PARAMETRIC MIXTURE MODEL OF THREE COMPONENTS FOR MODELLING HETEROGENEOUS SURVIVAL DATA

YUSUF ABBAKAR MOHAMMED

DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2015

Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok

Abstrak

Kajian yang lepas menunjukkan model kemandirian campuran dua komponen mencatatkan prestasi yang lebih baik berbanding model kemandirian berparameter klasik tulen. Namun terdapat juga keperluan yang penting bagi model kemandirian campuran tiga komponen kerana tingkah laku data kemandirian heterogen yang lazimnya merangkumi lebih dari dua taburan. Oleh itu dalam kajian ini dua model bagi tiga komponen telah dibina. Model 1 adalah model kemandirian campuran berparameter tiga komponen bertaburan Gamma dan Model 2 adalah model kemandirian campuran berparameter tiga komponen bertaburan Eksponen, Gamma dan Weibull. Kedua-dua model telah dianggar menggunakan Pemaksimuman jangkaan (EM) dan pengesahan prestasi model melalui kajian simulasi dan empirikal. Simulasi telah diulang 300 kali dengan mengambil kira tiga saiz sampel berbeza: 100, 200, 500; tiga peratus penapisan yang berbeza: 10%, 20%, 40%; dan dua set kebarangkalian bercampur secara: menaik (10%, 40%, 50%) dan secara menurun (50%, 30%, 20%). Beberapa set data sebenar telah digunakan dalam kajian empirikal dan perbandingan model-model telah dilaksanakan. Model 1 telah dibandingkan dengan model kemandirian berparameter klasik tulen, model kemandirian berparameter campuran dua dan empat komponen bertaburan Gamma. Model 2 telah dibandingkan dengan model kemandirian berparameter klasik tulen dan model kemandirian berparameter campuran tiga komponen bertaburan sama. Persembahan grafik, *log likelihood* (LL), Kriteria Maklumat Akaike (AIC), Min Ralat Kuasa Dua (MSE) dan Punca Min Ralat Kuasa Dua (RMSE) telah digunakan bagi menilai prestasi. Dapatan simulasi menunjukkan bahawa kedua-dua model mencatatkan prestasi yang baik pada saiz sampel yang besar, peratus tertapis yang kecil dan pada kebarangkalian bercampur secara menaik. Kedua-dua model menghasilkan ralat yang kecil berbanding dengan model kemandirian jenis lain dalam kajian empirikal. Ini menunjukkan bahawa kedua-dua model yang dibina adalah lebih tepat dan merupakan pilihan yang lebih baik untuk menganalisis data kemandirian heterogen.

Kata kunci: data survival, heterogen, tiga komponen, eksponen, Gamma, Weibull, Pengmaksimuman Jangkaan

Abstract

Previous studies showed that two components of survival mixture model performed better than pure classical parametric survival model. However there are crucial needs for three components of survival mixture model due to the behaviour of heterogeneous survival data which commonly comprises of more than two distributions. Therefore in this study two models of three components of survival mixture model were developed. Model 1 is three components of parametric survival mixture model of Gamma distributions and Model 2 is three components of parametric survival mixture model of Exponential, Gamma and Weibull distributions. Both models were estimated using the Expectation Maximization (EM) and validated via simulation and empirical studies. The simulation was repeated 300 times by incorporating three different sample sizes: 100, 200, 500; three different censoring percentages: 10%, 20%, 40%; and two different sets of mixing probabilities: ascending (10%, 40%, 50%) and descending (50%, 30%, 20%). Several sets of real data were used in the empirical study and models comparisons were implemented. Model 1 was compared with pure classical parametric survival model, two and four components parametric survival mixture models of Gamma distribution, respectively. Model 2 was compared with pure classical parametric survival models and three components parametric survival mixture models of the same distribution. Graphical presentations, log likelihood (LL), Akaike Information Criterion (AIC), Mean Square Error (MSE) and Root Mean Square Error (RMSE) were used to evaluate the performance. Simulation findings revealed that both models performed well at large sample size, small percentage of censoring and ascending mixing probabilities. Both models also produced smaller errors compared to other type of survival models in the empirical study. These indicate that both of the developed models are more accurate and provide better option to analyse heterogeneous survival data.

Keywords: survival data, heterogeneous, three components, Exponential, Gamma, Weibull, Expectation Maximization.

Acknowledgement

All praise and gratitude are due to Allah (SWT) the Lord of heavens for providing me the strength and sparing my life to witness the completion of this work. I would like to express my endless thanks to my wonderful supervisors Associate Professor Dr Bidin Yatim and Dr Suzilah Ismail, for their inspiration, motivation, guidance, and commitment. Without their supervision this thesis would not have been completed to this level. Once again, I express my deep gratitude and appreciation. Special thanks go to UUM CAS staff, especially the staff of the School of Quantitative Sciences (SQS) for their kind assistance whenever needed. My special thanks also go to my postgraduate colleagues both the local and international students from different countries which include Nigeria, Yemen, Iraq, Jordan and Sudan to mention a few, for the good spirit of cooperation, motivation and encouragement among them, and wish them all the best. My appreciations go to the Nigerian community for their cooperation and encouragement. Great appreciation and gratitude go to my mother Maryam Mohammed and my father Abbakar Mohammed for their continued prayers and inspiration. Thanks to my brothers and sisters, for their continued encouragement. Thanks and appreciations go to my loving wife, Dr Zainab Abdallah for her prayers, encouragement and patience. My thanks to my wonderful children Mohammed, Abubakr, Rayyan and Abdallah for their inspiration and patience. Finally, I would like to thank and express my appreciation to all those who in one way or the other contributed to the completion of this work and I could not mention their names, but still they are in my heart. Thank you very much.

Table of Contents

Permission to Use i
Abstrak ii
Abstract iii
Acknowledgement iv
Table of Contents v
List of Tables viii
List of Figures xi
List of Appendices xv
Glossary of Terms xvi
List of Abbreviations xvii

CHAPTER ONE INTRODUCTION 1

1.1 Background of the Study 1
1.2 Problem Statement 2
1.3 Objectives of the Study 6
1.4 Significance of the Study 7
1.5 Outline and Summary of the Thesis 8

CHAPTER TWO REVIEW OF THE LITERATURE 10

2.1 Introduction 10
2.2 Censoring 14
    2.2.1 Some Types of Right Censoring 15
2.3 Parametric Methods in Survival Analysis 16
    2.3.1 The Exponential Distribution 16
    2.3.2 The Gamma Distribution 20
    2.3.3 The Weibull Distribution 22
    2.3.4 The Lognormal Distribution 24
    2.3.5 The Gompertz Distribution 26
    2.3.6 Summary of Parametric Methods in Survival Analysis 28
2.4 Non-Parametric Methods in Survival Analysis 29
2.5 Mixture Models 31
2.6 Three Components Parametric Survival Mixture Models of Same Distribution 38

2.7 Three Components Parametric Survival Mixture Model of Different Distributions 39
2.8 Summary 41
CHAPTER THREE METHODOLOGY 43
3.1 Introduction 43
3.2 Development of Model 1 and Model 2 46
3.2.1 Validating the Performance of the Models Using Simulated Data 47
3.2.2 Validating the Models Using Real Data 52
3.3 Summary 56
CHAPTER FOUR THREE COMPONENTS PARAMETRIC SURVIVAL MIXTURE MODEL OF THE GAMMA DISTRIBUTIONS 58
4.1 Introduction 58
4.2 Theoretical Development of Model 1 58
4.2.1 General Formulation of Three Component Parametric Survival Mixture Model 59
4.2.2 Estimation of Parameters Using the EM 61
4.2.3 EM for Parametric Survival Mixture Model 66
4.3 Validation of the performance of Model 1 Using Simulated Data 71
4.3.1 Model 1 with Mixing Probabilities in Ascending Order 71
4.3.1.1 Sample of Size 100 Observations 72
4.3.1.2 Sample of Size 200 Observations 78
4.3.1.3 Sample of Size 500 observations 83
4.3.2 Model 1 with Mixing Probabilities in Descending Order 88
4.3.2.1 Sample of Size 100 observations 88
4.3.2.2 Sample of Size 200 observations 94
4.3.2.3 Sample of Size 500 observations 100
4.3.3 Special Case of Model 1 105
4.4 Validating Model 1 Using Real Data 107
4.4.1 Bone Marrow Transplant Data 107
4.4.2 Kidney Catheter Data 113
4.4.3 The Special Case of Model 1 119

4.5 Summary 121

CHAPTER FIVE THREE COMPONENTS PARAMETRIC SURVIVAL MIXTURE MODEL OF THE EXPONENTIAL, GAMMA AND WEIBULL DISTRIBUTIONS 123

5.1 Introduction 123

5.2 Theoretical Development of Model 2 123

5.3 Validation of the performance of Model 2 Using Simulated Data 131

5.3.1 Model 2 with Mixing Probabilities in Ascending Order 131

5.3.1.1 Sample of Size 100 observations 132

5.3.1.2 Sample of Size 200 observations 138

5.3.1.3 Sample of Size 500 observations 143

5.3.2 Model 2 with Mixing Probabilities in Descending Order 148

5.3.2.1 Sample of Size 100 observations 148

5.3.2.2 Sample of Size 200 observations 154

5.3.2.3 Sample of Size 500 observations 159

5.4 Kidney Catheter Data 164

5.4.1 Model 2 versus the Pure Classical Parametric Survival Models 164

5.4.2 Model 2 versus Three Components Parametric Survival Mixture Models of the Same Distributions 169

5.5 Summary 174

CHAPTER SIX CONCLUSION 176

6.1 Summary 176

6.2 Problems and Limitations 177

6.3 Future Research 178

REFERENCES 180

List of Tables

Table 2.1 18
Some Important Parametric Distributions in Survival Data Analysis 18
Table 3.1 48
Some Previous Simulation and Real Data Applications 48
Table 4.1 75
The Estimated Parameters the Simulated Data of Postulated Model 1 with 10%, 20% and 40% Censoring Observations 75
Table 4.2 77
The Repeated Simulation of Set of 100 Observations 77
Table 4.3 80
The Estimated Parameters the Simulated Data of size 200 with 10% Censoring Observations 80
Table 4.4 82
The Repeated Simulation of Set of 200 Observations 82
Table 4.5 85
The Estimated Parameters the Simulated Data of size 500 with 10% Censoring Observations 85
Table 4.6 87
The Repeated Simulation of Set of 500 Observations 87
Table 4.7 91
The Estimated Parameters the Simulated Data of Postulated Model 1 with 10% Censoring Observations 91
Table 4.8 93
The Repeated Simulation of Set of 100 Observations 93
Table 4.9 97
The Estimated Parameters the Simulated Data of size 200 with 10% Censoring Observations 97
Table 4.10 99
The Repeated Simulation of Set of 200 Observations 99
Table 4.11 102
The Estimated Parameters the Simulated Data of size 500 with 10% Censoring Observations 102

Table 4.12 104
The Repeated Simulation of Set of 500 Observations 104
Table 4.13 107
The Estimated Parameters of Simulated Data of the Pure Classical Survival Model of Gamma Distribution.. 107
Table 4.14 108
The Estimated Parameters of Model 1 of Marrow Transplant Data 108
Table 4.15 110
The Estimated Parameters of Model 1 for Bone Marrow Transplant Data 110
Table 4.16 112
The Estimated Parameters of (G1_G2) of the Bone Marrow Transplant Data 112
Table 4.17 112
The Estimated Parameters of of (G1_G2_G3_G4) of the Bone Marrow Transplant Data .. 112
Table 4.18 113
The LL and AIC Values of the Parametric Survival Mixture Models of the Gamma Distribution 113
Table 4.19 114
The Estimated Parameters of Model 1 of Kidney Catheter Data 114
Table 4.20 116
The Estimated Parameters of Model 1 for Kidney Catheter Data 116
Table 4.21 118
The Estimated Parameters of (G1_G2) of the Kidney Catheter Data 118
Table 4.22 118
The Estimated Parameters of of (G1_G2_G3_G4) of the Kidney Catheter Data 118
Table 4.23 119
The LL and AIC Values of the Parametric Survival Mixture Models of the Gamma Distribution 119
Table 4.24 120
The Estimated Parameters of Vaginal Cancer Data of the Pure Classical Gamma Distribution 120
Table 5.1 135
The Estimated Parameters of the Simulated Data of Postulated Model 2 with 10% Censoring Observations 135
Table 5.2 137
The Repeated Simulation of Set of 100 Observations 137

Table 5.3 140
The Estimated Parameters the Simulated Data of size 200 with 10% Censoring Observations 140
Table 5.4 142
The Repeated Simulation of Set of 200 Observations 142
Table 5.5 145
The Estimated Parameters the Simulated Data of size 500 with 10% Censoring Observations 145
Table 5.6 147
The Repeated Simulation of Set of 500 Observations 147
Table 5.7 151
The Estimated Parameters the Simulated Data of size 100 with 10% Censored Observations 151
Table 5.8 153
The Repeated Simulation of Set of 100 Observations 153
Table 5.9 156
The Estimated Parameters the Simulated Data of size 200 with 10% Censoring Observations 156
Table 5.10 158
The Repeated Simulation of Set of 200 Observations 158
Table 5.11 161
The Estimated Parameters the Simulated Data of size 500 with 10% Censoring Observations 161
Table 5.12 163
The Repeated Simulation of Set of 500 Observations 163
Table 5.13 166
The Estimated Parameters of Model 2 Using Kidney Catheter Data 166
Table 5.14 167
The LL and AIC Values for the Kidney Catheter Data 167
Table 5.15 171
Parameters, LL and AIC of Model 2 and Parametric Survival Mixture Models Corresponding to Each Component of Model 2 171

List of Figures

Figure 2.1 Hazard Function of Exponential Distribution λ = 0.25 19
Figure 2.2.Survival Function of Exponential Distribution λ = 0.25 20
Figure 2.3.Probability Density Function of Gamma Distribution $\alpha = 3$ and $\beta = 1$............ 21
Figure 2.4.Survival Function of the Gamma Distribution α =3 and β =1 22
Figure 2.5.Hazard Function of the Weibull Distribution $\alpha = 4$ 23
Figure 2.6.Survival Function of the Weibull Distribution $\alpha = 4$ 24
Figure 2.7 Density Function of the Lognormal Distribution $\mu = 0$ and $\sigma = 1$ 25
Figure 2.8.Survival Function of the Lognormal Distribution $\mu = 5$ and $\sigma = 3$ 26
Figure 2.9.Probability Density Function of the Gompertz Distribution $\alpha = 1.2$ and $\beta = 1$. 27
Figure 2.10.Survival Function of the Gompertz Distribution $\alpha = 3$ and $\alpha = 1$ 28
Figure 2.11.Kaplan-Meier Survival Function 30
Figure 2.12.Nelson-Aalen Cumulative Hazard Function.. 31
Figure 3.1.Methodology Frameworks of Model 1 44
Figure 3.2.Methodology Frameworks of Model 2 45
Figure 4.1 Probability Density Function of the Simulated Data of Model 1 with 100 Observations and 10% Censoring. 73
Figure 4.2 Probability Density Function of the Simulated Data of Model 1 with 100 Observations and 20% Censoring. 73
Figure 4.3 Probability Density Function of the Simulated Data of Model 1 with 100 Observations and 40% Censoring. 74
Figure 4.4 The Hazard Functions of the Simulated Data of Model 1 Corresponding to 10%, 20% and 40% Censored Observation.. 76
Figure 4.5 Probability Density Function of the Simulated Data of Size 200 Observations and 10% Censoring 78
Figure 4.6 Probability Density Function of the Simulated Data of Size 200 and 20% Censored Observations. 79
Figure 4.7 Probability Density Function of the Simulated Data of Size 200 Observations and 40% Censoring... 79
Figure 4.8 the hazard functions of the simulated data of size 200 corresponding to 10%, 20% and 40% censored observation.. 81
Figure 4.9 Probability Density Function of the Simulated Data of Size 500 Observations and 10% Censoring... 83

Figure 4.10 Probability Density Function of the Simulated Data of Size 500 Observations and 20% Censoring. 84
Figure 4.11 Probability Density Function of the Simulated Data of Size 500 Observations and 40% Censoring. 84
Figure 4.12 the hazard functions of the simulated data of size 500 corresponding to 10%, 20% and 40% censored observation. 86
Figure 4.13 Probability Density Function of the Simulated Data of Model 1 with 100 Observations and 10% Censoring. 89
Figure 4.14 Probability Density Function of the Simulated Data of Model 1 with 100 Observations and 20% Censoring. 90
Figure 4.15 Probability Density Function of the Simulated Data of Model 1 with 100 Observations and 40% Censoring 90
Figure 4.16 the Hazard Functions of the Simulated Data of Model 1 Corresponding to 10%, 20% and 40% Censored Observations. 92
Figure 4.17 Probability Density Function of the Simulated Data of Size 200 Observations and 10% Censoring. 95
Figure 4.18 Probability Density Function of the Simulated Data of Size 200 Observations and 20% Censoring. 95
Figure 4.19 Probability Density Function of the Simulated Data of Size 200 Observations and 40% Censoring. 96
Figure 4.20 The Hazard Functions of the Simulated Data of Size 200 Corresponding to 10%, 20% and 40% Censored Observation. 98
Figure 4.21 Probability Density Function of the Simulated Data of Size 500 Observations and10% Censoring. 100
Figure 4.22 Probability Density Function of the Simulated Data of Size 500 Observations and 20% Censoring. 101
Figure 4.23 Probability Density Function of the Simulated Data of Size 500 Observations and 40% Censoring. 101
Figure 4.24 The Hazard Functions of the Simulated Data of Size 500 Corresponding to 10%, 20% and 40% Censored Observations. 103
Figure 4.25The Density Functions of Pure Classical Parametric Survival Model of the Gamma Distribution 106
Figure 4.26 probability density function of Model 1 using Bone Marrow Transplant Data 108

Figure 4.27.Model 1 vs the Pure Classical Survival Model of the Gamma Distribution (G0) Using Bone Marrow Transplant Data ... 109
Figure 4.28 K-M, the Survival function of Model 1 and the Pure Survival Model ... 111
Figure 4.29 the probability density function of Model 1 using Kidney Catheter Data ... 114
Figure 4.30.Model 1 vs the Pure Classical Survival Model of the Gamma Distribution (G0) Using Kidney Catheter Data ... 115
Figure 4.31 K-M, the Survival function of Model 1 and the Pure Survival Model ... 117
Figure 4.32 The Pure Classical Parametric Survival Model of the Gamma Distribution for Vaginal Cancer Data ... 120
Figure 5.1 Probability Density Function of the Simulated Data of Model 2 with 100 Observations and 10% Censoring. ... 133
Figure 5.2 Probability Density Function of the Simulated Data of Model 2 with 100 Observations and 20% Censored Observations. ... 133
Figure 5.3 Probability Function of the Simulated Data of Model 2 of size 100 Observations and 40% Censored Observations. ... 134
Figure 5.4 the Hazard Functions of the Simulated Data of Model 1 Corresponding to 10%, 20% and 40% Censored Observations. ... 136
Figure 5.5 Probability Density Function of the Simulated Data of Size 200 Observations and 10% Censoring. ... 138
Figure 5.6 Probability Density Function of the Simulated Data of Size 200 and 20% Censored Observations. ... 139
Figure 5.7 Probability Density Function of the Simulated Data of Size 200 Observations and 40% Censoring. ... 139
Figure 5.8 the hazard functions of the simulated data of size 200 corresponding to 10%, 20% and 40% censored observation. ... 141
Figure 5.9 Probability Density Function of the Simulated Data of Size 500 Observations and 10% Censoring. ... 143
Figure 5.10 Probability Density Function of the Simulated Data of Size 500 Observations and 20% Censoring. ... 144
Figure 5.11 Probability Density Function of the Simulated Data of Size 500 Observations and 40% Censoring. ... 144
Figure 5.12 the hazard functions of the simulated data of size 500 corresponding to 10%, 20% and 40% censored observation. ... 146
Figure 5.13 Probability Density Function of the Simulated Data of Model 2 with 100 and 10% Censored Observations. ... 149

Figure 5.14 Probability Density Function of the Simulated Data of Model 2 with 100 Observations and 20% Censoring. 149
Figure 5.15 Probability Density Function of the Simulated Data of Model 2 with 100 and 40% Censored Observations. 150
Figure 5.16 the hazard functions of the simulated data of Model 2 corresponding to 10%, 20% and 40% censored observation. 152
Figure 5.17 Probability Density Function of the Simulated Data of Size 200 and 10% Censored Observations. 154
Figure 5.18 Probability Density Function of the Simulated Data of Size 200 and 20% Censored Observations. 155
Figure 5.19 Probability Density Function of the Simulated Data of Size 200 and 40% Censored Observations. 155
Figure 5.20 the hazard functions of the simulated data of size 200 corresponding to 10%, 20% and 40% censored observation. 157
Figure 5.21 Probability Density Function of the Simulated Data of Size 500 and 10% Censored Observations. 159
Figure 5.22 Probability Density Function of the Simulated Data of Size 500 and 20% Censored Observations. 160
Figure 5.23 Probability Density Function of the Simulated Data of Size 500 Observations and 40% Censoring. 160
Figure 5.24 the hazard functions of the simulated data of size 500 corresponding to 10%, 20% and 40% censored observation. 162
Figure 4.25.Model 2 vs the Pure Classical Parametric Survival Models (E, G and W) for the Kidney Catheter Data 165
Figure 4.26.Model 2 vs the Pure Classical Parametric Survival Models (E0, G0 and W0) for the Kidney Catheter Data 166
Figure 5.27 K-M, the Survival function of Model 2 and the Pure Survival Models Corresponding to Each Component 168
Figure 5.28.Model 2 vs the Parametric Survival Mixture Models Corresponding to Each Component Model 2 170
Figure 5.29.Model 2 vs the Parametric Survival Mixture of Exponential and the Pure Classical Distribution of Each Component 173
Figure 5.31.Model 2 vs the Parametric Survival Mixture of Weibull and the Pure Classical Distribution of Each Component 173

List of Appendices

Appendix A Real Data for Model 1 186
Appendix B Real Data for Model 2 191
Appendix C Computer Coding for Model 1 193
Appendix D Computer Coding for Model 2 201
Appendix E Log-likelihood Derivations Model 1 209
Appendix F Log-likelihood Derivations Model 2 212
Appendix G Published Articles 219

Glossary of Terms

E1_E2_E3	PARAMETRIC SURVIVAL MIXTURE MODEL OF EXPONENTIAL_EXPONENTIAL_EXPONENTIAL DISTRIBUTIONS.
E_G_W	PARAMETRIC SURVIVAL MIXTURE MODEL OF EXPONENTIAL_GAMMA_WEIBULL DISTRIBUTIONS.
G1_G2_G3	PARAMETRIC SURVIVAL MIXTURE MODEL OF GAMMA_GAMMA_GAMMA DISTRIBUTIONS.
W1_W2_W3	PARAMETRIC SURVIVAL MIXTURE MODEL OF WEIBULL_WEIBULL_ WEIBULL DISTRIBUTIONS.

List of Abbreviations

AIC	AKAIKE INFORMATION CRITERION
EEG	EXTENDED EXPONENTIAL-GEOMETRIC
EM	EXPECTATION MAXIMIZATION
KM	KAPLAN-MEIER
LL	LOG LIKELIHOOD
MCM	MIXTURE CURE MODELS
ML	MAXIMUM LIKELIHOOD
MSE	MEAN SQUARE ERROR
MTIWD	MIXTURE OF TWO INVERSE WEIBULL DISTRIBUTION
RMSE	ROOT MEAN SQUARE ERROR

CHAPTER ONE
INTRODUCTION

1.1 Background of the Study

Survival data analysis is the analysis of time to occurrence of a particular event of interest. The data are usually related to clinical studies of human, or laboratory studies of animal, or studies to test the life time of some devices. Major applications are in the areas of human clinical studies and industrial life testing (Kalbfleisch & Prentice, 2002).

The event of interest in clinical studies could be death, remission, or some other clinical events. The event of interest could be time taken to learning a new skill, exit from unemployment, divorce of a couple or failure of a device, to mention a few. The variable of interest, the time to occurrence of particular event T, which is a positive random variable, should clearly be defined in the study at hand. The start and end with the length of the time period in-between corresponding to T, should also be clearly defined prior to the commencement of the study (Lee & Wang, 2003).

Generally, in survival analysis, some individuals or objects do not experience the event of interest for one reason or the other, either they are lost to follow up during the period of the study or they do not experience the event until the end of the study. In such situation, the information about this particular individual will not be exactly known, and such individuals are referred to as censored observations or censored times.

The contents of the thesis is for internal user only

REFERENCES

Abu Bakar, M. Z., Daud, I, & Ibrahim, N. A. (2006). Estimating a logistic Weibull mixture models with long-Term survivors. *Jurnal Tecknologi,* 45(C) , 57-66.

Abu -Zinadah, H. H. (2010). A study on mixture of exponentiated pareto and exponential distributions. *Journal of Applied Sciences Research, 6*(4), 358-376.

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, *19*(6), 716-723.

Al-Hussaini, E. K., Al-Dayian, G. R., & Adham, S. A. (2000). On finite mixture of two-component Gompertz lifetime model. *Journal of Statistical Computation and Simulation, 67*(1), 1-20.

Birnbaun, Z. W. & Saunders S. C. (1958).“A statistical model for life-length of materials”. *Journal of the American Statistical Association.* 53, 151-160.

Blackstone, E. H., Naftel, D. C., & Turner, M. E. Jr. (1986). The decomposition of time-varying hazard into phases, each incorporating a separate stream of concomitant information. *Journal of the American Statistical Association, 81*(395), 615-624.

Bohning, D., & Seidel, W. (2003). Editorial: recent developments in mixture models. *Computational Statistics & Data Analysis, 41*(3-4), 349-357.

Brown, G. W., & Flood, M. M. (1947). Tumbler mortality. *Journal of the American Statistical Association*. 42, 562-574.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological methods & research*, *33*(2), 261-304.

Cai, C., Zou, Y., Peng, Y., Zhang, J., & Cai, M. C. (2012). Package ‘smcure’.

Chang, S. C. (1998). Using parametric statistical models to estimate mortality structure: The case of Taiwan. *Journal of Actuarial Practice*, *6*(1).

Cheng, S. W., & Fu, J. C. (1982). Estimation of mixed Weibull parameters in life testing. *Reliability, IEEE Transactions on, R-31*(4), 377-381.

Cohen, A. C., Jr. (1951). Estimating parameters of logarithmic-normal distributions by maximum likelihood. *Journal of the American Statistical Association, 46*(254), 206-212.

Copas, J. B., & Heydari, F. (1997). Estimating the risk of reoffending by using exponential mixture models. *Journal of the Royal Statistical Society*. *Series A (Statistics in Society), 160*(2), 237-252.

Davis, D. J. (1952). An analysis of some failure data. *Journal of the American Statistical Association, 47*(258), 113-150.

Dempster, A. P. Laird, N. M., & Rubin, D. B.(1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion)”. *Journal of Royal Statistical Society. Series B, 39, 1-38.*

Epstein, B. & Sobel, M. (1953). Life testing. *Journal of the American Statistical Association*, 48, 486-502.

Erisoglu, U., Erisoglu, M. & Erol, H. (2011). A mixture model of two different distributions approach to the analysis of heterogeneous survival data. *International Journal of Computational and Mathematical Sciences 5(2).*

Erisoglu, U., Erisoglu, M., & Erol, H. (2012). Mixture model approach to the analysis of heterogeneous survival data. *Pakistan Journal of Statistics 28*(1), 115-130.

Erişoğlu, Ü., & Erol, H. (2010). Modelling heterogeneous survival data using mixture of extended exponential-geometric distributions. *Communications in Statistics - Simulation and Computation, 39*(10), 1939-1952.

Escobar, L. A., & Meeker, W. Q., Jr. (1992). Assessing Influence in Regression Analysis with Censored Data. *Biometrics, 48*(2), 507-528. doi: 10.2307/2532306.

Everitt, B. S., & Hand, D. J., (1981). *Finite mixture distributions.* Chapman and Hall Inc. New York

Farcomeni, A., & Nardi, A. (2010). A two-component Weibull mixture to model early and late mortality in a Bayesian framework. *Computational Statistics & amp; Data Analysis, 54*(2), 416-428.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics, 38*(4), 1041-1046.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association, 97*(458), 611-631.

Fruhwirth-Schnatter, S. (2006). *Finite mixture and markovs switching models*. New York: Springer.

Harter, H. L., & Moore, A. H. (1966). Local-maximum-likelihood estimation of the parameters of three-parameter Lognormal populations from complete and censored samples. *Journal of the American Statistical Association, 61*(315), 842-851.

Ibrahim, J. G., Chen, M. H., & Sinha, D. (2001). *Bayesian survival analysis*. New York: Springer-verlag.

Jaheen, Z. (2005). On record Statistics from a mixture of two exponential distributions. *Journal of Statistical Computation & Simulation, 75*(1), 1-11.

Jensen, J. & Petersen, N. E. (1982). *Burn-in: an engineering approach to the design and analysis of burn-in procedures,* wiley , New york.

Jewell, N. P. (1982). Mixtures of exponential distributions. *The Annals of Statistics, 10*(2), 479-484.

Jiang, S. & Kececioglu, D (1992a). Graphical representation of two mixed-Weibull distributions. *IEEE Transaction on Reliability*, vol. 41,241-247.

Jiang, S. & Kececioglu, D (1992b). Maximum likelihood estimates, from censored data, for mixed-Weibull distributions. *IEEE Transaction on Reliability*, vol. 41,248-255.

Jiang, R., & Murthy, D. N. P. (1995). modelling failure-data by mixture of 2 Weibull distributions: a graphical approach. *Reliability, IEEE Transactions on, 44*(3), 477-488.

Jiang, R., & Murthy, D. N. P. A mixture model involving three weibull distributions, Proceedings of the Second Australia-Japan Workshop on Stochastic Models in Engineering. Technology and Management (Gold Coast, Australia).

Kalbfleisch J. D. & Prentice R. L. (2002). *The statistical analysis of failure time data* (second ed.), John Wiley & Sons, Inc. Hoboken, New Jersey.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*(282), 457-481.

Kersey, J. H., Weisdorf, D., Nesbit, M. E., LeBien, T. W., Woods, W. G., McGlave, P. B., . . . Bostrom, B. (1987). Comparison of autologous and allogeneic bone marrow transplantation for treatment of high-risk refractory acute lymphoblastic leukemia. *New England Journal of Medicine, 317*(8), 461-467.

Khalid, Z. M. & Morgan, J. T.(2008). Cross-sectional and longitudinal approaches in a survival mixture model, *Matematika*, Vol. 24, 231-242.

Koti, K. M. (2001). Failure-time mixture models: yet another way to establish efficacy. *Drug Information Journal, 35*(4), 1253-1260.

Kouassi, D. A. & Singh J. (1997). A semi-parametric approach to hazard estimation with randomly censored observations. *Journal of American Statistical Association* 92, pp.1351-1355.

Kuk, A. Y. C., & Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika, 79*(3), 531-541.

Larson, M. G., & Dinse, G. E. (1985). A mixture model for the regression analysis of competing risks data. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 34*(3), 201-211.

Lawless J. F. (2003). *Statistical models and methods of lifetime data*, (2nd ed.) John Wiley and Sons, Inc. Hoboken, New Jersey.

Lee, E. T. & Wang, J. W.(2003). *Statistical methods for survival data analysis* (3rd ed.). John Wiley & son.

Leisch, F. (2004). Exploring the structure of mixture model components. In J Antoch (ed.), "Compstat 2004- proceedings in Computational Statistics", pp. 1405-1412. Physica Verlag, Heidelberg. ISBN 3-7908-1554-3.

Leng, O. Y., & Khalid, Z. M. (2010). *A comparative study of maximum likelihood and Bayesian estimation approaches in estimating frailty mixture survival model parameters.* Paper presented at the Proceedings of the 6th IMT-GT Conference on Mathematics, Statistics and its Applications (ICMSA2010), Universiti Tunku Abdul Rahman, Kuala Lumpur, Malaysia.

Li, L., & Choe, M. K. (1997). A mixture model for duration data: analysis of second births in China. *Demography, 34*(2), 189-197.

Ling, D., Huang, H.-Z., & Liu, Y. (26, 26-29 Jan. 2009). *A method for parameter estimation of mixed Weibull distribution.* Paper presented at the Reliability and Maintainability Symposium, 2009. RAMS 2009. Annual.

Marín, J. M., Rodríguez-Bernal, M. T., & Wiper, M. P. (2005). Using Weibull mixture distributions to model heterogeneous survival data. *Communications in Statistics: Simulation and Computation, 34*(3), 673-684.

McGilchrist, C. A., & Aisbett, C. W. (1991). Regression with frailty in survival analysis. *Biometrics, 47*, 461-466.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*: John Wiley & Sons, Inc.

McLachlan, G. J., & Krishnan, T. (2008). *The EM algorithm and extensions* (Second ed.). Hoboken New Jersey: John Wiley & Sons, Inc.

Moltoft, J. (1983). Behind the "bathtub" curve, a new model and its consequences, *Microeclectonics & Reliability*, 23, 489-500.

Murthy D. N. P., Xie, M. & Jiang, R. (2004). *Weibull models.* John Wiley & son.

Ng, A. S. K., McLachlan, G. J., Yau, K. K. W., & Lee, A. H. (2004). Modelling the distribution of ischaemic stroke-specific survival time using an EM-based mixture approach with random effects adjustment. *Statistics in Medicine, 23*(17), 2729-2744.

Olkin, I., & Spiegelman, C. H.(1987). A semi-parametric approach to density estimation, *Journal of the American Statistical Association, 82, 858-865.*

Othus, M. Li, Y & Tiwari, R. C. (2009). A class of semi-paramertic mixture cure survival models with dependent censoring. *Journal of American Statistical Association,* 104(487). 1241-1250.

Phillips, N., Coldman, A., & McBride, M. L. (2002). Estimating cancer prevalence using mixture models for cancer survival. *Statistics in Medicine, 21*(9), 1257-1270.

Razali, A. M., & Salih, A. A. (2009). Combining two Weibull distributions using a mixing parameter. *European Journal of Scientific Research, 31*(2), 296-305.

Rider, P. R. (1961). The method of moments applied to a mixture of two exponential distributions. *The Annals of Mathematical Statistics, 32*(1), 143-147.

Seppa, K. Hakulinen, T., Kim, J. J. & Laara, E. (2010). Cure fraction model with random effects for regional variation in cancer survival. *Statistics in Medicine*, 29. 2781-2793.

Sultan, K. S., Ismail, M. A., & Al-Moisheer, A. S. (2007). Mixture of two inverse Weibull distributions: properties and estimation. *Computational Statistics & Data Analysis, 51*(11), 5377-5387.

Sun, J. (2006). *The statistical analysis of interval-cencored failure time data*. New York: Springer Science, Business Media.

Tableman, M., & Kim, J. S. (2004). *Survival analysis using S: analysis of time- to-event data*: Chapman & Hall/CRC.

Taylor, J. M. G. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics, 51*(3), 899-907.

Team, R. C. (2005). R: A language and environment for statistical computing: ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria, 2013. url: http://www. R-project. org.

Therneau, T. (2013). A Package for Survival Analysis in S. R package version 2.37-4. Retrieved from http://CRAN.R-project.org/package=survival

Tukey, J. W. (1977) *Explanatory data analysis*. Addison Wesley publishing company Inc. Philippines.

Vernic, R., Teodorescu, S., & Pelican, E. (2009). Two Lognormal models for real data. *Annals of Statistics Ovidius Constanta, 17*(3), 263-279.

Weibull, W. (1939). A statistical theory of strength of materials. *Ingeniorsvetens Kapsakadeniens Handlingar.*

Weibull, & W. (1951). A statistical distribution function for wide applicability. *Journal of Applied Mathematics*(18), 293-297.

Wiper, M., Insua, D. R., & Ruggeri, F. (2001). Mixtures of gamma distributions with applications. *Journal of Computational and Graphical Statistics, 10*(3).

Young, D. S., Benaglia, T., Chauveau, D., Hunter, D. R., Elmore, R. T., Xuan, F., ... & Thomas, H. (2007). The mixtools package: tools for mixture models. *R Package Version 0.2. 0.*

Yu, B., & Peng, Y. (2008). “Mixture cure models for multivariate survival time data” *Computational Statistics & Data Analysis*, 52, 1524-1532.

Zelen, M. (1966). Application of exponential models to problems in cancer research. *Journal of the Royal Statistical Society. Series A (General), 129*(3), 368-398.

Zhang Y. (2008). Parametric mixture models in survival analysis with application, (Doctoral Dissertation) UMI Number: 3300387, Graduate School, Temple University.

Zhang, X., Wang, Y., & Lu, D. (2011, 26-28 July 2011). *A new algorithm for parameters estimations of multivariate mixed Weibull distributions with censoring data.* Paper presented at the 2011 International Conference on Multimedia Technology, ICMT 2011.