributions. *Applied Mathematical Sciences*. 45(4):2207-2240.

Jaheen, Z. F. and Al-Matrafi, B. N. (2002). Bayesian prediction bounds from the scaled Burr type X model. *Computers and Mathematics with Applications*. 44:587-594.

Jeff Wu, C. F. (1983). On the Convergence Properties of the EM Algorithm. *The Annals of Statistics*, 11(1):95-103.

Joe, H. (1997). Multivariate Models and Dependence Concepts. *Monographs on Statistics and Applied Probability,* 73. Chapman &Hall, London.

Kleiber, C. (2007). A Guide to the Dagum Distributions. Chotikapanich, D. *Modeling Income Distributions and Lorenz Curves*. pp 97-117. London: Springer-Verlag

Li, X. (2000). On Default Correlation: A Copula Function Approach. *Journal of Fixed Income*, 9(4):43–54.

MacKenzie, D. and Spears, T. (2012). 'The Formula That Killed Wall Street'? The Gaussian Copula and the Material Cultures of Modelling. University of Edinburgh

Miura, K. (2011).An Introduction to Maximum Likelihood Estimation and Information Geometry. *Interdisciplinary Information Sciences*. 17(3):155–174

Nielsen , S. F. (2000). On simulated EM algorithms. *Journal of Econometrics*, 96:267-292.

Olapade, A. K. (2008). On a six-parameter generalized Burr XII distribution. arXiv:0806.1579v1 [math.ST]

Rastogi, M. K and Tripathi, Y. M. (2011). Inference on unknown parameters of a Burr distribution under hybrid censoring. *International Journal of Quality & Reliability Management*, 28 (8):885-893

Rodriguez, R. N. (1977). A guide to the Burr Type XII distributions. *Biometrika*. 64:129-134.

Schweizer, B. and Wolff, E. (1981). On nonparametric measures of dependence for random variables. The Annals of Statistics, 9:879-885.

Shao, Q. (2004). Notes on maximum likelihood estimation for the three-parameter Burr XII distribution. *Computational Statistics and Data Analysis*. 45:675-687.

Shao, Q., Chen, Y. D. and Zhang, L. (2008). An extension of three-parameter Burr III distribution for low-flow frequency analysis. *Computational Statistics & Data Analysis*. 52:1304 – 1314.

Shawky, A. I. and Al-Kashkari, F. H. (2007). On a stress strength model in Burr of type III. *International Journal of Statistics*. 3(LXV):371-385.

Sklar, A. (1959). Fonctions de repartition a *n* dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229-231.

Stigler, S. M. (2007). The Epic Story of Maximum Likelihood. *Statistical Science* 22(4):598-620.

Tadikamalla, P. R. (1980). A look at the Burr and Related Distributions. *International Statistical Review*. 48:337-344.

Tejeda, H. A. and Goodwin, B. K. (2008). Modeling Crop prices through a Burr distribution and Analysis of Correlation between Crop Prices and Yields using a Copula method. *Paper presented at Agricultural and Applied Economics Association.Annual Meeting*. 27-29 July Orlando, Florida, 1-39

Wang, F. K. and Cheng, Y. F. (2010). EM Algorithm for Estimating the Burr XII Parameters with Multiple Censored Data. *Quality and Reliability Engineering International*. 26:615-630.

Wang, F. K. and Cheng, Y. F. (2012). Estimating the Burr XII Parameters in Constant-Stress Partially Accelerated Life Tests under Multiple Censored Data. *Communications in Statistics-Simulation and Computation*, 41: 1711-1727.

Wang, F. K., Keats, J. B. and Zimmer, W. J. (1996). Maximum likelihood estimation of the Burr XII parameters with censored and uncensored data. *Microelectronics and Reliability*. 36(3):359-362.

Watkins, A. J. (1999). An algorithm for maximum likelihood estimation in the three parameter Burr XII distribution. *Computational Statistics and Data Analysis*. 32(1):19-27.

Wingo, D. R. (1993). Maximum likelihood methods for fitting the burr type XII distribution to multiply (progressively) censored life test data. *Metrika*. 40:203-210.

Zoraghi, N., Abbasi, B., Niaki, S. T. A and Abdi, M. (2012). Estimating the four parameters of the Burr III distribution using a hybrid method of variable neighborhood search and iterated local search algorithms. *Applied Mathematics and Computation*. 218:9664–9675.