

CONFIDENCE INTERVALS IN
SURVIVAL ANALYSIS

TAN SHAY KEE

DISSERTATION SUBMITTED IN FULFILMENT
OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

FEBRUARY 2010

CONFIDENCE INTERVALS IN
SURVIVAL ANALYSIS

TAN SHAY KEE

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR
FEBRUARY 2010

ABSTRACT

In obtaining confidence interval for the survivor function using Greenwood's formula and in performing log-rank test for comparing the survivor functions of two groups of individuals, only the information given by the first two moments of the relevant statistics are used. Presently we show that by incorporating the information given by the third and fourth moments of the statistics, the performance of the confidence interval and statistical test can be improved.

When the survivor function can be described by the Weibull distribution, the knowledge regarding the survivor function can be obtained through the estimation of the Weibull scale and shape parameters and the Weibull quantiles. In the present work we use the method based on hypothesis testing to construct confidence intervals for these parameters. In implementing the procedure based on hypothesis testing, we have made use of the distribution of the relevant statistic evaluated at a value of the parameter vector which satisfies the conditions under the null hypothesis and yet is nearest to the estimated parameter vector. It is found that compared to the bootstrap confidence intervals, the confidence intervals based on hypothesis testing tend to perform better.

ABSTRAK

Dalam proses untuk mencari selang keyakinan bagi fungsi masa hidup dengan menggunakan formula Greenwood dan melakukan ujian log-rank bagi membandingkan fungsi masa hidup daripada dua kumpulan individu berasingan, maklumat yang digunakan adalah momen pertama dan kedua bagi statistik berkenaan. Kini, dengan menggunakan momen ketiga dan keempat bagi statistik berkenaan, ditunjukkan bahawa prestasi selang keyakinan dan ujian berstatistik boleh dipertingkatkan.

Ketika fungsi masa hidup dapat ditentukan dengan menggunakan taburan Weibull, pengetahuan tentang fungsi masa hidup dapat diperolehi melalui anggaran bagi parameter skala, parameter bentuk dan kuantil untuk taburan Weibull. Dalam kajian ini, kaedah berdasarkan ujian hipotesis digunakan untuk membina selang keyakinan bagi parameter tersebut. Dalam proses menggunakan kaedah yang berdasarkan ujian hipotesis, maklumat penting yang diperlukan ialah taburan bagi statistik yang berkenaan pada nilai untuk vektor parameter yang memenuhi syarat dalam hipotesis nul dan terdekat dengan jangkauan vektor parameter tersebut. Didapati bahawa sekiranya dibandingkan dengan selang keyakinan *bootstrap*, selang keyakinan yang berdasarkan ujian hipotesis adalah lebih memuaskan.

ACKNOWLEDGEMENTS

I am most grateful to my supervisor Professor Dr. Pooi Ah Hin who has provided substantial advice and encouragement during the term of my research.

Thanks also to my friends Miss Fam Pei Shan and Dr. Ng Kok Haur who have given me comments and suggestions in JAVA programming. My sincere thanks are also due to the members of the Institute of Mathematical Sciences, especially Miss Ng Lee Ling, for her substantial technical advice.

Most importantly, I wish to thank my family members for their unconditional love, support and encouragement.

TABLES OF CONTENTS

ABSTRAK		ii
ABSTRACT		iii
ACKNOWLEDGEMENTS		iv
TABLE OF CONTENTS		v
LIST OF FIGURES		viii
LIST OF TABLES		ix
CHAPTER 1	INTRODUCTION	1
1.1	Survival Analysis	1
1.2	Summary of the Dissertation	4
CHAPTER 2	CONFIDENCE INTERVALS FOR VALUES OF SURVIVOR FUNCTION	5
2.1	Introduction	5
2.2	Standard Error of the Survivor Function	5
2.3	Approximate Confidence Intervals Based on Higher Order Moment	7
2.4	Numerical Results	13
2.5	Concluding Remarks	14

CHAPTER 3	THE LOG RANK TEST	21
3.1	Introduction	21
3.2	The Approximate Distribution of U_L	23
3.3	Numerical Results	24
CHAPTER 4	APPROXIMATE DISTRIBUTIONS FOR THE ESTIMATORS OF THE WEIBULL QUANTILES	26
4.1	Introduction	26
4.2	The Maximum Likelihood Estimates of the Weibull Parameters	27
4.3	Approximate Distribution of $\hat{\lambda}, \hat{\gamma}$	32
4.4	Multivariate Quadratic-Normal Distribution	32
4.5	Adequacy of Approximation given by the Multivariate Quadratic-Normal Distribution	35
CHAPTER 5	CONFIDENCE INTERVALS FOR THE WEIBULL PARAMETERS	39
5.1	Confidence Intervals for the Scale and Shape Parameters	39
5.2	Confidence Intervals for the Quantile	43
5.3	Numerical Results	45
CHAPTER 6	CONCLUDING REMARKS	53
APPENDIX A	DERIVATION OF THE MOMENT $E d_{1j}^k$	54

APPENDIX B	EVALUATION OF THE MOMENT	$E w_0^{I_0} w_1^{I_1} w_2^{I_2} w_3^{I_3} w_4^{I_4}$	58
REFERENCES			70

LIST OF FIGURES

Figure 2.3.1: Probability density function of ε_i .	10
Figure 2.3.2: Probability density function of ε_i .	11
Figure 2.3.3: Probability density function of ε_i .	12
Figure 4.5.1: Histogram for $\hat{\gamma}$ computed from \mathbf{t} which is generated directly from Weibull (0.155, 0.5) ($n = 20, N = 10000$).	36
Figure 4.5.2: Histogram for $\hat{\gamma}$ computed from $(w_0, w_1, w_2, w_3, w_4)$ which is generated from the multivariate quadratic-normal distribution ($\lambda_0 = 0.155, \gamma_0 = 0.5, n = 20, N = 10000$).	36
Figure 4.5.3: Histogram for $\hat{\gamma}$ computed from \mathbf{t} which is generated directly from Weibull (7.750E-03, 1.5) ($n = 20, N = 10000$).	37
Figure 4.5.4: Histogram for $\hat{\gamma}$ computed from $(w_0, w_1, w_2, w_3, w_4)$ which is generated from the multivariate quadratic-normal distribution ($\lambda_0 = 7.750E-03, \gamma_0 = 1.5, n = 20, N = 10000$).	37
Figure 4.5.5: Histogram for $\hat{\gamma}$ computed from \mathbf{t} which is generated directly from Weibull (3.875E-04, 2.5) ($n = 20, N = 10000$).	38
Figure 4.5.6: Histogram for $\hat{\gamma}$ computed from $(w_0, w_1, w_2, w_3, w_4)$ which is generated from the multivariate quadratic-normal distribution ($\lambda_0 = 3.875E-04, \gamma_0 = 2.5, n = 20, N = 10000$).	38

LIST OF TABLES

<p>Table 2.4.1 : Coverage probabilities and expected lengths of the 95% confidence intervals for the survivor function $S(t)$ when $T \sim \text{Weibull}(\lambda, \gamma)$ [Median of T is 20, $t = I^*(\text{Median of } T)/10$, $n = 30$, $N = 1000$].</p>	15
<p>Table 3.3.1 : Rejection probabilities of the tests based on quadratic-normal and normal distributions ($\alpha = 0.05$, $r = 10$, $d_j = 1$, $1 \leq j \leq r$, $N = 10000$).</p>	25
<p>Table 4.2.1 : The estimate $\hat{\gamma}$ calculated using Equations (4.2.6) and (4.2.11) for the j-th generated value of \mathbf{t}, $r = n$ (Median of the Weibull distribution is 20. The values in parentheses are the values based on Equation (4.2.11)).</p>	30
<p>Table 4.2.2 : The Estimate $\hat{\gamma}$ Calculated using Equations (4.2.6) and (4.2.11) for the j-th generated value of \mathbf{t}, $r = (n - 1)$ (Median of the Weibull distribution is 20. The values in parentheses are the values based on Equation (4.2.11)).</p>	31
<p>Table 5.3.1 : Coverage probabilities and expected lengths of confidence intervals for γ (Median = 20, $n = 20$, $r = n - 1$, $N = 1000$).</p>	46
<p>Table 5.3.2 : Coverage probabilities and expected lengths of confidence intervals for λ (Median = 20, $n = 20$, $r = n - 1$, $N = 1000$).</p>	48
<p>Table 5.3.3 : Coverage probabilities and expected lengths of confidence intervals for Q_p (Median = 20, $n = 20$, $r = n - 1$, $N = 1000$).</p>	49