THE CRAMER-VON MISES TEST STATISTIC OF THE GENERALIZED EXTREME VALUE DISTRIBUTION AND WEIBULL DISTRIBUTION

NUR HAMIZAH BINTI ABDUL GHANI

A thesis submitted in fulfillment of the requirements for the award of the Master of Science (Mathematics)

Faculty of Science
Universiti Teknologi Malaysia

JANUARY 2014

Specially dedication for,

To my beloved father: Abdul Ghani bin Sulaiman

To my beloved mother: Suaibah binti Muhamad Saleh

To my supervisor: Dr. Ani Shabri

To my beloved sister and brother: Hazim, Haziq, Hazrin,Hazirah, Hannani, Hazwani

and

All my friends

Thanks for everything.

ACKNOWLEDGEMENT

First of all, praise and thanks to Allah the Beneficients, the Merciful for hearing what I wanted and giving me the strength that I needed in order to complete this research on time.

A special thanks to my supervisor, Dr. Ani Shabri for her coaching, guidance and support during this research being carried out.

I would like to thank each and every individual who have either directly or indirectly helped me throughout the efforts of this report in the form of encouragement, advice or kind reminders. I also thank my friend, Hafidzatun Nadwa for her valuable editing and suggestions.

Lastly, I would like to thank my incredible family for endless support and belief in all my adventures. Without all the helps provided, this research might not proceed smoothly.

May Allah bless you.

Thank You.

ABSTRACT

Goodness of fit (GOF) test is a statistical technique in selection of an appropriate probability distribution for a given sample data. This study use a goodness of fit test, which is the Cramer Von Mises (CVM) test, in order to analyze the GEV distribution and Weibull distribution that best fit to model the annual rainfall data in Peninsular Malaysia. There are two objectives covered in this study, which are to develop the calculation critical values of CVM test statistic for GEV distribution and Weibull distribution by using simulation experiment, and to compare the accuracy of CVM test between both distributions in real data. The result of goodness of fit test using the CVM test statistic shows that the Weibull distribution has the greater rejection power at three different significance levels compare to the GEV distribution. It can be conclude that, the result of goodness of fit test using CVM test for this rainfall data in Peninsular Malaysia is most appropriate by using the Weibull distribution to fit the model of this rainfall data.

.

ABSTRAK

Pengujian kesahihan pemadanan adalah satu teknik statistik yang boleh digunakan untuk menentukan taburan kebarangkalian stastik untuk sampel data yang diberikan.Kajian ini menggunakan satu pengujian kesahihan pemadanan iaitu ujian Cramer-Von Mises (CVM), untuk menganalisis taburan Generalized Extreme Value (GEV) dan taburan Weibull, bagi memodelkan data tahunan hujan di Semenanjung Malaysia. Terdapat dua objektif dalam kajian ini, iaitu membangunkan nilai-nilai kritikal bagi ujian CVM bagi kedua-dua taburan GEV dan taburan Weibull dengan menggunakan eksperimen simulasi. Hasil keputusan Pengujian kesahihan pemadanan menggunakan ujian CVM, menunjukkan bahawa taburan Weibull mempunyai kuasa penolakan yang lebih besar pada tiga peringkat penting berbanding taburan GEV. Kesimpulannya, hasil keputusan Pengujian kesahihan pemadanan menggunakan ujian CVM untuk data tahunan hujan di Semenanjung Malaysia adalah lebih sesuai menggunakan taburan Weibull untuk memodelkan data taburan hujan ini.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AD	Anderson Darling
CVM	Cramer Von Mises
GEV	Generalived Extreme Value
GOF	Goodness of fit
KS	Kolmogorov-Smirnov
MAD	Modified Anderson Darling
MLE	Maximum Likelihood Estimation
MOM	Method of Moments
PWM	Probability Weighted Moments

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

One of the basic problems in statistic is the selection of an appropriate probability distribution for a given sample data. The key in selecting the distribution model is generally based on goodness of fit (GOF) tests. GOF test is a statistical procedure to examine how well a sample data agrees with an empirical frequency distribution and the assumed probability distribution. In other words, it involves observing a random sample from a distribution in order to test the null hypothesis of that distribution is in fact. The problem can be formulated under the test hypothesis involving $H_0 : F(y) = F_0(y_{(i,n)})$ where F_0 the hypothesized continuous random variable with known parameter against is $H_1 : F(y) \neq F_0(y_{(i,n)})$.

GOF test is conducted for modeling and decision making process. Any prediction and conclusion to be drawn in any research depend very much on the selection of the distribution. Choosing an inappropriate model can make the analysis become meaningless or useless. Thus, it is important to select appropriate distribution to avoid wrong decision and interpretation of the data.

Many researchers such as Tamer (1993), Shin (2012) and Abidin (2012) have focused on the power study of the GOF test for a variety of distributions to assess the performance and efficiency several of goodness of fit tests which are Anderson Darling (AD) test, Modified Anderson Darling (MAD) test, Cramer-von Mises (CVM) test and Kolmogorov-Smirnov (KS) test. These methods evaluate the goodness of fit test by comparing the assumed cumulative density function to the empirical density function or comparing the assumed cumulative density function to the empirical density function.

This study only focuses on one method of goodness of fit test, which is the CVM test statistic to test the best distribution for a sample data. This study applies two distribution, which are Generalized Extreme Value (GEV) distribution and Weibull distribution, in order to know which distribution show the best fit to the data. The GEV distribution is a family of continuous probability distributions which develops within extreme value theory. Extreme value theory provides theory of modelling and it measures the extreme events which occur in a very small probability. Meanwhile, the Weibull distribution is one of continuous probability distribution, that come from the class of GEV distribution. There are many applications of GEV distribution and Weibull distribution has been found especially in hydrology. It has been widely used to model variety of natural extremes such as floods, wind speeds, rainfall, wave height and others.

"Parameter estimation is a discipline that provides tools for the efficient use of data in the estimation of constants appearing in mathematical model to aid in the modelling of phenomena", Tamer (1993). Based on the previous research, there are several possible methods to estimate the parameters of each distribution such as method of moments (MOM), method of maximum likelihood (MLE) and Probability Weighted Moments (PWM). In reference to the same study, Martins (2000) recommends to use PWM for the parameter estimation as it is more stable and unbiased compared to other methods. Hosking (1985) also suggests using PWM as it is fast and straightforward to compute and always yield feasible values for the estimated parameters. Thus, PWM method is used as the preferred method of parameter estimation for this study.

Furthermore, the critical value of goodness of fit test plays role of reference value in deciding which distribution fits with the data. Thus, the availability of software package which is MATHCAD software could helps in developing calculation of critical value of each test. Then, other mathematical simulations can also be employed to produce the statistic value for each GOF test.

1.2 Statement of Problem

Goodness of fit (GOF) test is a test that widely used in statistic, to select an appropriate probability distribution for a sample data. There are a lot of GOF tests that can be used. To the best of my knowledge, this study will focus on one of the GOF test, which is the CVM test statistic.. CVM test is a most powerful test against a large class of alternative hypotheses. “They all involve integration over the whole range of data, rather than use of a supremum, so they are best-suited for situations where the true alternative distribution deviates a little over the whole support rather than having large deviations over a small section of the support (Arnold and Emerson, 2011)”.

This dissertation explores the goodness of fit test for CVM test in GEV distribution and Weibull distribution. Many research about the goodness of fit test for GEV distribution and Weibull distribution has been done. But, there is still discrepancies and lack in published the critical values for both distributions. So, the critical values of both distributions are developed in this study. The result of this study will be used to evaluate the CVM test of GEV distribution and Weibull distribution to a sample data. Furthermore, in order to test the suitable distribution function, the researchers would encounter the problem of getting the table of critical value for each distribution. Critical value table is needed to analyse how well a goodness of fit test against specific alternative distribution. So, this study develops formulas and mathematical calculation of critical values by using MATHCAD software in computing goodness of fit test for the GEV distribution and Weibull distribution.

1.3 Research Objectives

The objectives of this study are:

1.3.1 To develop calculation critical values of CVM test statistic for GEV distribution and Weibull distribution.

1.3.2 To compare the accuracy of CVM test between GEV distribution and Weibull distribution in real data.

1.4 Scope of the Study

This study focuses on the CVM test for GEV distribution and Weibull distribution. The probability weighted moments or L-moments method is used to obtain the parameters estimation of these distributions. The mathematical calculation of critical value is developed using simulation experiment. Then, a set of rainfall data is applied in order to evaluate the assessment of CVM goodness of fit test. In order to know which distribution best fit to the rainfall data, the power test using CVM test statistic is performed.

1.5 Significance of the Study

Study of goodness of fit test for extreme events is useful to determine the distribution function in modeling extremes of natural phenomena such as floods, wind speeds and rainfall. This study develops formulas and mathematical calculation of critical values by using MATHCAD software in computing the CVM test statistic for two distributions, which are the GEV distribution and Weibull distribution. So, this research helps other researchers in computing critical values for CVM test in both distributions. Furthermore, a rainfall data is applied to this study in order to know which distribution fit the rainfall data. Thus, this study is useful to know which distribution best fit the rainfall data.

1.5 Research Organization

This final year project is divided into five chapters, which are:

Chapter 1 briefly discusses the introduction, background of the study, research objectives, scope of the study, and significance of the study.

Chapter 2 explores on the literature review of goodness of fit for CVM test statistic. This chapter also discusses the literature review of GEV distribution and Weibull distribution, L-Moments parameter estimation and the application of both distributions.

Chapter 3 describes the methodology of this study, which involves the GEV distribution and Weibull distribution, the parameter estimation of L-moment method, the properties of CVM goodness of fit test and the power of test for this study.

Chapter 4 develops the completed critical values table of CVM test statistic for GEV distribution and Weibull distribution, which are developed from simulation experiments

Chapter 5 is about data analysis. In this research, a set of data on annual maximums of rainfalls data for seventy five stations in peninsular Malaysia is applied in illustrating the application of GEV distribution and Weibull distribution by using the L-moments approach. Then, power test is done to compare the performance of test statistic for the sample data between both distributions.

Lastly, Chapter 6 summarizes and concludes the whole study with some suggestions for further research.

REFERENCES

Ahmad, M. I., Sinclair, C. D., and Spurr, B. D., *Assessment of Flood Frequency Models Using Empirical Distribution Function Statistics*, Water Resources Research, August *1988*, vol. 24, No. 8: 1323-1328.

Arnold, T. B. and Emerson, J. W., *Nonparametric Goodness of Fit Test for Discrete Null Distributions*, The R Journal, Dec 2011, vol. 3/2.

Awale, D., *Comparison of Probability Distribution for the use in Flood Frequency Analysis at Selected Rivers in Bangladesh*, Asian University for Women, 2013.

Bhattarai, K.P., *An investigation of the use of the partial L-moments for analysing censored flood samples*, National University of Ireland, 2002.

Coppa, M. C., *A New Goodness of Fit Test for the Weibull Distribution Based On Spacing*, Air Force Institute of Technology, 1993.

Dauda, T. O, Asiribo, O. E., Akintoye, N. A., *On the Estimation of a Scaled of Rainfall Data of South West Nigeria*, International Journal of Modern Mathematical Sciences, 2013,7(2):176-184.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R., *Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressable in Inverse Form*, Water Resources, 1979, vol. 15: 1049-1054

Gubareva, T. S. and Gartsman, B. I., *Estimating Distribution Parameters of Extreme Hydrometeorological Characteristics by L-Moment Method*, Water Resources, 2010, vol. 37, no 4.

Hanson, L. R., *The Probability Distribution of Daily Rainfall in the United States*, Department of Civil and Environmental Engineering , Tufts University, 2003.

Hosking, J. R. M., *Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments*, American Statistical Association and the American Society for Quality Control, 1985, vol.27, no 3.

Hosking, J. R. M., L-Moments: *Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics*, J. Royal Statistical Soc. Ser. B, 1990, vol. 52:105-124.

Jamaluddin, S. and Jemain, A. A., *Fitting Daily Rainfall Amount in Peninsular Malaysia Using Several Types of Exponential Distribution*, Journal of Applied Science Research, 2007.

James, *et al*, *Two and Three Parameter Weibull Goodness of Fit Tests*, Department of Agriculture, United States, 1989.

Justus, C. G., Hargraves, W. R., Mikhail, A., and Graber, D., *Methods for estimating Wind Sped Frequency Distribution*, Journal of Apllied Methodology, 1977, vol. 7.

Laio, F., *Cramer-Von Mises and Anderson-Darling goodness of fit tests for exteme value distributions with unknown parameters*, Water Resources Research, 2004, vol. 40.

Liliefors, H. W., *On the Kolmogorov –Smirnov test for the exponential distribution with mean unknown*, Journal Statistic Association, 1969, vol. 64: 387-389

Martins, E. S. and Stedinger, J. R., *Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data*, Water Resources Research, 2000, vol. 36:737-744.

Pocherich, M. J., *Application of Extreme Value Theory and Threshold Models to Hydrological Events, University of Colarado*, 2002.

Shabri, A. and Jemain, A. A., *Goodness of fit test for extreme value Type I Distribution*, Matematika, 2009, vol. 25:53-66.

Shin, J., Jung, Y., Jeong, C. and Heo, J.H., *Assessment of modified Anderson-Darling test statistics for the generalized extreme value and generalized logistic distributions*, Stoch Environ Res Assess. 2012, vol. 26 :105-114.

Stedinger, J. R. and Lu, L. H., *Goodness of Fit Tests for Regional Generalized Extreme Value Flood Distributions*, Water Resources Research, 1991, vol. 27, no. 7:1765-1776.

Tamer, O., *A Modified Anderson Darling Goodness of Fit Test for the Gamma Distribution with unknown Scale and Location Parameters*, Air Force Institute of Technology, 1993.

Wang, H. M., *Comparison of the Goodness of Fit Tests: the Pearson Chi-square and Kolmogorov-Smirnov Tests*, Ling Tung University.

Yusof, F., *Use of Statistical Distribution for Drought Analysis, Applied Mathematical Sciences*, 2012, vol. 6:1031-1051.