12-2009

A Two-Sample Adaptive Procedure Based on the Log-Rank and Peto and Peto's Wilcoxon Tests

Annie A. Tordilla
Western Michigan University

A TWO-SAMPLE ADAPTIVE PROCEDURE BASED ON THE LOG-RANK AND PETO AND PETO'S WILCOXON TESTS

by

Annie A. Tordilla

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Statistics
Advisor: Joshua Naranjo, Ph.D.

Western Michigan University
Kalamazoo, Michigan
December 2009

A TWO-SAMPLE ADAPTIVE PROCEDURE BASED ON THE LOG-RANK AND PETO AND PETO'S WILCOXON TESTS

Annie A. Tordilla

Western Michigan University, 2009

It has been shown that under a location-scale model $y = \mu + \beta z + \sigma\epsilon$ where y is right censored, the Log-Rank test is asymptotically efficient for the Extreme minimum value error distribution while Peto and Peto's Wilcoxon test is asymptotically efficient for the Logistic error distribution. We propose a two-sample adaptive test, which first selects between Extreme minimum value and Logistic error distribution as to which is a better fit to the data, then performs the asymptotically efficient test (Log-Rank or Peto and Peto's Wilcoxon test) for the selected distribution. The performance of the adaptive test is compared with the Log-Rank and Peto and Peto's Wilcoxon tests through simulation.

ACKNOWLEDGMENTS

Completing this dissertation while holding a full-time job would have been nearly impossible had it not been for the following people:

Dr. Joshua Naranjo, whose practical insights and ideas have always brought me back on track whenever my brain digressed. Thank you for accommodating my odd consultation hours. Having you as my dissertation advisor definitely made it much less difficult for me.

My reading committee members: Dr. Steve Denham, who encouraged and supported me in finishing my doctorate degree since his first week as my supervisor; Dr. Jung Chao Wang, whose R expertise I called on when my debugging skills fell short, and Dr. Joe McKean whose thought-provoking questions and suggestions since the conception of this research motivated me to continually make improvements. Thank you all for finding the time in your busy schedules to review my paper.

MPI Research Inc., who financed my dissertation credits and has nourished my intellectual maturity as a statistician.

Dr. Henry Escuadro, who helped me refresh my Calculus skills as I was working on my proofs. To the rest of my friends, thank you for the poker games, bowling and movie nights that helped me stay sane!

My parents, Pablo, Jr. and Corazon Tordilla, whose ceaseless love and support have brought me to where I am right now. To the rest of my large and extended family who have been praying for this to come for too long now, you can finally move on to the next petition in our list.

My loving husband, Jao, who is never contented with cheering from the sidelines, that he learned LaTeX and ended up encoding most of this paper. WE DID IT!!! Thank you for believing in me especially when I was starting to lose faith in myself.

Most importantly, GOD, who gave me the strength to push harder when things seemed impossible.

To everyone else, who lent a hand during this project or said a prayer or two, I offer my deepest gratitude.

Annie A. Tordilla

TABLE OF CONTENTS

Table of Contents—Continued

CHAPTER

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Background and Motivation

Survival Analysis is a statistical tool for analyzing time to "event" data. An "event" may be the death of a laboratory mouse in a carcinogenicity study, the remission of a cancer patient or the failure of a machine.

One of the fundamental interests in survival analysis is to determine if the risk of the "event" happening in one group is the same as that in the other group. The literature offers numerous inferential procedures for the comparison of survival data from two groups with right censoring. Two of the most popular tests are the nonparametric Log-Rank and Wilcoxon tests. Suppose $t_1 < t_2 < \ldots < t_D$ be the distinct failure times in the pooled samples. The two tests are based on the statistic

$$Z = \sum_{i=1}^{D} Y_{i1} W(t_i) \left[\frac{d_{i1}}{Y_{i1}} - \frac{d_i}{Y_i} \right]$$

where d_{i1} = number of failures in sample 1 at time t_i

Y_{i1} = number of individuals at risk in sample 1 at time t_i

d_i = number of failures in the combined samples at time t_i

Y_i = number of individuals at risk in the combined samples at time t_i.

The Log-Rank test (Peto and Peto, 1972; Cox, 1972; Mantel, 1966), which is a generalization of the Savage (1956) test for right-censored observations, gives equal weights to hazard differences, $W(t) = 1$ for all t. Gehan's Wilcoxon test (1965) on the other hand, which is a generalization of the Mann-Whitney Wilcoxon test, assigns more weight to early hazard differences than to late hazard differences, $W(t_i) = Y_i$. Peto and Peto (1972) suggested an alternative generalization of the Wilcoxon test which uses the estimated survival function, $\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{Y_i + 1}\right)$, as the weight to hazard differences. The advantage of Peto and Peto's Wilcoxon test over Gehan's Wilcoxon test was shown by Prentice and Marek (1979) through a case study wherein the latter statistic gave misleading results due to its dependence on censoring rates.

True to any statistical procedure, none of these tests is optimal for all data distributions. Prentice (1978) showed that under a location-scale model, $y = \mu + \beta z + \sigma\epsilon$, Log-Rank test is asymptotically fully efficient for the Extreme minimum value error distribution $f(\epsilon) = exp(\epsilon - e^{\epsilon})$, while Peto and Peto's Wilcoxon test is asymptotically fully efficient for the Logistic error distribution $f(\epsilon) = \frac{e^{\epsilon}}{(1 + e^{\epsilon})^2}$. Since the efficiency of rank tests is invariant under monotone increase data transformations, Prentice's results also mean that Log-Rank test is fully efficient when the failure time follows a Weibull distribution and Peto and Peto's Wilcoxon test is fully efficient when the failure time follows a Log-logistic distribution.

In 1975, Hogg proposed a two-sample adaptive distribution-free test which uses the data first to select the most appropriate model from a class of models and then makes an inference based on the chosen model. This paper uses Hogg's approach by proposing a preliminary test to determine which test is more appropriate for the data. Prentice (1975) proposed a discrimination procedure that embeds both Extreme minimum value and Logistic distributions in a larger parametric family of distributions. A test statistic based on parameter estimates of this distribution will serve as a pretest for Log-Rank versus Peto-Peto's Wilcoxon test. Prentice claimed asymptotic normality of the test statistic but did not provide formal proof of this result. In this paper, we will prove the asymptotic normality of Prentice's test statistic. We will also conduct a simulation study to validate Type I error rate and compare relative efficiency of the adaptive procedure against either Log-Rank or Wilcoxon test.

1.2 Accelerated Failure Time

Let T denote the failure time and Z a vector of fixed-time explanatory covariates. The relationship between the distribution of T and Z is described by the accelerated failure time model as follows:

$$S(t|Z) = S_0[e^{(\beta' Z)}t], \text{ for all t} \tag{1.1}$$

or equivalently,

$$h(t|Z) = e^{(\beta' Z)}h_0[e^{(\beta' Z)}t], \text{ for all t} \tag{1.2}$$

where β is a vector of regression coefficients

S_0 is the baseline survival function

h_0 is the baseline hazard rate.

Note the difference between (1.2) and the hazard function under the popular Cox proportional hazards model which is

$$h(t|Z) = h_0(t)e^{(\beta' Z)}.$$

The accelerated failure time model can also be represented as a linear relationship between the logarithm of failure time and the vector of covariates as shown below:

$$y = \ln T = \mu + \beta' Z + \sigma\epsilon \tag{1.3}$$

where ϵ is the error distribution.

For the two-sample problem, Z is a binary covariate which serves as an indicator of the two samples

$$Z = \begin{cases} 0, & \text{if placebo} \\ 1, & \text{if treatment} \end{cases}$$

The linear log-time models for the two samples according to (1.3) can then be expressed as

Placebo group:

$$\begin{aligned} \ln T &= \mu + \beta \times (0) + \sigma\epsilon \\ &= \mu + \sigma\epsilon \end{aligned} \tag{1.4}$$

Treatment group:

$$\begin{aligned} \ln T &= \mu + \beta \times (1) + \sigma\epsilon \\ &= \mu + \beta + \sigma\epsilon. \end{aligned} \tag{1.5}$$

Based on these log-linear models, treatment will increase or decrease the logarithm of the failure time by β or equivalently, the failure time by e^{β}. If $\beta = 0.19$, then $e^{0.19} = 1.20$. One can tell a patient that he is expected to live 20% times longer if he took the treatment. The factor $e^{(\beta' Z)}$, which tells us how much a covariate affects the expected failure time, is called an "acceleration factor".

Since the effect of the covariate is modelled directly on the failure time, the results can be easily explained to a patient. This ease of interpretability is one of the appeals of the accelerated failure time model over the Cox proportional hazards model which models the covariate effect on hazard ratios (Reid, 1994).

1.3 Statement of the Problem

The accelerated failure time model encompasses a wide range of survival time distributions, depending on the error distribution. In this paper, we will focus on two of the most popular survival distributions as discussed by Klein and Moeschberger (2003).

1. If ϵ is the Extreme minimum value distribution, then T follows the Weibull distribution.

The survival function of $T \sim Weibull\ (shape = \alpha, scale = \lambda)$ is

$$S(t) = exp(-\lambda t^{\alpha}) \tag{1.6}$$

which implies that the survival function of $y = \ln T$ is

$$S(y) = exp(-\lambda e^{\alpha y}). \tag{1.7}$$

Now, if the parameters are redefined as

$$\lambda = e^{(-\mu/\sigma)} \text{ and } \sigma = 1/\alpha, \tag{1.8}$$

the survival function of y can be rewritten as

$$\begin{aligned} S(y) &= exp(-e^{(-\mu/\sigma)}e^{(y/\sigma)}) \\ &= exp(-e^{(y-\mu)/\sigma}). \end{aligned} \tag{1.9}$$

Hence from (1.9), $y = \ln T$ can be expressed as

$$y = \ln T = \mu + \sigma\epsilon \tag{1.10}$$

where ϵ is the Extreme minimum value distribution with the survival function,

$$S(\epsilon) = exp(-e^{\epsilon}). \tag{1.11}$$

2. If ϵ is the Logistic distribution, then T follows the Log-logistic distribution.

The survival function of $T \sim$ *Log-logistic* $(shape = \alpha, scale = \lambda)$ is

$$S(t) = \frac{1}{1 + \lambda t^{\alpha}}. \tag{1.12}$$

It follows then that the survival function of $y = \ln T$ is given by

$$S(y) = \frac{1}{1 + \lambda e^{\alpha y}}. \tag{1.13}$$

In redefining the parameters as

$$\lambda = e^{(-\mu/\sigma)} \text{ and } \sigma = 1/\alpha, \tag{1.14}$$

the survival function of y becomes

$$\begin{aligned} S(y) &= \frac{1}{1 + e^{(-\mu/\sigma)} e^{(y/\sigma)}} \\ &= \frac{1}{1 + e^{(y - \mu)/\sigma}}. \end{aligned} \tag{1.15}$$

Therefore, from (1.15), $y = \ln T$ can be described as

$$y = \ln T = \mu + \sigma\epsilon \tag{1.16}$$

where ϵ is the standard Logistic distribution with the survival function,

$$S(\epsilon) = \frac{1}{1 + e^{\epsilon}}. \tag{1.17}$$

The Weibull distribution exhibits a monotonic hazard function, i.e. the hazard rate is monotone increasing ($\sigma < 1$), decreasing ($\sigma > 1$) or constant ($\sigma = 1$). In contrast, the hazard rate of the Log-logistic distribution may be non-monotonic. When $\sigma < 1$, it increases at early times and then decreases at later times. It is monotone decreasing when $\sigma \geq 1$. This paper will tackle the problem of testing for the appropriateness of the Weibull and Log-logistic distributions on the failure time data. Once we have obtained a formal test that will tell us which of these two distributions is a better fit to the data, then we will know the answer to the main problem of this research: Is Log-Rank or Peto-Peto's Wilcoxon test more appropriate for the data?

1.4 Research Objectives

This research aims to fulfill the following goals:

1. Formally derive and prove the asymptotic properties of the test statistic originally proposed by Prentice (1975),
2. Prepare a procedure for using the discrimination test in the two-sample survival data framework,
3. Investigate optimal choices of a critical value for the discrimination test statistic, and

4. Compare the finite sample performance of the proposed adaptive test with the non-adaptive tests (Log-Rank and Peto-Peto's Wilcoxon tests) through simulation.

1.5 Outline

This paper consists of five chapters including background and motivation, and statement of the problem in Chapter 1. In Chapter 2, we will discuss the generalized family of distributions used by Prentice (1975) to discriminate between Logistic and Extreme minimum value distributions. In Chapter 3, we will provide a formal proof of asymptotic normality of the discrimination test statistic. We will also prepare a scheme for extending it to discriminate between error distributions in the two-sample accelerated failure time model. In Chapter 4, we will examine the finite sample properties of the adaptive tests. Type I error and power will be compared with the Log-Rank and Peto-Peto's Wilcoxon tests. Conclusion of this research effort and proposed further work will be covered in Chapter 5.

Chapter 2

Weibull Versus Log-Logistic Distribution

Discriminating between Weibull and Log-logistic distribution under the accelerated failure time model, is equivalent to discriminating between their error distributions: Extreme minimum value and Logistic distributions. In this chapter, we will present Prentice's discrimination procedure that embeds both Extreme minimum value and Logistic distributions in a larger parametric family of distributions.

2.1 Parametric Family of Distributions

Consider the location-scale model, $y = \mu + \sigma\epsilon$, such that

$$f(\epsilon) = \frac{1}{B(m_1, m_2)} \left(\frac{m_1}{m_2}\right)^{m_1} e^{\epsilon\, m_1} \left(1 + \frac{m_1 e^{\epsilon}}{m_2}\right)^{-(m_1 + m_2)} \tag{2.1}$$

where $m_1 > 0$, $m_2 > 0$ and B is the beta function.

Hence, the density of y is

$$f(y) = \frac{1}{\sigma B(m_1, m_2)} \left(\frac{m_1}{m_2}\right)^{m_1} e^{\left(\frac{y-\mu}{\sigma}\right) m_1} \left(1 + \frac{m_1 e^{\left(\frac{y-\mu}{\sigma}\right)}}{m_2}\right)^{-(m_1+m_2)}, \quad (2.2)$$

a parametric family which reduces to common statistical models for specific values of (m_1, m_2). Some of these reduced models are exhibited in Table 2.1.

Table 2.1: Distributions of y for Some Values of m_1 and m_2

(m_1,m_2)	Distribution of y
(1,1)	Logistic
$(1,\infty)$	Extreme minimum value
$(\infty,1)$	Extreme maximum value
(∞,∞)	Degenerate normal

Notice from Table 2.1 that the Logistic and Extreme minimum value distributions both have $m_1 = 1$ but have different m_2 values: $m_2 = 1$ for Logistic and $m_2 = \infty$ for Extreme minimum value. Therefore, a discrimination test between these two models is based on the maximum likelihood estimate of m_2 being closer to 1 or ∞, given that $m_1 = 1$.

2.2 Maximum Likelihood Estimator of m_2

Let $y_1, y_2, \ldots, y_n$ be a sample of n *iid* variables with density (2.2). Hence, the likelihood function at (μ, σ, m_1, m_2) is

$$
\begin{aligned}
L(y;\mu,\sigma,m_1,m_2) &= \prod_{i=1}^{n} f(y_i) \\
&= \frac{1}{[\sigma B(m_1,m_2)]^n}\left(\frac{m_1}{m_2}\right)^{nm_1} e^{m_1 \sum_{i=1}^{n}\left(\frac{y_i-\mu}{\sigma}\right)} \\
&\quad \times \prod_{i=1}^{n}\left[1+\frac{m_1 e^{\left(\frac{y_i-\mu}{\sigma}\right)}}{m_2}\right]^{-(m_1+m_2)} .
\end{aligned} \tag{2.3}
$$

Given $m_1 = 1$, (2.3) reduces to

$$
\begin{aligned}
L(y;\mu,\sigma,m_2) &= \frac{1}{[\sigma B(1,m_2)]^n}\left(\frac{1}{m_2}\right)^{n} e^{\sum_{i=1}^{n}\left(\frac{y_i-\mu}{\sigma}\right)} \\
&\quad \times \prod_{i=1}^{n}\left[1+\frac{e^{\left(\frac{y_i-\mu}{\sigma}\right)}}{m_2}\right]^{-(1+m_2)} .
\end{aligned} \tag{2.4}
$$

The maximum likelihood estimator (MLE) of $\theta = (\mu, \sigma, m_2)$ is

$$
\hat{\theta} = \arg\max_{\theta} \; L(y;\mu,\sigma,m_2) \tag{2.5}
$$

or more conveniently,

$$\hat{\theta} = \arg\max_{\theta} \; \ln L(y; \mu, \sigma, m_2) \tag{2.6}$$

since the maxima is not affected by monotone transformations.

To obtain these maximum likelihood estimators, let us take first the log-likelihood function

$$\begin{aligned} \ell(y; \mu, \sigma, m_2) &= -n \ln m_2 + \sum_{i=1}^{n} \left(\frac{y_i - \mu}{\sigma} \right) \\ &\quad -(1 + m_2) \sum_{i=1}^{n} \ln \left[1 + \frac{1}{m_2} e^{\left(\frac{y_i - \mu}{\sigma} \right)} \right] \\ &\quad -n \ln \sigma - n \ln B(1, m_2). \end{aligned} \tag{2.7}$$

From (2.6), the MLE $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{m}_2)$ maximizes (2.7) and hence, is the solution to the following set of equations

$$\frac{\partial \ell}{\partial \mu} = -\frac{n}{\sigma} + \frac{(1 + m_2)}{\sigma} \sum_{i=1}^{n} \frac{1}{1 + \frac{1}{m_2} e^{\epsilon_i}} \left(\frac{e^{\epsilon_i}}{m_2} \right) = 0, \tag{2.8}$$

$$\frac{\partial \ell}{\partial \sigma} = -\frac{\sum_{i=1}^{n} \epsilon_i}{\sigma} + \frac{(1 + m_2)}{\sigma} \sum_{i=1}^{n} \frac{\epsilon_i}{1 + \frac{1}{m_2} e^{\epsilon_i}} \left(\frac{e^{\epsilon_i}}{m_2} \right) - \frac{n}{\sigma} = 0, \tag{2.9}$$

$$\frac{\partial \ell}{\partial m_2} = -\frac{n}{m_2} + (1+m_2)\sum_{i=1}^{n} \frac{e^{\epsilon_i}}{m_2{}^2 + m_2 e^{\epsilon_i}} - \sum_{i=1}^{n} \ln\left(1 + \frac{1}{m_2} e^{\epsilon_i}\right)$$

$$-n[\psi(m_2) - \psi(1+m_2)] = 0 \qquad (2.10)$$

where $\epsilon_i = \left(\frac{y_i - \mu}{\sigma}\right)$.

To obtain (2.10), note that $\partial B(a,b)/\partial b = B(a,b)[\psi(b) - \psi(a+b)]$, where $\psi(k) = \frac{\partial \ln \Gamma(k)}{\partial(k)} = \int_0^\infty \left(\frac{e^{-t}}{t} - \frac{e^{-kt}}{1-e^{-t}}\right) \partial t$ is the digamma function (Abramowitz and Stegun, 1964).

The MLE $\hat{m}_2$ can be obtained simultaneously with $\hat{\mu}$ and $\hat{\sigma}$ using numerical methods. For our simulation, we will use R's *optim* function which outputs the three maximum likelihood estimates based on the inputted log-likelihood function and the estimating equations (2.8), (2.9) and (2.10).

To verify if $\hat{m}_2$ can indeed discriminate between the Extreme minimum value and Logistic distributions, we simulated 50 observations from each distribution and produced the corresponding $\hat{m}_2$ values. The simulation was replicated 5000 times for each of the following cases:

Case A. $\mu = 0$, $\sigma < 1$ (probability density function (pdf) of both distributions is narrow and tall)

Case B. $\mu = 0$, $\sigma > 1$ (probability density function (pdf) of both distributions is broad and shallow).

One can observe from Figures 2.1 and 2.2 that the $\hat{m}_2$ values from the logistic distribution are clustered around 0. On the other hand, the $\hat{m}_2$ values from the extreme minimum value distribution are very large. There are also some large $\hat{m}_2$ values from the logistic distribution, but based on the percentiles in Table 2.2, only 20% of the values are greater than 2. In contrast, only 20% of the $\hat{m}_2$ estimates from the extreme minimum value distribution are less than 10.

This clear distinction between the two distributions with respect to $\hat{m}_2$ provides evidence that $\hat{m}_2$ can discriminate between the two distributions.

Figure 2.1: Scatterplots of $\hat{m}_2$ for Both Extreme Minimum Value ($m_2 = \infty$) and Logistic ($m_2 = 1$) Using $\mu = 0$ and $\sigma = 0.5$

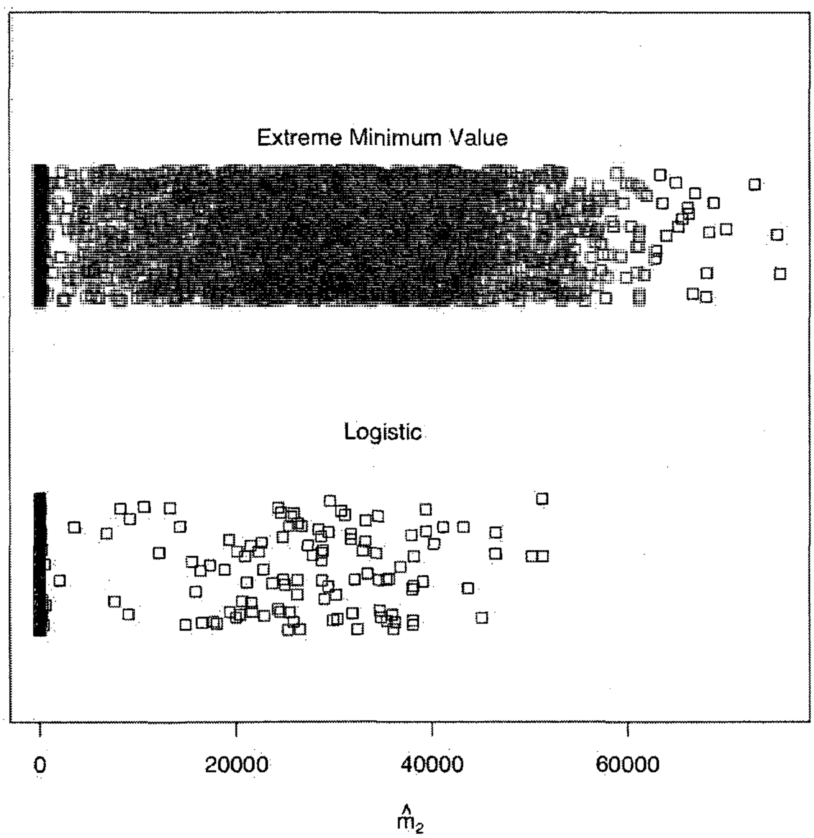


Figure 2.2: Scatterplots of $\hat{m}_2$ for Both Extreme Minimum Value ($m_2 = \infty$) and Logistic ($m_2 = 1$) Using $\mu = 0$ and $\sigma = 2$

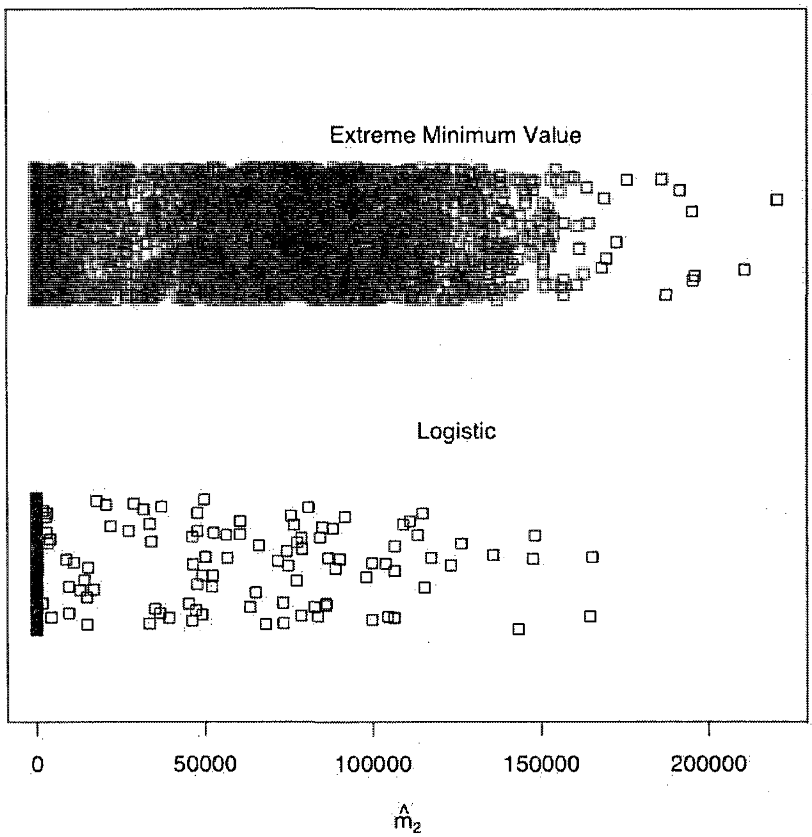


Table 2.2: Percentiles of $\hat{m}_2$ at $\mu = 0$

	$\sigma = 0.5$		$\sigma = 2$	
Percentile	Extreme minimum value ($m_2 = \infty$)	Logistic ($m_2 = 1$)	Extreme minimum value ($m_2 = \infty$)	Logistic ($m_2 = 1$)
5^{th}	2.2716	0.4310	2.1767	0.4418
10^{th}	3.1288	0.5283	3.2156	0.5402
20^{th}	5.8244	0.6590	6.0527	0.6724
30^{th}	13.7306	0.7733	13.5070	0.7921
40^{th}	3768.6133	0.8871	3523.0438	0.9048
50^{th}	18159.3947	1.0165	40544.4337	1.0444
60^{th}	25137.5983	1.1816	61759.8040	1.2092
70^{th}	30609.9432	1.4183	75824.0997	1.4498
80^{th}	36062.4104	1.7928	89861.8513	1.8903
90^{th}	42404.6878	3.0453	106864.0991	3.1853
95^{th}	47324.8637	6.1804	120990.5581	6.1366

2.3 Some Known Results (Prentice, 1975)

In the next chapter, we will present three discrimination tests that are based on $\hat{m}_2$. But before we do that, let us first discuss some of Prentice's results which we will use in that chapter.

Let ϵ of the location-scale model $y = \mu + \sigma\epsilon$ be distributed with density function (2.1).

R1. As $m_2 \to \infty$,

$$f(\epsilon) = \frac{m_1{}^{m_1}\, e^{(\epsilon m_1 - e^{\epsilon} m_1)}}{\Gamma(m_1)}$$

R2. The information matrix S corresponding to $(\alpha, \sigma, m_1, m_2)$, where $\alpha = \mu + \sigma \log\left(\frac{m_1}{m_2}\right)$, has the following elements:

$$\begin{aligned}
S_{11} &= m_1 m_2 (m_1 + m_2 + 1)^{-1} \sigma^{-2}, \\
S_{12} &= [m_1 m_2 \{\psi(m_1) - \psi(m_2)\} + (m_2 - m_1)] \\
&\quad \times (m_1 + m_2 + 1)^{-1} \sigma^{-2}, \\
S_{13} &= m_2 (m_1 + m_2)^{-1} \sigma^{-1}, \\
S_{14} &= -m_1 (m_1 + m_2)^{-1} \sigma^{-1}, \\
S_{22} &= (m_1 m_2 [\psi'(m_1) + \psi'(m_2) + \{\psi(m_1) - \psi(m_2)\}^2] \\
&\quad + 2(m_2 - m_1)\{\psi(m_1) - \psi(m_2)\} - 2) \\
&\quad \times\ (m_1 + m_2 + 1)^{-1} \sigma^{-2} + \sigma^{-2}, \\
S_{23} &= [m_2 \{\psi(m_1) - \psi(m_2)\} - 1]\, (m_1 + m_2)^{-1} \sigma^{-1}, \\
S_{24} &= [-m_1 \{\psi(m_1) - \psi(m_2)\} - 1]\, (m_1 + m_2)^{-1} \sigma^{-1}, \\
S_{33} &= \psi'(m_1) - \psi'(m_1 + m_2), \\
S_{34} &= -\psi'(m_1 + m_2), \\
S_{44} &= \psi'(m_2) - \psi'(m_1 + m_2)
\end{aligned}$$

where $\psi'(k) = \dfrac{\partial^2 \log \Gamma(k)}{\partial k^2}$.

The information matrix for θ is then CSC', where

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \log\left(\frac{m_2}{m_1}\right) & 1 & 0 & 0 \\ -\frac{\sigma}{m_1} & 0 & 1 & 0 \\ \frac{\sigma}{m_2} & 0 & 0 & 1 \end{bmatrix}.$$

R3. To establish the discrimination test between Extreme minimum value and Logistic distribution, the density function (2.2) was first reparameterized by defining $r_1 = \frac{1}{m_1}$ and $r_2 = \frac{1}{m_2}$.

a. $\frac{\partial \ell}{\partial r_1}$ and $\frac{\partial \ell}{\partial r_2}$ are finite and not identically zero at the boundary $r_1 = 0$ or $r_2 = 0$, provided r_1 and r_2 are not simultaneously zero. That is, estimation of r_2 is regular at $r_2 = 0$, provided $r_1 \neq 0$. Similarly, estimation of r_1 is regular at $r_1 = 0$ provided $r_2 \neq 0$.

For fixed $m_1 \neq \infty$ (i.e. $r_1 \neq 0$),

$$\lim_{r_2 \to 0} \frac{\partial \ell}{\partial r_2} = \lim_{m_2 \to \infty} \left(-{m_2}^2 \frac{\partial \ell}{\partial m_2}\right)$$

$$= \frac{1}{2}{m_1}^2 e^{\left[2\left(\frac{y-\mu}{\sigma}\right)\right]} - {m_1}^2 e^{\left(\frac{y-\mu}{\sigma}\right)} + \frac{1}{2}{m_1}^2 - \frac{1}{2}m_1.$$

b. At $r_1 = 1$ and $r_2 = 0$, the information that y contains about (μ, σ, r_1, r_2) is

$$I(\mu, \sigma, r_1, r_2) = \begin{bmatrix} \frac{1}{\sigma^2} & \frac{1-\gamma}{\sigma^2} & 0 & \frac{1}{\sigma} \\ \frac{1-\gamma}{\sigma^2} & \frac{\frac{\pi^2}{6} + (1-\gamma)^2}{\sigma^2} & \frac{1}{\sigma} & \frac{2-\gamma}{\sigma} \\ 0 & \frac{1}{\sigma} & \frac{\pi^2}{6} - 1 & \frac{1}{2} \\ \frac{1}{\sigma} & \frac{2-\gamma}{\sigma} & \frac{1}{2} & 2 \end{bmatrix}.$$

c. Logistic versus Extreme minimum value discrimination

Let $\ell(r_2)$ represent the log likelihood at $(r_1 = 1,\ r_2)$ maximized over $(\mu,\ \sigma)$. Then at $r_2 = 1$, the Logistic,

$$\begin{aligned} \hat{r}_2 &\sim N\left\{1,\ \frac{4(\pi^2+3)}{n(\pi^2-6)}\right\} \\ &\sim N\left(1,\ \frac{13.30}{n}\right). \end{aligned}$$

At the boundary $r_2 = 0$, the Extreme minimum value,

$$\begin{aligned} \hat{r}_2 &\sim N\left\{0,\ \frac{1}{n\left(1 - \frac{6}{\pi^2}\right)}\right\} \\ &\sim N\left(0,\ \frac{2.55}{n}\right) \end{aligned}$$

with probability $\frac{1}{2}$, that $\hat{r}_2 \leq 0$ amassed at $\hat{r}_2 = 0$.

Based on these asymptotic results, a one-sided 0.05 level test for the Extreme minimum value distribution versus a Logistic alternative rejects the hypothesis if

$$\hat{r}_2 > 1.64 \sqrt{\frac{2.55}{n}} .$$

Prentice did not provide formal proof of the asymptotic results in (R3c). It seems he omitted the nuisance parameters μ and σ, then obtained from the reduced log likelihood function $\ell(r_2)$ the discrimination test statistic $\hat{r}_2$ and its asymptotic distribution. In this paper, the complete likelihood function $L(\mu, \sigma, r_2)$, along $r_1 = 1$, is used instead to obtain the test statistic $\hat{r}_2$. It will be shown in the next chapter that the resulting asymptotic distribution of $\hat{r}_2$ is the same as (R3c) even if the nuisance parameters are not eliminated from the likelihood function.

Chapter 3

Adaptive Survival Test

The discrimination test between Logistic and Extreme minimum value is based on $\hat{r}_2$ at $r_1 = 1$, where $r_1 = \dfrac{1}{m_1}$ and $r_2 = \dfrac{1}{m_2}$. In this chapter, we shall examine the estimator of r_2 and its asymptotic distribution under the Logistic model (i.e. $r_1 = r_2 = 1$) and the Extreme minimum value model (i.e. $r_1 = 1,\ r_2 = 0$). Using the asymptotic results, we will then present a few discrimination tests between the two models. Lastly, we shall extend these tests to the two-sample accelerated failure time model setting and propose an adaptive survival test for equality of failure times.

3.1 Estimation of $r_2 = 1/m_2$

We provide in this section the theoretical derivation of the MLE $\hat{r}_2$ and its asymptotic distribution. As opposed to Prentice's approach, we will not eliminate the nuisance parameters μ and σ from the likelihood function prior to r_2 estimation. Rather, we will use the complete likelihood function $L(\mu, \sigma, r_2)$, along $r_1 = 1$ to obtain $\hat{r}_2$.

Theorem 1 : (Hogg and Craig, 1995) *Let $\eta = h(\theta)$ define a one-to-one transformation. Then the value of η, say $\hat{\eta}$, that maximizes the likelihood function $L(\theta)$, or equivalently $L(\theta = h^{-1}(\eta))$ is $\hat{\eta} = h(\hat{\theta})$, where $\hat{\theta}$ is the maximum likelihood estimator of θ. This result is called the invariance property of a maximum likelihood estimator.*

Recall from Section 2.2 that if $m_1 = 1$, i.e. $r_1 = 1$, $\hat{\mu}$, $\hat{\sigma}$ and m_2 are the solutions to equations (2.8), (2.9) and (2.10). Thus, due to the invariance property of MLE (Theorem 1), $\hat{r}_2 = \dfrac{1}{\hat{m}_2}$. Consider the next theorem for the asymptotic distribution of $\hat{r}_2$.

Theorem 2 : (Lehmann and Casella, 1998) *Let $X_1, ..., X_n$ be iid, each with a density $f(x|\theta)$ which satisfies the regularity conditions. Then, with probability tending to 1 as $n \to \infty$, there exist solutions $\hat{\theta}_n = \hat{\theta}_n(x_1, ..., x_n)$ of the likelihood equations such that $\hat{\theta}_{jn}$ is asymptotically efficient in the sense that*

$$\sqrt{n}\,(\hat{\theta}_{jn} - \theta_j) \xrightarrow{d} N\{0, [I(\theta)]^{-1}_{jj}\}$$

Using theorem (Theorem 2), we shall obtain the information matrices $I(\mu, \sigma, r_2 = 1)$ and $I(\mu, \sigma, r_2 = 0)$ at $r_1 = 1$ to establish the asymptotic distribution of $\hat{r}_2$ under the Logistic and Extreme minimum value hypotheses, respectively.

3.1.1 Logistic Model ($r_1 = r_2 = 1$)

Theorem 3 : (Sorensen and Gianola, 2002) *Let the distribution of the random vector y be indexed by a parameter θ having more than one element. Consider* *the one-to-one transformation* $\eta = f(\theta)$ *such that the inverse function* $\theta = f^{-1}(\eta)$ *exists, and suppose that the likelihood is differentiable with respect to* η *at least twice. Then, the expected information matrix for* η *is*

$$I(\eta) = A\ I[f^{-1}(\eta)]A'$$

$$\text{where} \quad A = \frac{\partial \theta'}{\partial \eta} = \begin{bmatrix} \frac{\partial \theta_1}{\partial \eta_1} & \frac{\partial \theta_2}{\partial \eta_1} & \cdots & \frac{\partial \theta_p}{\partial \eta_1} \\ \frac{\partial \theta_1}{\partial \eta_2} & \frac{\partial \theta_2}{\partial \eta_2} & \cdots & \frac{\partial \theta_p}{\partial \eta_2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \frac{\partial \theta_1}{\partial \eta_p} & \frac{\partial \theta_2}{\partial \eta_p} & \cdots & \frac{\partial \theta_p}{\partial \eta_p} \end{bmatrix},$$

$$p = \text{number of elements in } \theta.$$

In our case, the original parameterization was in terms of $\theta' = [\mu, \sigma, m_1, m_2]$ with $I(\theta) = CSC'$ as described in (R2) of Section 2.3. The new parameterization consists of the vector

$$\eta = \begin{bmatrix} \mu \\ \sigma \\ r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} \mu \\ \sigma \\ \dfrac{1}{m_1} \\ \dfrac{1}{m_2} \end{bmatrix}$$

so that the old parameter vector in terms of the new parameters is $\theta' = [f^{-1}(\eta)]' = [\mu,\ \sigma,\ 1/r_1,\ 1/r_2]$ and the old information matrix in terms of the new parameters is $I[f^{-1}(\eta)] = C^* S^* C^{*\prime}$, where

$$C^* = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \log\left(\dfrac{r_1}{r_2}\right) & 1 & 0 & 0 \\ -\sigma r_1 & 0 & 1 & 0 \\ \sigma r_2 & 0 & 0 & 1 \end{bmatrix}$$

and S^* has the following elements

$$\begin{aligned} S^*_{11} &= \frac{1}{r_1 r_2}\left(\frac{1}{r_1} + \frac{1}{r_2} + 1\right)^{-1} \frac{1}{\sigma^2}, \\ S^*_{12} &= \left[\frac{1}{r_1 r_2}\left\{\psi\left(\frac{1}{r_1}\right) - \psi\left(\frac{1}{r_2}\right)\right\} + \left(\frac{1}{r_2} - \frac{1}{r_1}\right)\right] \\ &\quad \times \left(\frac{1}{r_1} + \frac{1}{r_2} + 1\right)^{-1} \frac{1}{\sigma^2}, \end{aligned}$$

$$
\begin{aligned}
S^{\star}_{13} &= \frac{1}{r_2}\left(\frac{1}{r_1}+\frac{1}{r_2}\right)^{-1}\frac{1}{\sigma}, \\
S^{\star}_{14} &= -\frac{1}{r_1}\left(\frac{1}{r_1}+\frac{1}{r_2}\right)^{-1}\frac{1}{\sigma}, \\
S^{\star}_{22} &= \left(\frac{1}{r_1 r_2}\left[\psi^{'}\left(\frac{1}{r_1}\right)+\psi^{'}\left(\frac{1}{r_2}\right)+\left\{\psi\left(\frac{1}{r_1}\right)-\psi\left(\frac{1}{r_2}\right)\right\}^2\right]\right. \\
&\quad \left. +\,2\left(\frac{1}{r_2}-\frac{1}{r_1}\right)\left\{\psi\left(\frac{1}{r_1}\right)-\psi\left(\frac{1}{r_2}\right)\right\}-2\right) \\
&\quad \times \left(\frac{1}{r_1}+\frac{1}{r_2}+1\right)^{-1}\frac{1}{\sigma^2}+\frac{1}{\sigma^2},
\end{aligned}
$$

$$
\begin{aligned}
S^{\star}_{23} &= \left[\frac{1}{r_2}\left\{\psi\left(\frac{1}{r_1}\right)-\psi\left(\frac{1}{r_2}\right)\right\}-1\right]\left(\frac{1}{r_1}+\frac{1}{r_2}\right)^{-1}\frac{1}{\sigma}, \\
S^{\star}_{24} &= \left[-\frac{1}{r_1}\left\{\psi\left(\frac{1}{r_1}\right)-\psi\left(\frac{1}{r_2}\right)\right\}-1\right]\left(\frac{1}{r_1}+\frac{1}{r_2}\right)^{-1}\frac{1}{\sigma}, \\
S^{\star}_{33} &= \psi^{'}\left(\frac{1}{r_1}\right)-\psi^{'}\left(\frac{1}{r_1}+\frac{1}{r_2}\right), \\
S^{\star}_{34} &= -\psi^{'}\left(\frac{1}{r_1}+\frac{1}{r_2}\right), \\
S^{\star}_{44} &= \psi^{'}\left(\frac{1}{r_2}\right)-\psi^{'}\left(\frac{1}{r_1}+\frac{1}{r_2}\right).
\end{aligned}
$$

Thus, based on theorem (Theorem 3), the information matrix under the new parameterization is

$$
\begin{aligned}
I(\mu, \sigma, r_1, r_2) &= A(C^* S^* C^{*\prime})A' \\
&= (AC^*)S^*(AC^*)'
\end{aligned}
\tag{3.1}
$$

where matrix AC^* is given by

$$
AC^* = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -\dfrac{1}{r_1^2} & 0 \\ 0 & 0 & 0 & -\dfrac{1}{r_2^2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ \log\left(\dfrac{r_1}{r_2}\right) & 1 & 0 & 0 \\ -\sigma r_1 & 0 & 1 & 0 \\ \sigma r_2 & 0 & 0 & 1 \end{bmatrix}
$$

$$
= \begin{bmatrix} 1 & 0 & 0 & 0 \\ \log\left(\dfrac{r_1}{r_2}\right) & 1 & 0 & 0 \\ \dfrac{\sigma}{r_1} & 0 & -\dfrac{1}{r_1^2} & 0 \\ -\dfrac{\sigma}{r_2} & 0 & 0 & -\dfrac{1}{r_2^2} \end{bmatrix}.
$$

From equation (3.1), $I(\mu, \sigma, r_1, r_2)$ denoted by T is equal to

$$T = I(\mu,\sigma,r_1,r_2) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \log\left(\frac{r_1}{r_2}\right) & 1 & 0 & 0 \\ \frac{\sigma}{r_1} & 0 & -\frac{1}{r_1^2} & 0 \\ -\frac{\sigma}{r_2} & 0 & 0 & -\frac{1}{r_2^2} \end{bmatrix} \begin{bmatrix} S^{\star}_{11} & S^{\star}_{12} & S^{\star}_{13} & S^{\star}_{14} \\ S^{\star}_{21} & S^{\star}_{22} & S^{\star}_{23} & S^{\star}_{24} \\ S^{\star}_{31} & S^{\star}_{32} & S^{\star}_{33} & S^{\star}_{34} \\ S^{\star}_{41} & S^{\star}_{42} & S^{\star}_{43} & S^{\star}_{44} \end{bmatrix}$$

$$\times \begin{bmatrix} 1 & \log\left(\frac{r_1}{r_2}\right) & \frac{\sigma}{r_1} & -\frac{\sigma}{r_2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -\frac{1}{r_1^2} & 0 \\ 0 & 0 & 0 & -\frac{1}{r_2^2} \end{bmatrix}$$

$$= \begin{bmatrix} T_{11} & T_{12} & T_{13} & T_{14} \\ T_{21} & T_{22} & T_{23} & T_{24} \\ T_{31} & T_{32} & T_{33} & T_{34} \\ T_{41} & T_{42} & T_{43} & T_{44} \end{bmatrix} \qquad (3.2)$$

where the elements are defined as

$$
\begin{aligned}
T_{11} &= S^{\star}_{11} \\
T_{12} &= S^{\star}_{11}\log\left(\frac{r_1}{r_2}\right) + S^{\star}_{12} \\
T_{13} &= \frac{\sigma}{r_1}S^{\star}_{11} - \frac{1}{r_1^2}\,S^{\star}_{13} \\
T_{14} &= -\frac{\sigma}{r_1}S^{\star}_{11} - \frac{1}{r_2^2}\,S^{\star}_{14} \\
T_{22} &= \left[S^{\star}_{11}\log\left(\frac{r_1}{r_2}\right) + S^{\star}_{21}\right]\log\left(\frac{r_1}{r_2}\right) + S^{\star}_{12}\log\left(\frac{r_1}{r_2}\right) + S^{\star}_{22} \\
T_{23} &= \frac{\sigma}{r_1}\left[S^{\star}_{11}\log\left(\frac{r_1}{r_2}\right) + S^{\star}_{21}\right] - \frac{1}{r_1^2}\left[S^{\star}_{13}\log\left(\frac{r_1}{r_2}\right) + S^{\star}_{23}\right] \\
T_{24} &= -\frac{\sigma}{r_2}\left[S^{\star}_{11}\log\left(\frac{r_1}{r_2}\right) + S^{\star}_{21}\right] - \frac{1}{r_2^2}\left[S^{\star}_{14}\log\left(\frac{r_1}{r_2}\right) + S^{\star}_{24}\right] \\
T_{33} &= \frac{\sigma}{r_1}\left[\frac{\sigma}{r_1}S^{\star}_{11} - \frac{1}{r_1^2}S^{\star}_{31}\right] - \frac{1}{r_1^2}\left[\frac{\sigma}{r_1}S^{\star}_{13} - \frac{1}{r_1^2}S^{\star}_{33}\right] \\
T_{34} &= -\frac{\sigma}{r_2}\left[\frac{\sigma}{r_1}S^{\star}_{11} - \frac{1}{r_1^2}S^{\star}_{31}\right] - \frac{1}{r_2^2}\left[\frac{\sigma}{r_1}S^{\star}_{14} - \frac{1}{r_1^2}S^{\star}_{34}\right] \\
T_{44} &= -\frac{\sigma}{r_2}\left[-\frac{\sigma}{r_2}S^{\star}_{11} - \frac{1}{r_2^2}S^{\star}_{41}\right] - \frac{1}{r_2^2}\left[-\frac{\sigma}{r_2}S^{\star}_{14} - \frac{1}{r_2^2}S^{\star}_{44}\right]
\end{aligned}
$$

Note that $T_{ij} = T_{ji}$, $i = 1, ..., 4$ and $j = 1, ..., 4$.

If r_1 is known and equal to 1, by definition of the expected information matrix,

$$I(\mu,\sigma,r_1,r_2) = \begin{bmatrix} cov\left(\frac{\partial \ell}{\partial \mu},\frac{\partial \ell}{\partial \mu}\right) & cov\left(\frac{\partial \ell}{\partial \mu},\frac{\partial \ell}{\partial \sigma}\right) & cov\left(\frac{\partial \ell}{\partial \mu},\frac{\partial \ell}{\partial r_1}\right) & cov\left(\frac{\partial \ell}{\partial \mu},\frac{\partial \ell}{\partial r_2}\right) \\ & cov\left(\frac{\partial \ell}{\partial \sigma},\frac{\partial \ell}{\partial \sigma}\right) & cov\left(\frac{\partial \ell}{\partial \sigma},\frac{\partial \ell}{\partial r_1}\right) & cov\left(\frac{\partial \ell}{\partial \sigma},\frac{\partial \ell}{\partial r_2}\right) \\ & & cov\left(\frac{\partial \ell}{\partial r_1},\frac{\partial \ell}{\partial r_1}\right) & cov\left(\frac{\partial \ell}{\partial r_1},\frac{\partial \ell}{\partial r_2}\right) \\ & & & cov\left(\frac{\partial \ell}{\partial r_2},\frac{\partial \ell}{\partial r_2}\right) \end{bmatrix},$$

all covariance terms pertaining to r_1, i.e. the third row and column are ignored. Hence, the information matrix (3.2) reduces to

$$I(\mu,\sigma,r_2)\Big|_{r_1=1} = \begin{bmatrix} T_{11} & T_{12} & T_{14} \\ T_{21} & T_{22} & T_{24} \\ T_{41} & T_{42} & T_{44} \end{bmatrix}. \tag{3.3}$$

Further at $r_2 = 1$, the elements in (3.3) simplify to

$$\begin{aligned} T_{11} &= S^{\star}_{11} = \frac{1}{3\sigma^2} \\ T_{12} &= S^{\star}_{12} = 0 \\ T_{14} &= -\sigma S^{\star}_{11} - S^{\star}_{14} = -\frac{1}{3\sigma} + \frac{1}{2\sigma} = \frac{1}{6\sigma} \end{aligned}$$

$$
\begin{aligned}
T_{22} &= S^{\star}_{22} = \left(2\psi'(1) - 2\right)\frac{1}{3\sigma^2} + \frac{1}{\sigma^2} = \frac{2\psi'(1) + 1}{3\sigma^2} \\
T_{24} &= -\sigma S^{\star}_{21} - S^{\star}_{24} = \frac{1}{2\sigma} \\
T_{44} &= \sigma^2 S^{\star}_{11} + \sigma S^{\star}_{41} + \sigma S^{\star}_{14} + S^{\star}_{44} \\
&= \frac{1}{3} + 2\sigma\left(-\frac{1}{2\sigma}\right) + \psi'(1) - \psi'(2) \\
&= -\frac{2}{3} + \psi'(1) - \psi'(2)
\end{aligned}
$$

so that matrix (3.3) becomes,

$$
I(\mu, \sigma, r_2 = 1)\Big|_{r_1=1} = \begin{bmatrix} \frac{1}{3\sigma^2} & 0 & \frac{1}{6\sigma} \\ 0 & \frac{\pi^2+3}{9\sigma^2} & \frac{1}{2\sigma} \\ \frac{1}{6\sigma} & \frac{1}{2\sigma} & \frac{1}{3} \end{bmatrix}. \tag{3.4}
$$

Note that T_{22} and T_{44} were simplified using the following properties of the trigamma function $\psi'(k) = \frac{\partial}{\partial k}\psi(k)$ (Abramowitz and Stegun, 1964):

a. $\psi'(1) = (-1)^2\, 1!\, \zeta(2)$, where $\zeta()$ is the Riemann zeta function

$$
= \frac{\pi^2}{6}
$$

b. Recurrence Formula

$$\begin{aligned}\psi'(k+1) &= \psi'(k) + (-1)\ 1!\ k^{-2} \\ &= \psi'(k) - \frac{1}{k^2}\end{aligned}$$

At $k = 1$,

$$\psi'(1+1) = \psi'(2) = \psi'(1) - 1 = \frac{\pi^2}{6} - 1$$

Now, since the inverse of (3.4) is

$$I^{-1}(\mu, \sigma, r_2 = 1)\Big|_{r_1=1} = \frac{108\sigma^4}{\pi^2 - 6} \begin{bmatrix} \frac{4\pi^2 - 15}{108\sigma^2} & \frac{1}{12\sigma^2} & \frac{-(\pi^2+3)}{54\sigma^3} \\ \frac{1}{12\sigma^2} & \frac{1}{12\sigma^2} & -\frac{1}{6\sigma^3} \\ \frac{-(\pi^2+3)}{54\sigma^3} & -\frac{1}{6\sigma^3} & \frac{\pi^2+3}{27\sigma^4} \end{bmatrix},$$

theorem (Theorem 2) implies that under the Logistic hypothesis, $r_2 = 1$,

$$\sqrt{n}\ (\hat{r_2} - 1) \ \xrightarrow{d}\ N\left\{0, \frac{108\sigma^4}{\pi^2 - 6}\left(\frac{\pi^2+3}{27\sigma^4}\right)\right\}.$$

That is,

$$\hat{r_2} \ \xrightarrow{d}\ N\left\{1, \frac{4}{n}\left(\frac{\pi^2+3}{\pi^2-6}\right)\right\}. \tag{3.5}$$

3.1.2 Extreme Minimum Value Model ($r_1 = 1, r_2 = 0$)

From (R3b) of Section 2.3, the information about (μ, σ, r_1, r_2) at $r_1 = 1$ and $r_2 = 0$ is

$$I(\mu, \sigma, r_1 = 1, r_2 = 0) = \begin{bmatrix} \frac{1}{\sigma^2} & \frac{1-\gamma}{\sigma^2} & 0 & \frac{1}{\sigma} \\ \frac{1-\gamma}{\sigma^2} & \frac{\frac{\pi^2}{6} + (1-\gamma)^2}{\sigma^2} & \frac{1}{\sigma} & \frac{2-\gamma}{\sigma} \\ 0 & \frac{1}{\sigma} & \frac{\pi^2}{6} - 1 & \frac{1}{2} \\ \frac{1}{\sigma} & \frac{2-\gamma}{\sigma} & \frac{1}{2} & 2 \end{bmatrix}. \quad (3.6)$$

Assuming that r_1 is known and equal to 1, the information matrix (3.6) reduces to

$$I(\mu, \sigma, r_2 = 0)\Big|_{r_1=1} = \begin{bmatrix} \frac{1}{\sigma^2} & \frac{1-\gamma}{\sigma^2} & \frac{1}{\sigma} \\ \frac{1-\gamma}{\sigma^2} & \frac{\frac{\pi^2}{6} + (1-\gamma)^2}{\sigma^2} & \frac{2-\gamma}{\sigma} \\ \frac{1}{\sigma} & \frac{2-\gamma}{\sigma} & 2 \end{bmatrix}. \quad (3.7)$$

This result can also be verified using the error density in (R1) and the limit of $\frac{\partial \ell}{\partial r_2}$ in (R3a) of Section 2.3, as shown in the Appendix.

Taking the inverse of matrix (3.7), we have

$$I^{-1}(\mu,\sigma,r_2=0)\Big|_{r_1=1} = \frac{\sigma^4}{\dfrac{\pi^2}{6}-1}\begin{bmatrix} \dfrac{\dfrac{\pi^2}{3}-2+\gamma^2}{\sigma^2} & \dfrac{\gamma}{\sigma^2} & \dfrac{1-\gamma-\dfrac{\pi^2}{6}}{\sigma^3} \\ \dfrac{\gamma}{\sigma^2} & \dfrac{1}{\sigma^2} & -\dfrac{1}{\sigma^3} \\ \dfrac{1-\gamma-\dfrac{\pi^2}{6}}{\sigma^3} & -\dfrac{1}{\sigma^3} & \dfrac{\dfrac{\pi^2}{6}}{\sigma^4} \end{bmatrix}.$$

Hence, it follows from theorem (Theorem 2) that under the Extreme minimum value hypothesis, $r_2 = 0$,

$$\sqrt{n}\,(\hat{r_2}-0) \xrightarrow{d} N\left\{0,\ \frac{\dfrac{\pi^2}{6}}{\dfrac{\pi^2}{6}-1}\right\}.$$

Consequently,

$$\hat{r_2} \xrightarrow{d} N\left\{0,\ \frac{1}{n\left(1-\dfrac{6}{\pi^2}\right)}\right\}. \tag{3.8}$$

Notice that the asymptotic results (3.5) and (3.8) are identical to Prentice's results (R3c) of Section 2.3 even if the nuisance parameters were not removed from the likelihood function.

3.2 Discrimination Between Logistic and Extreme Minimum Value Distributions

We propose in this section three discrimination tests between Logistic and Extreme minimum value distributions. The first test uses the Logistic distribution as default unless there is significant evidence otherwise. The second test uses the Extreme minimum value distribution as default unless there is significant evidence otherwise. Adding to these tests, we propose a third test which simply uses the midpoint of the null and alternative hypotheses values as the cut-off point in the rejection rule.

Let $y_i = \mu + \sigma\epsilon_i$, $i = 1, 2, ...n$ be a random sample of size n with an unknown distribution form. To determine if the Extreme minimum value or the Logistic distribution is a better fit to the data, one of the following α level tests may be used:

1. Hypotheses

   H_0: Logistic versus H_1: Extreme minimum value

   or equivalently,

   $H_0 : r_2 = 1$ versus $H_1 : r_2 < 1$

   Rejection Rule

   The asymptotic result (3.5) implies rejection of H_0 if

$$\frac{\hat{r}_2 - 1}{\sqrt{\frac{13.30}{n}}} < Z_\alpha.$$

2. Hypotheses

   H_0: Extreme minimum value versus H_1: Logistic

   or equivalently,

   $H_0 : r_2 = 0$ versus $H_1 : r_2 > 0$

   Rejection Rule

   The asymptotic result (3.8) implies rejection of H_0 if

$$\frac{\hat{r}_2}{\sqrt{\frac{2.55}{n}}} > Z_\alpha.$$

3. Hypotheses

   H_0: Extreme minimum value versus H_1: Logistic

   or equivalently,

   $H_0 : r_2 = 0$ versus $H_1 : r_2 = 1$

   Rejection Rule

   Reject H_0 if $\hat{r}_2 > 0.5$.

Keep in mind that these tests merely determine the model which better fits the data, and does not prove that the chosen model is the correct distribution form.

3.3 Two-Sample Accelerated Failure Time Test

First, we prepare a procedure for applying the discrimination test within the framework of a two-sample accelerated failure time model.

Let $T_1 = (T_{11}, T_{12}, ...T_{1n_1})$ and $T_2 = (T_{21}, T_{22}, ...T_{2n_2})$ be random samples of failure times whose log values are denoted by the following respective location-scale models:

$$\log T_1 = \mu + \sigma\epsilon \tag{3.9}$$

$$\log T_2 = \mu + \beta + \sigma\epsilon, \tag{3.10}$$

where the distribution form of ϵ is unknown.

Based from models (3.9) and (3.10), the relationship between $\log T_1$ and $\log T_2$ is characterized by the location-change model

$$\log T_2 = \beta + \log T_1 \tag{3.11}$$

while the relationship between T_1 and T_2 is characterized by the scale-change model

$$T_2 = e^{\beta}\, T_1 \tag{3.12}$$

In testing $\beta = 0$, it is known that the Log-Rank test is asymptotically fully efficient for the Extreme minimum value error distribution while the Peto-Peto's Wilcoxon test is asymptotically fully efficient for the Logistic error distribution.

Now we can use the discrimination test in Section 3.2 on the combined log failure times to compare the fit of the Extreme minimum value and Logistic discributions on the data. However, we first need to standardize the two samples to the same location and scale so the discrimination test can select the more appropriate shape. Standardization can be attained by either recentering the log failure times to the same location or by rescaling the failure times to the same spread. We will use the latter approach in this paper.

From model (3.12), the scale-change parameter or acceleration factor, is $e^{\beta} = \frac{T_2}{T_1}$, which can be intuitively estimated by the ratio of the two standard deviations, i.e. $\frac{SD(T_2)}{SD(T_1)}$. Hence, multiplying T_1 by this ratio,

$$T_1^{\star} = \frac{SD(T_2)}{SD(T_1)} T_1 ,$$

results to a rescaled $T_1^{\star}$ which exhibits the same variability as T_2.

We now define an adaptive test procedure for $H_0 : \beta = 0$ as follows:

1. Let T_1 and C_1 be the failure times and censored times of the first sample, respectively.

2. Let T_2 and C_2 be the failure times and censored times of the second sample, respectively.

3. Rescale the failure times from the two samples so they have the same variability:

$$T_1^{\star} = T_1 \times \frac{SD(T_2)}{SD(T_1)} .$$

4. Combine rescaled failure times T_c =(T_1^* and T_2).

5. Log-transform T_c and compute $\hat{m}_2$ using R's *optim* function.

6. Take $\hat{r}_2 = \frac{1}{\hat{m}_2}$ and perform one of the three discrimination tests described in Section 3.2. If the Extreme minimum value distribution is selected, use Log-Rank test. Otherwise, if the Logistic distribution is chosen, use Peto-Peto's Wilcoxon test.

The adaptive procedure produces three tests which differ by their discrimination scheme for choosing between the Log-Rank and Peto-Peto's Wilcoxon tests:

1. Test DL (Default Log-Rank): Use Log-Rank unless

$$\frac{\hat{r}_2}{\sqrt{\frac{2.55}{n}}} > Z_\alpha.$$

2. Test DW (Default Wilcoxon): Use Wilcoxon unless

$$\frac{\hat{r}_2 - 1}{\sqrt{\frac{13.30}{n}}} < Z_\alpha.$$

3. Test EQ (Equal Discrimination):

   If $\hat{r}_2 > 0.5$ use Wilcoxon. Else, use Log-Rank.

Chapter 4

Simulation

We examine the finite sample performance of the three adaptive survival tests for the two-sample accelerated failure time problem. The tests differ by their preliminary procedure for choosing between the Log-Rank and Peto-Peto's Wilcoxon tests:

1. Test DL (Default Log-Rank): Use Log-Rank unless
$$\frac{\hat{r}_2}{\sqrt{\frac{2.55}{n}}} > Z_\alpha. \tag{4.1}$$

2. Test DW (Default Wilcoxon): Use Wilcoxon unless
$$\frac{\hat{r}_2 - 1}{\sqrt{\frac{13.30}{n}}} < Z_\alpha. \tag{4.2}$$

3. Test EQ (Equal Discrimination):
$$\text{If } \hat{r}_2 > 0.5 \text{ use Wilcoxon. Else, use Log-Rank.} \tag{4.3}$$

In this chapter, we will assess the empirical validity of these tests. We will also compare performance among the adaptive tests, and between the adaptive and non-adaptive tests based on their size and power.

4.1 Simulation Models

We describe in detail the failure time and censoring models considered in the simulation study.

Failure times, T_1 and T_2, were generated from the following distributions:

1. Weibull

$$f(t) = \alpha\lambda t^{\alpha-1}\exp(-\lambda x^{\alpha}), \qquad \alpha, \lambda > 0$$

$$S(t) = e^{-\lambda t^{\alpha}}$$

or equivalently,

$$f(t) = \frac{1}{\sigma}\, e^{(-\mu/\sigma)}\, t^{(1/\sigma-1)} \exp\left[-e^{(-\mu/\sigma)}\, t^{(1/\sigma)}\right], \qquad \sigma > 0$$

$$S(t) = \exp\left[-e^{(-\mu/\sigma)}\, t^{(1/\sigma)}\right]$$

Note that $\alpha = \frac{1}{\sigma}$ and $\lambda = e^{(-\mu/\sigma)}$.

2. Log-logistic

$$f(t) = \frac{\alpha\lambda t^{\alpha-1}}{(1+\lambda t^{\alpha})^2}, \qquad \alpha, \lambda > 0$$

$$S(t) = \frac{1}{1+\lambda t^{\alpha}}$$

or equivalently,

$$f(t) = \frac{\frac{1}{\sigma} e^{(-\mu/\sigma)} t^{(1/\sigma - 1)}}{[1 + e^{(-\mu/\sigma)} t^{(1/\sigma)}]^2}, \qquad \sigma > 0$$

$$S(t) = \frac{1}{1 + e^{(-\mu/\sigma)} t^{(1/\sigma)}}$$

Again, $\alpha = \frac{1}{\sigma}$ and $\lambda = e^{(-\mu/\sigma)}$.

3. Log-normal

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\left[\frac{1}{2}\left(\frac{\ln(t)-\mu}{\sigma}\right)^2\right]}, \qquad \sigma > 0$$

$$S(t) = 1 - \Phi\left[\frac{\ln(t) - \mu}{\sigma}\right]$$

For each of these failure time distributions, the shape parameter σ is set to (a) greater than 1, (b) equal to 1, and (c) less than 1 to represent different degrees of skewness and tail weight of the survival function. The scale parameter for S_1 is $\mu = 0$ in all cases. In the size study, the scale parameter for S_2 is also $\mu = 0$ while in the power study, the scale parameter for S_2 is $\mu > 0$. The various shapes of the probability density and survival functions in the power study are presented in Figures 4.1 and 4.2, respectively.

Figure 4.1: Shapes of the Probability Density Function in the Power Simulation Study

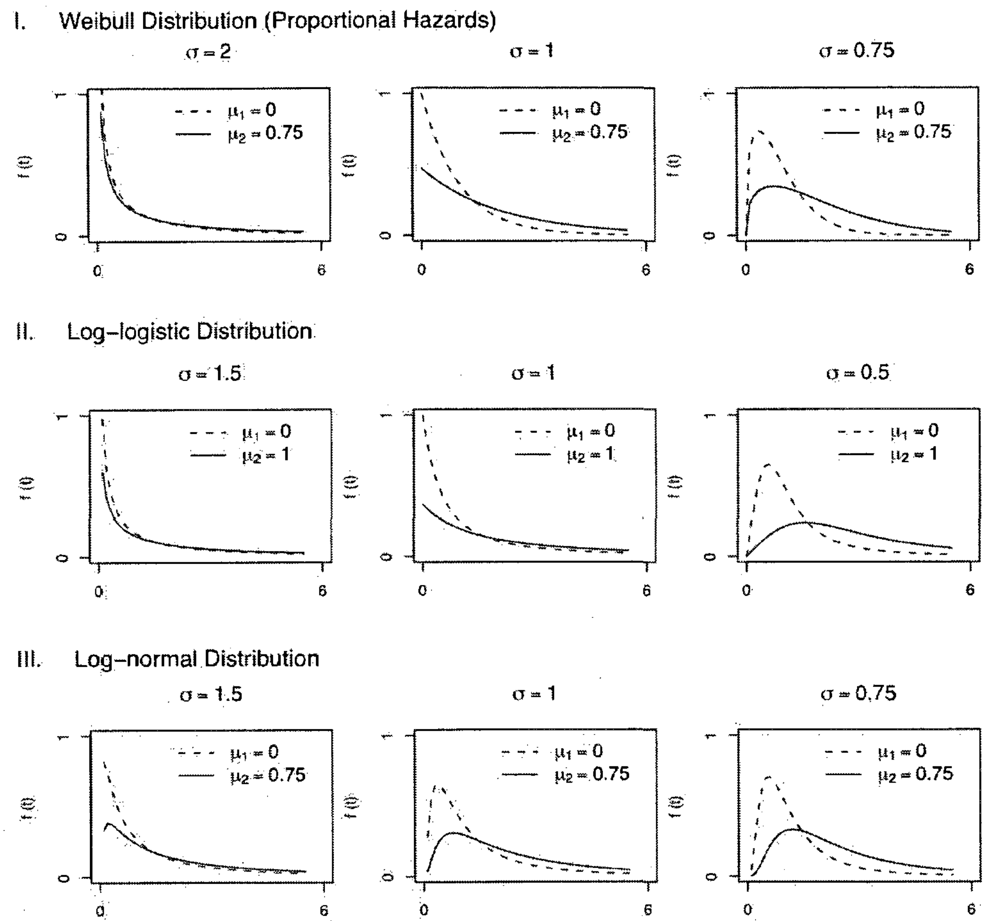
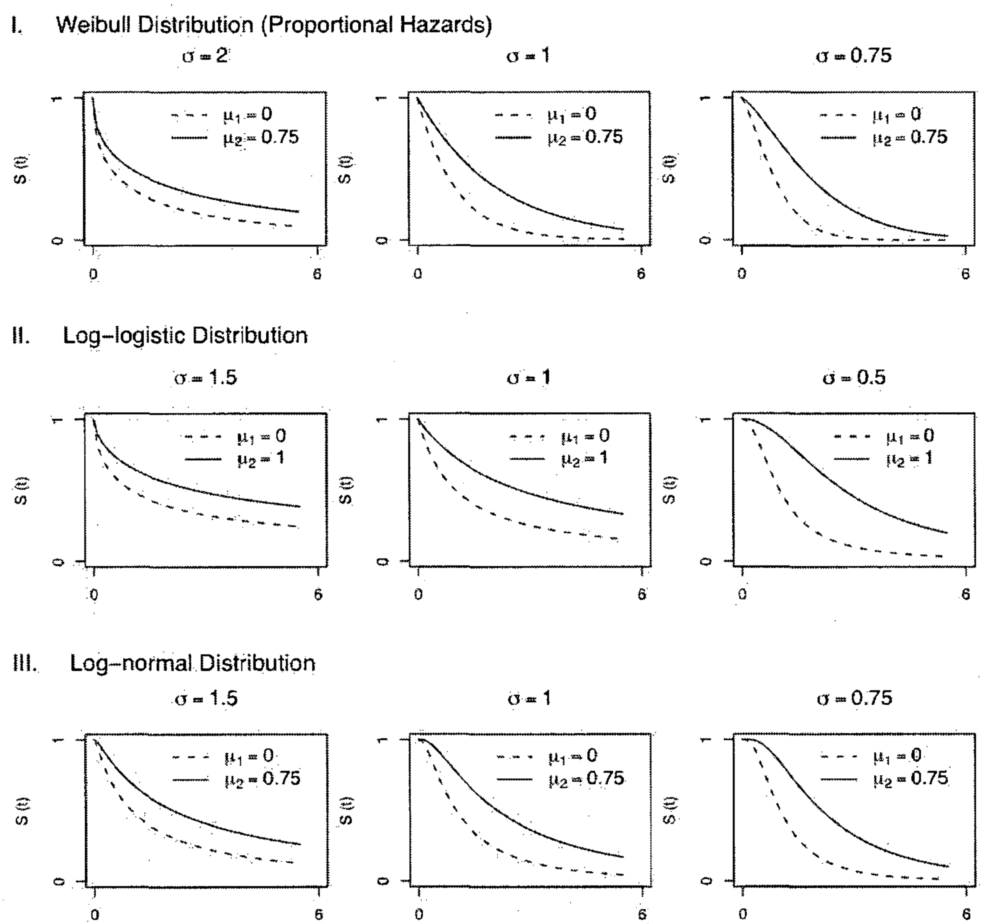


Figure 4.2: Shapes of the Survival Function in the Power Simulation Study



The performance of the tests was examined on both uncensored and right-censored failure times. Under the censored setting, the censoring distribution of the two groups were equal and uniform $U(0, c)$. Various values of c, which correspond to the percentage of censored observations, were considered in the simulation.

4.2 Size Study

The adaptive and non-adaptive tests were performed on 10,000 samples under each survival and censoring configuration. Results at the 5% significance level for sample sizes 20 and 50 are presented in Tables 4.1 and 4.2, respectively.

Test DL (Default Log-Rank), Test DW (Default Wilcoxon) and Test EQ (Equal Discrimination) are based on the discrimination scheme in equations (4.1), (4.2) and (4.3), respectively. All simulations were done in R, using the *optim* function to compute the $\hat{r_2}$ test statistic and the *survdiff* function to obtain the Log-Rank and Peto-Peto's Wilcoxon test statistics.

Table 4.1: Size Simulation Results at Significance Level $\alpha = 0.05$, from 10,000 Replications with Equal Censoring, $n_1 = n_2 = 20$

Relative Skewness	Survival Time	Average % Censored	Test DL	Test DW	Test EQ	Log-Rank	Wilcoxon
$S(t)$=Weibull Distribution							
		$C_1 = 0.0$, $C_2 = 0.0$	0.0603	0.0586	0.0599	0.0586	0.0522
heavy	$\mu_1 = 0, \sigma_1 = 2$	$C_1 = 23.2$, $C_2 = 23.4$	0.0517	0.0519	0.0519	0.0519	0.0505
	$\mu_2 = 0, \sigma_2 = 2$	$C_1 = 29.7$, $C_2 = 29.7$	0.0540	0.0538	0.0540	0.0538	0.0503
		$C_1 = 0.0$, $C_2 = 0.0$	0.0603	0.0583	0.0596	0.0583	0.0560
moderate	$\mu_1 = 0, \sigma_1 = 1$	$C_1 = 16.6$, $C_2 = 16.6$	0.0561	0.0555	0.0558	0.0556	0.0520
	$\mu_2 = 0, \sigma_2 = 1$	$C_1 = 24.3$, $C_2 = 24.6$	0.0559	0.0561	0.0558	0.0560	0.0511
		$C_1 = 0.0$, $C_2 = 0.0$	0.0612	0.0601	0.0613	0.0600	0.0559
light	$\mu_1 = 0, \sigma_1 = 0.75$	$C_1 = 15.3$, $C_2 = 15.3$	0.0570	0.0566	0.0571	0.0564	0.0514
	$\mu_2 = 0, \sigma_2 = 0.75$	$C_1 = 22.8$, $C_2 = 22.8$	0.0562	0.0559	0.0565	0.0558	0.0549
$S(t)$=Log-logistic Distribution							
		$C_1 = 0.0$, $C_2 = 0.0$	0.0673	0.0624	0.0678	0.0594	0.0538
heavy	$\mu_1 = 0, \sigma_1 = 1.5$	$C_1 = 22.4$, $C_2 = 22.3$	0.0571	0.0572	0.0571	0.0572	0.0523
	$\mu_2 = 0, \sigma_2 = 1.5$	$C_1 = 35.3$, $C_2 = 35.4$	0.0525	0.0523	0.0523	0.0525	0.0503
		$C_1 = 0.0$, $C_2 = 0.0$	0.0673	0.0614	0.0678	0.0587	0.0507
moderate	$\mu_1 = 0, \sigma_1 = 1$	$C_1 = 15.1$, $C_2 = 15.2$	0.0517	0.0517	0.0526	0.0515	0.0472
	$\mu_2 = 0, \sigma_2 = 1$	$C_1 = 29.6$, $C_2 = 29.5$	0.0545	0.0536	0.0544	0.0533	0.0510

Table 4.1 — Continued

Relative Skewness	Survival Time	Average % Censored	Test DL	Test DW	Test EQ	Log-Rank	Wilcoxon
		$C_1 = 0.0$, $C_2 = 0.0$	0.0662	0.0584	0.0658	0.0561	0.0517
light	$\mu_1 = 0, \sigma_1 = 0.5$	$C_1 = 7.6$, $C_2 = 7.6$	0.0633	0.0575	0.0633	0.0570	0.0517
	$\mu_2 = 0, \sigma_2 = 0.5$	$C_1 = 20.4$, $C_2 = 20.5$	0.0582	0.0551	0.0578	0.0545	0.0484
$S(t)$=Log-normal Distribution							
		$C_1 = 0.0$, $C_2 = 0.0$	0.0647	0.0595	0.0653	0.0574	0.0531
heavy	$\mu_1 = 0, \sigma_1 = 1.5$	$C_1 = 16.3$, $C_2 = 16.4$	0.0587	0.0569	0.0587	0.0567	0.0508
	$\mu_2 = 0, \sigma_2 = 1.5$	$C_1 = 31.2$, $C_2 = 31.1$	0.0554	0.0552	0.0557	0.0556	0.0522
		$C_1 = 0.0$, $C_2 = 0.0$	0.0694	0.0624	0.0690	0.0608	0.0527
moderate	$\mu_1 = 0, \sigma_1 = 1$	$C_1 = 10.9$, $C_2 = 10.8$	0.0606	0.0564	0.0605	0.0555	0.0505
	$\mu_2 = 0, \sigma_2 = 1$	$C_1 = 25.2$, $C_2 = 25.4$	0.0612	0.0600	0.0612	0.0595	0.0565
		$C_1 = 0.0$, $C_2 = 0.0$	0.0635	0.0586	0.0639	0.0564	0.0513
light	$\mu_1 = 0, \sigma_1 = 0.75$	$C_1 = 8.9$, $C_2 = 8.9$	0.0644	0.0606	0.0645	0.0591	0.0509
	$\mu_2 = 0, \sigma_2 = 0.75$	$C_1 = 21.9$, $C_2 = 21.8$	0.0602	0.0569	0.0603	0.0561	0.0499

Table 4.2: Size Simulation Results at Significance Level $\alpha = 0.05$, from 10,000 Replications with Equal Censoring, $n_1 = n_2 = 50$

Relative Skewness	Survival Time	Average % Censored	Test DL	Test DW	Test EQ	Log-Rank	Wilcoxon
$S(t)$=Weibull Distribution							
heavy	$\mu_1 = 0, \sigma_1 = 2$	$C_1 = 0.0$, $C_2 = 0.0$	0.0550	0.0542	0.0545	0.0542	0.0485
	$\mu_2 = 0, \sigma_2 = 2$	$C_1 = 29.8$, $C_2 = 29.7$	0.0552	0.0552	0.0552	0.0552	0.0556
		$C_1 = 41.4$, $C_2 = 41.3$	0.0489	0.0489	0.0489	0.0489	0.0484
moderate	$\mu_1 = 0, \sigma_1 = 1$	$C_1 = 0.0$, $C_2 = 0.0$	0.0567	0.0556	0.0556	0.0522	0.0496
	$\mu_2 = 0, \sigma_2 = 1$	$C_1 = 16.6$, $C_2 = 16.4$	0.0540	0.0541	0.0540	0.0541	0.0510
		$C_1 = 43.3$, $C_2 = 43.3$	0.0503	0.0502	0.0502	0.0502	0.0488
light	$\mu_1 = 0, \sigma_1 = 0.75$	$C_1 = 0.0$, $C_2 = 0.0$	0.0567	0.0556	0.0556	0.0556	0.0487
	$\mu_2 = 0, \sigma_2 = 0.75$	$C_1 = 22.9$, $C_2 = 23.0$	0.0481	0.0479	0.0477	0.0479	0.0480
		$C_1 = 43.9$, $C_2 = 43.7$	0.0501	0.0501	0.0501	0.0501	0.0503
$S(t)$=Log-logistic Distribution							
heavy	$\mu_1 = 0, \sigma_1 = 1.5$	$C_1 = 0.0$, $C_2 = 0.0$	0.0536	0.0543	0.0569	0.0529	0.0498
	$\mu_2 = 0, \sigma_2 = 1.5$	$C_1 = 30.5$, $C_2 = 30.5$	0.0511	0.0512	0.0513	0.0512	0.0476
		$C_1 = 43.4$, $C_2 = 43.5$	0.0521	0.0520	0.0519	0.0520	0.0491
moderate	$\mu_1 = 0, \sigma_1 = 1$	$C_1 = 0.0$, $C_2 = 0.0$	0.0540	0.0549	0.0597	0.0530	0.0485
	$\mu_2 = 0, \sigma_2 = 1$	$C_1 = 24.0$, $C_2 = 24.0$	0.0530	0.0526	0.0525	0.0526	0.0518
		$C_1 = 40.4$, $C_2 = 40.4$	0.0485	0.0485	0.0483	0.0485	0.0462

Table 4.2 — Continued

Relative Skewness	Survival Time	Average % Censored	Test DL	Test DW	Test EQ	Log-Rank	Wilcoxon
		$C_1 = 0.0$, $C_2 = 0.0$	0.0515	0.0556	0.0581	0.0549	0.0474
light	$\mu_1 = 0, \sigma_1 = 0.5$	$C_1 = 14.7$, $C_2 = 14.8$	0.0567	0.0545	0.0555	0.0543	0.0514
	$\mu_2 = 0, \sigma_2 = 0.5$	$C_1 = 33.2$, $C_2 = 33.1$	0.0536	0.0531	0.0526	0.0531	0.0510
$S(t)$=Log-normal Distribution							
		$C_1 = 0.0$, $C_2 = 0.0$	0.0553	0.0548	0.0592	0.0530	0.0505
heavy	$\mu_1 = 0, \sigma_1 = 1.5$	$C_1 = 25.8$, $C_2 = 25.8$	0.0522	0.0515	0.0519	0.0516	0.0525
	$\mu_2 = 0, \sigma_2 = 1.5$	$C_1 = 39.6$, $C_2 = 39.4$	0.0566	0.0559	0.0563	0.0559	0.0552
		$C_1 = 0.0$, $C_2 = 0.0$	0.0555	0.0569	0.0606	0.0559	0.0498
moderate	$\mu_1 = 0, \sigma_1 = 1$	$C_1 = 19.5$, $C_2 = 19.6$	0.0525	0.0528	0.0541	0.0529	0.0491
	$\mu_2 = 0, \sigma_2 = 1$	$C_1 = 35.1$, $C_2 = 35.0$	0.0589	0.0580	0.0583	0.0578	0.0546
		$C_1 = 0.0$, $C_2 = 0.0$	0.0525	0.0536	0.0564	0.0519	0.0469
light	$\mu_1 = 0, \sigma_1 = 0.75$	$C_1 = 16.5$, $C_2 = 16.5$	0.0554	0.0525	0.0579	0.0524	0.0503
	$\mu_2 = 0, \sigma_2 = 0.75$	$C_1 = 32.0$, $C_2 = 32.0$	0.0560	0.0538	0.0555	0.0535	0.0533

In general, all tests exhibited acceptable observed significance levels, with the Wilcoxon test being the most conservative in almost all cases. Also, the observed significance levels of the adaptive tests do not deviate substantially from the observed significance levels of the non-adaptive tests.

At $n = 20$, Test DW (Default Wilcoxon) demonstrated the most superior performance among the three adaptive tests under the Log-logistic and Log-normal survival distributions, with its observed significance levels being closest to the nominal level. At $n = 50$, the performance of the three adaptive tests were comparable.

4.3 Power Study

The power of the tests at a 5% significance level were obtained from 10,000 samples for each survival and censoring setting. Results for sample sizes 20 and 50 are displayed on Tables 4.3 thru 4.8.

Table 4.3: Power Simulation Results for Weibull Distribution at Significance Level $\alpha = 0.05$, from 10,000 Replications with Equal Censoring, $n_1 = n_2 = 20$

Relative Skewness	Survival Time	Average % Censored	Test DL	Test DW	Test EQ	Log-Rank	Wilcoxon
		$C_1 = 0.0$, $C_2 = 0.0$	0.2161	0.2141	0.2158	0.2141	0.1736
heavy	$\mu_1 = 0, \sigma_1 = 2$	$C_1 = 23.5$, $C_2 = 35.4$	0.1654	0.1656	0.1655	0.1655	0.1485
	$\mu_2 = 0.75, \sigma_2 = 2$	$C_1 = 29.6$, $C_2 = 42.2$	0.1574	0.1573	0.1575	0.1574	0.1482
		$C_1 = 0.0$, $C_2 = 0.0$	0.6260	0.6252	0.6258	0.6252	0.5222
moderate	$\mu_1 = 0, \sigma_1 = 1$	$C_1 = 16.8$, $C_2 = 33.2$	0.5119	0.5137	0.5122	0.5137	0.4469
	$\mu_2 = 0.75, \sigma_2 = 1$	$C_1 = 24.5$, $C_2 = 44.9$	0.4632	0.4643	0.4630	0.4644	0.4083
		$C_1 = 0.0$, $C_2 = 0.0$	0.8495	0.8485	0.8500	0.8485	0.7489
light	$\mu_1 = 0, \sigma_1 = 0.75$	$C_1 = 15.4$, $C_2 = 32.1$	0.7530	0.7544	0.7535	0.7546	0.6688
	$\mu_2 = 0.75, \sigma_2 = 0.75$	$C_1 = 22.9$, $C_2 = 45.9$	0.6984	0.7004	0.6990	0.7003	0.6294

Result: According to literature, Log-Rank is fully efficient under the Weibull survival distribution. The simulation confirms this with Log-Rank outperforming Wilcoxon in all cases. The difference in their power is minimal under the heavily skewed Weibull distribution with censoring but increases up to 10% as the degree of skewness decreases. All adaptive tests are as powerful as Log-Rank.

Table 4.4: Power Simulation Results for Weibull Distribution at Significance Level $\alpha = 0.05$, from 10,000 Replications with Equal Censoring, $n_1 = n_2 = 50$

Relative Skewness	Survival Time	Average % Censored	Test DL	Test DW	Test EQ	Log-Rank	Wilcoxon
		$C_1 = 0.0$, $C_2 = 0.0$	0.2336	0.2324	0.2329	0.2324	0.1897
heavy	$\mu_1 = 0, \sigma_1 = 2$	$C_1 = 29.7$, $C_2 = 38.1$	0.1713	0.1713	0.1713	0.1713	0.1611
	$\mu_2 = 0.5, \sigma_2 = 2$	$C_1 = 41.4$, $C_2 = 49.7$	0.1526	0.1527	0.1527	0.1527	0.1440
		$C_1 = 0.0$, $C_2 = 0.0$	0.6843	0.6851	0.6853	0.6851	0.5648
moderate	$\mu_1 = 0, \sigma_1 = 1$	$C_1 = 24.6$, $C_2 = 37.7$	0.5338	0.5344	0.5344	0.5344	0.4698
	$\mu_2 = 0.5, \sigma_2 = 1$	$C_1 = 43.4$, $C_2 = 57.9$	0.4027	0.4027	0.4027	0.4027	0.3840
		$C_1 = 0.0$, $C_2 = 0.0$	0.8955	0.8969	0.8970	0.8969	0.8061
light	$\mu_1 = 0, \sigma_1 = 0.75$	$C_1 = 22.9$, $C_2 = 37.1$	0.7738	0.7758	0.7756	0.7758	0.7028
	$\mu_2 = 0.5, \sigma_2 = 0.75$	$C_1 = 43.7$, $C_2 = 61.6$	0.6196	0.6208	0.6203	0.6208	0.5802

Result: Similar to the results for $n = 20$, Log-Rank beats Wilcoxon under the Weibull survival distribution in all cases, confirming what the literature says. All adaptive tests are as powerful as Log-Rank.

Table 4.5: Power Simulation Results for Log-logistic Distribution at Significance Level $\alpha = 0.05$, from 10,000 Replications with Equal Censoring, $n_1 = n_2 = 20$

Relative Skewness	Survival Time	Average % Censored	Test DL	Test DW	Test EQ	Log-Rank	Wilcoxon
heavy	$\mu_1 = 0, \sigma_1 = 1, 5$ $\mu_2 = 1, \sigma_2 = 1.5$	$C_1 = 0.0$, $C_2 = 0.0$	0.2346	0.2071	0.2340	0.1980	0.2233
		$C_1 = 22.5$, $C_2 = 34.6$	0.2020	0.2004	0.2017	0.2001	0.2119
		$C_1 = 35.2$, $C_2 = 50.1$	0.2042	0.2043	0.2044	0.2043	0.2068
moderate	$\mu_1 = 0, \sigma_1 = 1$ $\mu_2 = 1, \sigma_2 = 1$	$C_1 = 0.0$, $C_2 = 0.0$	0.4343	0.3945	0.4337	0.3802	0.4270
		$C_1 = 15.3$, $C_2 = 28.9$	0.3878	0.3828	0.3872	0.3825	0.4116
		$C_1 = 29.6$, $C_2 = 49.3$	0.3790	0.3776	0.3792	0.3776	0.3924
light	$\mu_1 = 0, \sigma_1 = 0.5$ $\mu_2 = 1, \sigma_2 = 0.5$	$C_1 = 0.0$, $C_2 = 0.0$	0.9189	0.8859	0.9165	0.8788	0.9378
		$C_1 = 7.6$, $C_2 = 19.6$	0.8968	0.8784	0.8940	0.8758	0.9236
		$C_1 = 20.2$, $C_2 = 46.6$	0.8727	0.8658	0.8712	0.8655	0.8979

Result: In theory, Wilcoxon is fully efficient under the Log-logistic survival distribution. The simulation results confirm this with Wilcoxon beating Log-Rank in all cases. Wilcoxon outperforms Log-Rank under the lightly skewed survival distribution by about 3% to 6% but not by much under the heavily skewed survival distribution. The adaptive tests are more powerful than Log-Rank, as most evident under the lightly skewed survival distribution. Among the three adaptive tests, Test DW is the least powerful.

Table 4.6: Power Simulation Results for Log-logistic Distribution at Significance Level $\alpha = 0.05$, from 10,000 Replications with Equal Censoring, $n_1 = n_2 = 50$

Relative Skewness	Survival Time	Average % Censored	Test DL	Test DW	Test EQ	Log-Rank	Wilcoxon
		$C_1 = 0.0$, $C_2 = 0.0$	0.3335	0.2842	0.3323	0.2790	0.3301
heavy	$\mu_1 = 0, \sigma_1 = 1.5$	$C_1 = 30.6$, $C_2 = 41.7$	0.2966	0.2965	0.2967	0.2965	0.3142
	$\mu_2 = 0.8, \sigma_2 = 1.5$	$C_1 = 43.5$, $C_2 = 55.5$	0.2720	0.2715	0.2716	0.2715	0.2841
		$C_1 = 0.0$, $C_2 = 0.0$	0.6171	0.5292	0.6114	0.5200	0.6210
moderate	$\mu_1 = 0, \sigma_1 = 1$	$C_1 = 23.9$, $C_2 = 38.0$	0.5515	0.5502	0.5502	0.5502	0.5896
	$\mu_2 = 0.8, \sigma_2 = 1$	$C_1 = 40.2$, $C_2 = 57.3$	0.5256	0.5251	0.5251	0.5251	0.5459
		$C_1 = 0.0$, $C_2 = 0.0$	0.9923	0.9775	0.9893	0.9758	0.9943
light	$\mu_1 = 0, \sigma_1 = 0.5$	$C_1 = 14.7$, $C_2 = 30.0$	0.9836	0.9743	0.9791	0.9742	0.9896
	$\mu_2 = 0.8, \sigma_2 = 0.5$	$C_1 = 33.1$, $C_2 = 59.1$	0.9659	0.9647	0.9647	0.9647	0.9769

Result: The simulation results support the theory that Wilcoxon is fully efficient under the Log-logistic survival distribution. Wilcoxon beats Log-Rank by up to 10% in the uncensored cases and up to 4% in the censored cases. Under the lightly skewed survival distribution, all tests are powerful with Wilcoxon leading Log-Rank by only about 1% to 2%. Tests DL and EQ are as powerful as Wilcoxon in the uncensored cases. For the censored cases, the adaptive tests are almost equivalent to Log-Rank, with Test DL having the highest power.

Table 4.7: Power Simulation Results for Log-normal Distribution at Significance Level $\alpha = 0.05$, from 10,000 Replications with Equal Censoring, $n_1 = n_2 = 20$

Relative Skewness	Survival Time	Average % Censored	Test DL	Test DW	Test EQ	Log-Rank	Wilcoxon
		$C_1 = 0.0$, $C_2 = 0.0$	0.3541	0.3266	0.3522	0.3189	0.3384
heavy	$\mu_1 = 0, \sigma_1 = 1.5$	$C_1 = 16.3$, $C_2 = 27.9$	0.3058	0.3003	0.3040	0.2992	0.3157
	$\mu_2 = 0.75, \sigma_2 = 1.5$	$C_1 = 31.3$, $C_2 = 47.1$	0.2761	0.2727	0.2755	0.2720	0.2891
		$C_1 = 0.0$, $C_2 = 0.0$	0.6232	0.5843	0.6211	0.5743	0.6207
moderate	$\mu_1 = 0, \sigma_1 = 1$	$C_1 = 11.1$, $C_2 = 21.9$	0.5771	0.5586	0.5765	0.5556	0.5919
	$\mu_2 = 0.75, \sigma_2 = 1$	$C_1 = 25.4$, $C_2 = 45.1$	0.5249	0.5166	0.5235	0.5161	0.5427
		$C_1 = 0.0$, $C_2 = 0.0$	0.8481	0.8212	0.8470	0.8146	0.8539
light	$\mu_1 = 0, \sigma_1 = 0.75$	$C_1 = 8.8$, $C_2 = 18.5$	0.8167	0.7975	0.8150	0.7931	0.8282
	$\mu_2 = 0.75, \sigma_2 = 0.75$	$C_1 = 21.7$, $C_2 = 42.8$	0.7643	0.7531	0.7634	0.7520	0.7849

Result: Under the Log-normal survival distribution, Wilcoxon beats Log-Rank by 2% to 5%, depending on the degree of skewness of the survival distribution and the censoring rate. The adaptive tests are more powerful than Log-Rank, with the adaptive tests' gain in power increasing as censoring rate decreases or as the survival distribution becomes less skewed. Among the three adaptive tests, Test DW is the least powerful.

Table 4.8: Power Simulation Results for Log-normal Distribution at Significance Level $\alpha = 0.05$, from 10,000 Replications with Equal Censoring, $n_1 = n_2 = 50$

Relative Skewness	Survival Time	Average % Censored	Test DL	Test DW	Test EQ	Log-Rank	Wilcoxon
heavy	$\mu_1 = 0, \sigma_1 = 1.5$ $\mu_2 = 0.5, \sigma_2 = 1.5$	$C_1 = 0.0$, $C_2 = 0.0$	0.3774	0.3499	0.3805	0.3445	0.3721
		$C_1 = 25.7$, $C_2 = 35.5$	0.3285	0.3205	0.3235	0.3204	0.3407
		$C_1 = 39.4$, $C_2 = 50.7$	0.2962	0.2953	0.2955	0.2951	0.3122
moderate	$\mu_1 = 0, \sigma_1 = 1$ $\mu_2 = 0.5, \sigma_2 = 1$	$C_1 = 0.0$, $C_2 = 0.0$	0.6805	0.6293	0.6790	0.6237	0.6799
		$C_1 = 19.7$, $C_2 = 30.1$	0.6141	0.5882	0.6041	0.5878	0.6343
		$C_1 = 35.0$, $C_2 = 49.8$	0.5612	0.5537	0.5570	0.5537	0.5902
light	$\mu_1 = 0, \sigma_1 = 0.75$ $\mu_2 = 0.5, \sigma_2 = 0.75$	$C_1 = 0.0$, $C_2 = 0.0$	0.8943	0.8544	0.8895	0.8508	0.8954
		$C_1 = 16.5$, $C_2 = 26.7$	0.8557	0.8188	0.8457	0.8176	0.8654
		$C_1 = 31.9$, $C_2 = 48.2$	0.7944	0.7753	0.7860	0.7749	0.8243

Result: Similar to the results for $n = 20$, Wilcoxon beats Log-Rank under the Log-normal survival distribution by 2% to 6%, depending on the degree of skewness of the survival distribution and the censoring rate. The adaptive tests are more powerful than Log-Rank. Generally, as the censoring rate decreases or as the survival distribution becomes less skewed, the gain in power of the adaptive tests against Log-Rank increases. Among the three adaptive tests, Test DW is the least powerful.

As expected, the power of the tests increases as the degree of skewness of the survival distribution decreases, i.e. $\sigma \downarrow$. Further, the power of the tests increases as the percentage of censored observations decreases.

In comparing the performance of the Log-Rank and Wilcoxon tests, the simulation results confirm that the Log-Rank test is more powerful than the Wilcoxon test under the Weibull survival distribution while the Wilcoxon test beats the Log-Rank test under the Log-logistic survival distribution. Under the Log-normal distribution, the Wilcoxon test also performs better than the Log-Rank test. This is probably because the hazard function of the Log-normal distribution is very similar to the Log-logistic distribution (Klein and Moeschberger 2003).

All the adaptive tests perform as well as the efficient Log-Rank test under the Weibull distribution. Under the Log-logistic and Log-normal distributions, the adaptive tests outperform Log-Rank test in most cases. Among the three adaptive tests, DL (Default Log-Rank) and EQ (Equal Discrimination) are preferable to DW (Default Wilcoxon) given their superior power in most cases. The competitive performance displayed by the adaptive tests under the Log-normal distribution indicates that they are also sensitive to survival differences even when the underlying survival distribution is other than the Weibull and Log-logistic distributions.

Chapter 5

Conclusion

In this paper, we have formally derived and proven the asymptotic normality of the discrimination test statistic originally proposed by Prentice (1975). We have prepared a scheme for extending the discrimination test in the two-sample survival data framework and introduced an adaptive procedure based on the Log-Rank and Peto-Peto's Wilcoxon tests. We have investigated optimal choices of a critical value for the discrimination test statistic to maintain the size and maximize the power of the adaptive procedure. And lastly, we have compared the finite sample performance of the adaptive test with the non-adaptive tests (Log-Rank and Peto-Peto's Wilcoxon tests) through simulation.

The simulation results have shown that the adaptive two-sample test is more robust to the underlying survival distribution, relative to the Log-Rank and Wilcoxon tests. The adaptive test may not have exhibited the highest power across the various survival and censoring distributions, but it performed better than the less efficient test between the two non-adaptive procedures in most cases of the simulation tests. It has even competed well with the most efficient Log-Rank test under the Weibull survival distribution. Moreover, it has demonstrated validity by maintaining the Type I error rate. These results confirm the fact that it would be useful to do a pretest prior to performing the Log-Rank and Wilcoxon tests like what we have proposed in this paper.

The adaptive test can be extended to the K-sample problem by modifying the procedure's step which standardizes the failure times from the two samples. Such K-sample adaptive test can then be further developed into a stratified test to account for covariates.

This paper only investigated the performance of the adaptive test on three well-known survival distributions: Weibull, Log-logistic and Log-normal. Their performance on other survival distributions may also be evaluated in subsequent research. One may also explore further Prentice's model and come up with a more generalized preliminary test, which discriminates one distribution from the rest of the distributions embedded in the model.

REFERENCES

Abramowitz, M., Stegun, I. A. (1964). Gamma Function and Related Functions. Handbook of Mathematical Functions, 255-265.

Cox, D. R. (1972). Regression Models and Life-Tables (with discussion). J. R. Statist. Soc. B 34, 187-220.

Gehan, E. (1965). A generalized Wilcoxon test for comparing arbitrarily singly censored samples. Biometrika 52, 203-223.

Hogg, R. V., Craig, A. T. (1995). Introduction to Mathematical Statistics, 5th Ed., 265.

Hogg, R. V., Fisher, D. M., Randles, R. H. (1975). A Two-Sample Adaptive Distribution-Free Test. Journal of the American Statistical Association 70, 656-661.

Klein, J. P., Moeschberger, M. L. (2003). Survival Analysis: Techniques for Censored and Truncated Data, 36-44, 393-408.

Lehmann, E. L., Casella, G. (1998). Theory of Point Estimation, 2nd Ed., 461-468.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemotheraphy Rep. 50, 163-170.

Peto, R., Peto, J. (1972). Asymptotically Efficient Rank Invariant Test Procedures (with discussion). J. R. Statist. Soc. A 135, 185-206.

Prentice, R. L. (1975). Discrimination among some parametric models. Biometrika 62, 607-614.

Prentice, R. L. (1978). Linear rank tests with right censored data. Biometrika 65, 167-179.

Prentice, R. L., Marek, P. (1979). A Qualitative Discrepancy between Censored Data Rank Tests. Biometrics 35, 861-867.

Reid, N. (1994) A conversation with Sir David Cox. Statist. Sci., 9, 439-455.

Savage, I. R. (1956). Contributions to the theory of rank order statistics-the two-sample case. Ann. Math. Statist. 27, 590-615.

Sorensen, D., Gianola, D. (2002). Likelihood, Bayesian and MCMC Methods in Quantitative Genetics, 153-160.

APPENDIX

Derivation of Information Matrix $\mathbf{I}(\mu, \sigma, \mathbf{r_2} = \mathbf{0})$ Along $\mathbf{r_1} = \mathbf{1}$ (See Section 3.1.2)

Consider the location-scale model, $y = \mu + \sigma\epsilon$, such that

$$f(\epsilon) = \frac{1}{B(m_1, m_2)} \left(\frac{m_1}{m_2}\right)^{m_1} e^{\epsilon\, m_1} \left(1 + \frac{m_1 e^{\epsilon}}{m_2}\right)^{-(m_1 + m_2)}$$

where $m_1 > 0$, $m_2 > 0$ and B is the beta function.

Let $r_1 = \dfrac{1}{m_1}$ and $r_2 = \dfrac{1}{m_2}$. Then by definition of the information matrix, $I(\mu, \sigma, r_2 = 0)$ at $r_1 = 1$ is expressed as

$$I(\mu, \sigma, r_2 = 0)\Big|_{r_1=1} = \begin{bmatrix} cov\left(\frac{\partial\ell}{\partial\mu}, \frac{\partial\ell}{\partial\mu}\right) & cov\left(\frac{\partial\ell}{\partial\mu}, \frac{\partial\ell}{\partial\sigma}\right) & cov\left(\frac{\partial\ell}{\partial\mu}, \frac{\partial\ell}{\partial r_2}\right) \\ & cov\left(\frac{\partial\ell}{\partial\sigma}, \frac{\partial\ell}{\partial\sigma}\right) & cov\left(\frac{\partial\ell}{\partial\sigma}, \frac{\partial\ell}{\partial r_2}\right) \\ & & cov\left(\frac{\partial\ell}{\partial r_2}, \frac{\partial\ell}{\partial r_2}\right) \end{bmatrix},$$

where $\partial\ell/\partial\mu$, $\partial\ell/\partial\sigma$ and $\partial\ell/\partial r_2$ are partial derivatives of the log-likelihood function at $r_1 = 1$, with respect to μ, σ and r_2, respectively, evaluated at $r_2 = 0$.

As we will show later, the partial derivative $\partial\ell/\partial r_2$ at ($r_1 = 1$, $r_2 = 0$) can be directly obtained from Prentice's result (R3a) of Section 2.3. On the other hand, $\partial\ell/\partial\mu$ and $\partial\ell/\partial\sigma$ can be derived from the log-likelihood function evaluated at $(r_1 = 1,\, r_2 = 0)$.

1. Log-likelihood Function at $(\mathbf{r_1 = 1, r_2 = 0})$

From Prentice's result (R1) of Section 2.3, as $m_2 \to \infty$ or $r_2 = \dfrac{1}{m_2} \to 0$,

$$f(\epsilon) = \frac{{m_1}^{m_1}\, e^{(\epsilon m_1 - e^{\epsilon} m_1)}}{\Gamma(m_1)}.$$

Further, at $m_1 = 1$ or $r_1 = \dfrac{1}{m_1} = 1$,

$$f(\epsilon) = e^{\epsilon - e^{\epsilon}}$$

so that the density of y at $(m_1 = 1,\, m_2 = \infty)$ or $(r_1 = 1,\, r_2 = 0)$ is given by

$$f(y) = \frac{1}{\sigma}\, e^{\left[\frac{y-\mu}{\sigma} - e^{\left(\frac{y-\mu}{\sigma}\right)}\right]}$$

and the log-likelihood function is

$$\ell(y; \mu, \sigma) = \frac{y-\mu}{\sigma} - e^{\left(\frac{y-\mu}{\sigma}\right)} - \ln \sigma.$$

2. First-Order Partial Derivatives

$$a.\quad \frac{\partial \ell}{\partial \mu} = -\frac{1}{\sigma} - e^{\left(\frac{y-\mu}{\sigma}\right)}\left(-\frac{1}{\sigma}\right)$$

$$= -\frac{1}{\sigma} + \left(\frac{1}{\sigma}\right)e^{\left(\frac{y-\mu}{\sigma}\right)}$$

$$= \frac{1}{\sigma}\left[e^{\left(\frac{y-\mu}{\sigma}\right)} - 1\right]$$

$$b.\quad \frac{\partial \ell}{\partial \sigma} = -\frac{y-\mu}{\sigma^2} - e^{\left(\frac{y-\mu}{\sigma}\right)}\left(-\frac{y-\mu}{\sigma^2}\right) - \frac{1}{\sigma}$$

$$= -\frac{y-\mu}{\sigma^2} + \left(\frac{y-\mu}{\sigma^2}\right)e^{\left(\frac{y-\mu}{\sigma}\right)} - \frac{1}{\sigma}$$

$$= -\frac{1}{\sigma}\left[\frac{y-\mu}{\sigma} - \left(\frac{y-\mu}{\sigma}\right)e^{\left(\frac{y-\mu}{\sigma}\right)} + 1\right]$$

From Prentice's result (R3a),

$$c.\quad \lim_{r_2 \to 0} \frac{\partial \ell}{\partial r_2} = \lim_{m_2 \to \infty} \left(-\, m_2{}^2 \, \frac{\partial \ell}{\partial m_2} \right)$$

$$= \frac{1}{2} m_1{}^2 \, e^{\left[2\left(\frac{y-\mu}{\sigma}\right)\right]} - m_1{}^2 \, e^{\left(\frac{y-\mu}{\sigma}\right)} + \frac{1}{2} m_1{}^2 - \frac{1}{2} m_1$$

At $m_1 = 1$ or $r_1 = \dfrac{1}{m_1} = 1$,

$$\lim_{r_2 \to 0} \left. \frac{\partial \ell}{\partial r_2} \right|_{r_1 = 1} = \frac{1}{2} \, e^{\left[2\left(\frac{y-\mu}{\sigma}\right)\right]} - e^{\left(\frac{y-\mu}{\sigma}\right)}$$

3. Expected Values of Partial Derivatives

$$a.\quad E\left[\frac{\partial \ell}{\partial \mu}\right] = \int_{-\infty}^{\infty} \frac{1}{\sigma} \left[e^{\left(\frac{y-\mu}{\sigma}\right)} - 1 \right] \frac{1}{\sigma} e^{\left[\frac{y-\mu}{\sigma} - e^{\left(\frac{y-\mu}{\sigma}\right)}\right]} \, dy$$

$$= \frac{1}{\sigma} \int_{0}^{\infty} (w-1)\, e^{-w} \, dw \,, \quad \text{where } w = e^{\left(\frac{y-\mu}{\sigma}\right)}$$

$$= \frac{1}{\sigma} \left[\int_{0}^{\infty} w e^{-w} \, dw - \int_{0}^{\infty} e^{-w} \, dw \right]$$

$$\text{Let} \quad u = w \qquad \text{and} \quad dv = e^{-w}\, dw$$

$$du = dw \quad \text{and} \quad v = -e^{-w}$$

Thus, through integration by parts,

$$E\left[\frac{\partial \ell}{\partial \mu}\right] = \frac{1}{\sigma}\left\{ -we^{-w}\Big]_0^{\infty} + \int_0^{\infty} e^{-w}\, dw - \int_0^{\infty} e^{-w}\, dw \right\}$$

$$= -\frac{1}{\sigma}\, we^{-w}\Big]_0^{\infty}$$

$$= 0$$

$$b.\ E\left[\frac{\partial \ell}{\partial \sigma}\right] = \int_{-\infty}^{\infty} -\frac{1}{\sigma}\left[\frac{y-\mu}{\sigma} - \left(\frac{y-\mu}{\sigma}\right) e^{\left(\frac{y-\mu}{\sigma}\right)} + 1\right] \frac{1}{\sigma} e^{\left[\frac{y-\mu}{\sigma} - e^{\left(\frac{y-\mu}{\sigma}\right)}\right]} dy$$

$$= -\frac{1}{\sigma}\left\{ \int_{-\infty}^{\infty} \left(\frac{y-\mu}{\sigma}\right) \frac{1}{\sigma}\, e^{\left(\frac{y-\mu}{\sigma}\right)}\, e^{-\left[e^{\left(\frac{y-\mu}{\sigma}\right)}\right]} dy \right.$$

$$- \int_{-\infty}^{\infty} \left(\frac{y-\mu}{\sigma}\right) e^{\left(\frac{y-\mu}{\sigma}\right)} \frac{1}{\sigma}\, e^{\left(\frac{y-\mu}{\sigma}\right)}\, e^{-\left[e^{\left(\frac{y-\mu}{\sigma}\right)}\right]} dy$$

$$+\int_{-\infty}^{\infty} \frac{1}{\sigma} e^{\left(\frac{y-\mu}{\sigma}\right)} e^{-\left[e^{\left(\frac{y-\mu}{\sigma}\right)}\right]} dy \Bigg\}$$

$$= -\frac{1}{\sigma}\left[\int_{0}^{\infty} (\ln w) e^{-w}\, dw - \int_{0}^{\infty} (\ln w) w e^{-w}\, dw + \int_{0}^{\infty} e^{-w}\, dw\right],$$

where $w = e^{\left(\frac{y-\mu}{\sigma}\right)}$

2^{nd} Term: $-\int_{0}^{\infty} (\ln w) w e^{-w}\, dw$

Let $u = w(\ln w)$ and $dv = e^{-w}\, dw$

$du = (\ln w + 1)\, dw$ and $v = -e^{-w}$

Through integration by parts,

$$-\int_{0}^{\infty} (\ln w) w e^{-w}\, dw = -\left\{ -w(\ln w) e^{-w}\Big]_{0}^{\infty} + \int_{0}^{\infty} e^{-w}(\ln w + 1)\, dw \right\}$$

$$= w(\ln w) e^{-w}\Big]_{0}^{\infty} - \int_{0}^{\infty} (\ln w) e^{-w}\, dw - \int_{0}^{\infty} e^{-w}\, dw$$

Hence,

$$E\left[\frac{\partial \ell}{\partial \sigma}\right] = -\frac{1}{\sigma}\left\{\int_0^\infty (\ln w)e^{-w}\,dw + w(\ln w)e^{-w}\Big]_0^\infty - \int_0^\infty (\ln w)e^{-w}\,dw\right.$$

$$\left. - \int_0^\infty e^{-w}\,dw + \int_0^\infty e^{-w}\,dw\right\}$$

$$= -\frac{1}{\sigma}\Big[w(\ln w)e^{-w}\Big]_0^\infty$$

$$= 0$$

c. $$E\left[\frac{\partial \ell}{\partial r_2}\right] = \int_{-\infty}^{\infty}\left[\frac{1}{2}\,e^{\left[2\left(\frac{y-\mu}{\sigma}\right)\right]} - e^{\left(\frac{y-\mu}{\sigma}\right)}\right]\frac{1}{\sigma}\,e^{\left[\frac{y-\mu}{\sigma} - e^{\left(\frac{y-\mu}{\sigma}\right)}\right]}\,dy$$

$$= \int_{-\infty}^{\infty}\frac{1}{2}\,e^{\left[2\left(\frac{y-\mu}{\sigma}\right)\right]}\frac{1}{\sigma}\,e^{\left[\frac{y-\mu}{\sigma} - e^{\left(\frac{y-\mu}{\sigma}\right)}\right]}\,dy$$

$$- \int_{-\infty}^{\infty} e^{\left(\frac{y-\mu}{\sigma}\right)}\frac{1}{\sigma}\,e^{\left[\frac{y-\mu}{\sigma} - e^{\left(\frac{y-\mu}{\sigma}\right)}\right]}\,dy$$

$$= \int_0^\infty \frac{1}{2}\,w^2 e^{-w}\,dw - \int_0^\infty w e^{-w}\,dw\ , \quad \text{where } w = e^{\left(\frac{y-\mu}{\sigma}\right)}$$

$$\text{Let} \quad u = w^2 \qquad \text{and} \quad dv = e^{-w}\, dw$$

$$du = 2w\, dw \quad \text{and} \quad v = -e^{-w}$$

Hence, through integration by parts,

$$E\left[\frac{\partial \ell}{\partial r_2}\right] = \frac{1}{2}\left\{ -w^2 e^{-w}\Big]_0^\infty + \int_0^\infty 2\, we^{-w}\, dw \right\} - \int_0^\infty we^{-w}\, dw$$

$$= -\frac{1}{2}\, w^2 e^{-w}\Big]_0^\infty$$

$$= 0$$

d. $$E\left[\left(\frac{\partial \ell}{\partial \mu}\right)^2\right] = \int_{-\infty}^{\infty} \frac{1}{\sigma^2} \left[e^{\left(\frac{y-\mu}{\sigma}\right)} - 1\right]^2 \frac{1}{\sigma}\, e^{\left[\frac{y-\mu}{\sigma} - e^{\left(\frac{y-\mu}{\sigma}\right)}\right]} dy$$

$$= \frac{1}{\sigma^2} \int_0^\infty (w-1)^2\, e^{-w}\, dw\,, \qquad \text{where } w = e^{\left(\frac{y-\mu}{\sigma}\right)}$$

$$= \frac{1}{\sigma^2}\left[\int_0^\infty w^2 e^{-w}\, dw - 2\int_0^\infty we^{-w}\, dw + \int_0^\infty e^{-w}\, dw\right]$$

1^{st} term: $\int_0^\infty w^2 \, e^{-w} \, dw$

$$\text{Let} \quad u = w^2 \qquad \text{and} \quad dv = e^{-w} \, dw$$

$$du = 2w \, dw \quad \text{and} \quad v = -e^{-w}$$

Through integration by parts,

$$\int_0^\infty w^2 \, e^{-w} \, dw = -w^2 \, e^{-w} \Big]_0^\infty + \int_0^\infty e^{-w} \, 2w \, dw.$$

Accordingly,

$$E\left[\left(\frac{\partial \ell}{\partial \mu}\right)^2\right] = \frac{1}{\sigma^2}\left\{ -w^2 e^{-w} \Big]_0^\infty + \int_0^\infty 2we^{-w} \, dw \right.$$

$$\left. -2\int_0^\infty we^{-w} \, dw - e^{-w} \Big]_0^\infty \right\}$$

$$= \frac{1}{\sigma^2}$$

$$e.\ E\left[\left(\frac{\partial \ell}{\partial \sigma}\right)^2\right] = \int_{-\infty}^{\infty} \frac{1}{\sigma^2}\left[\frac{y-\mu}{\sigma} - \left(\frac{y-\mu}{\sigma}\right) e^{\left(\frac{y-\mu}{\sigma}\right)} + 1\right]^2$$

$$\times \frac{1}{\sigma}\, e^{\left[\frac{y-\mu}{\sigma} - e^{\left(\frac{y-\mu}{\sigma}\right)}\right]}\, dy$$

$$= \frac{1}{\sigma^2}\int_0^{\infty} [\ln w - (\ln w)w + 1]^2\ e^{-w}\, dw\ , \quad \text{where } w = e^{\left(\frac{y-\mu}{\sigma}\right)}$$

$$= \frac{1}{\sigma^2}\int_0^{\infty} [(\ln w)(1-w) + 1]^2\ e^{-w}\, dw$$

$$= \frac{1}{\sigma^2}\left[\int_0^{\infty} (\ln w)^2 (1-w)^2 e^{-w}\, dw\right.$$

$$\left. + 2\int_0^{\infty} (\ln w)(1-w)e^{-w}\, dw + \int_0^{\infty} e^{-w}\, dw\right]$$

$$= \frac{1}{\sigma^2}\left[\int_0^{\infty} (\ln w)^2 (1 - 2w + w^2) e^{-w}\, dw + 2\int_0^{\infty} (\ln w) e^{-w}\, dw\right.$$

$$\left. - 2\int_0^{\infty} w(\ln w)e^{-w}\, dw + \int_0^{\infty} e^{-w}\, dw\right]$$

$$= \frac{1}{\sigma^2}\left[\int_0^\infty (\ln w)^2 e^{-w}\, dw - 2\int_0^\infty (\ln w)^2 w e^{-w}\, dw\right.$$

$$+\int_0^\infty w^2(\ln w)^2 e^{-w}\, dw + 2\int_0^\infty (\ln w) e^{-w}\, dw$$

$$\left. -2\int_0^\infty w(\ln w) e^{-w}\, dw + \int_0^\infty e^{-w} dw\right]$$

To evaluate the above integral, the following equations will be used:

(1) $\int_0^\infty (\ln w)^2 e^{-w}\, dw = \frac{\pi^2}{6} + \gamma^2$, where γ = Euler's constant

(2) $\int_0^\infty (\ln w) e^{-w}\, dw = -\gamma$

Let us now simplify some of the terms in the integral.

1^{st} Term: $\int_0^\infty (\ln w)^2 e^{-w}\, dw = \frac{\pi^2}{6} + \gamma^2$

3^{rd} Term: $\int_0^\infty w^2(\ln w)^2 e^{-w}\, dw$

$$\text{Let} \quad u = w^2(\ln w)^2 \qquad \text{and} \quad dv = e^{-w}\, dw$$

$$du = [2w(\ln w)^2 + 2w(\ln w)]\; dw \quad \text{and} \quad v = -e^{-w}$$

Through integration by parts,

$$\int_0^\infty w^2(\ln w)^2 e^{-w}\, dw = -w^2(\ln w)^2 e^{-w}\Big]_0^\infty + \int_0^\infty 2w(\ln w)^2 e^{-w}\, dw$$

$$+ \int_0^\infty 2w(\ln w) e^{-w}\, dw$$

4^{th} Term: $2\int_0^\infty (\ln w) e^{-w}\, dw = -2\gamma$

Thus,

$$E\left[\left(\frac{\partial \ell}{\partial \sigma}\right)^2\right] = \frac{1}{\sigma^2}\left\{\frac{\pi^2}{6} + \gamma^2 - 2\int_0^\infty (\ln w)^2 w e^{-w}\, dw - w^2(\ln w)^2 e^{-w}\Big]_0^\infty\right.$$

$$+ \int_0^\infty 2w(\ln w)^2 e^{-w}\, dw + \int_0^\infty 2w(\ln w) e^{-w}\, dw - 2\gamma$$

$$-2\int_0^\infty w(\ln w)e^{-w}\,dw - e^{-w}\Big]_0^\infty \Bigg\}$$

$$= \frac{1}{\sigma^2}\left(\frac{\pi^2}{6} + \gamma^2 - 2\gamma + 1\right)$$

$$= \frac{1}{\sigma^2}\left[\frac{\pi^2}{6} + (1-\gamma)^2\right]$$

$$f.\ E\left[\left(\frac{\partial \ell}{\partial r_2}\right)^2\right] = \int_{-\infty}^{\infty}\left\{\frac{1}{2}e^{\left[2\left(\frac{y-\mu}{\sigma}\right)\right]} - e^{\left(\frac{y-\mu}{\sigma}\right)}\right\}^2 \frac{1}{\sigma}e^{\left[\frac{y-\mu}{\sigma} - e^{\left(\frac{y-\mu}{\sigma}\right)}\right]}\,dy$$

$$= \int_{-\infty}^{\infty}\left\{\frac{1}{4}e^{\left[4\left(\frac{y-\mu}{\sigma}\right)\right]} - e^{\left[3\left(\frac{y-\mu}{\sigma}\right)\right]} + e^{\left[2\left(\frac{y-\mu}{\sigma}\right)\right]}\right\}$$

$$\times \frac{1}{\sigma}\, e^{\left(\frac{y-\mu}{\sigma}\right)} e^{-\left[e^{\left(\frac{y-\mu}{\sigma}\right)}\right]}\,dy$$

$$= \int_0^\infty \frac{1}{4}w^4 e^{-w}\,dw - \int_0^\infty w^3 e^{-w}\,dw + \int_0^\infty w^2 e^{-w}\,dw$$

where $w = e^{\left(\frac{y-\mu}{\sigma}\right)}$

1^{st} Term: $\frac{1}{4}\int_0^\infty w^4 e^{-w}\,dw$

$$\text{Let} \quad u = w^4 \qquad \text{and} \quad dv = e^{-w}\,dw$$

$$du = 4w^3\,dw \quad \text{and} \quad v = -e^{-w}$$

Through integration by parts,

$$\frac{1}{4}\int_0^\infty w^4 e^{-w}\,dw = \frac{1}{4}\left\{ -w^4 e^{-w}\Big]_0^\infty + \int_0^\infty 4w^3 e^{-w}\,dw \right\}$$

3^{rd} Term: $\int_0^\infty w^2 e^{-w}\,dw$

Through repeated integration by parts,

$$\int_0^\infty w^2 e^{-w}\,dw = -w^2 e^{-w}\Big]_0^\infty + 2\int_0^\infty w e^{-w}\,dw$$

$$= \Big[- w^2 e^{-w} - 2we^{-w} \Big]_0^\infty + 2\int_0^\infty e^{-w}\,dw$$

$$= \Big[- w^2 e^{-w} - 2we^{-w} - 2e^{-w} \Big]_0^\infty$$

Therefore,

$$E\left[\left(\frac{\partial \ell}{\partial r_2}\right)^2\right] = \frac{1}{4}\Big[-w^4e^{-w}\Big]_0^\infty + \int_0^\infty w^3e^{-w}\,dw$$

$$-\int_0^\infty w^3e^{-w}\,dw - \Big[w^2e^{-w} + 2we^{-w} + 2e^{-w}\Big]_0^\infty$$

$$= -e^{-w}\left(\frac{1}{4}w^4 + w^2 + 2w + 2\right)\Big]_0^\infty$$

$$= 2$$

$$g.\ E\left[\left(\frac{\partial \ell}{\partial \mu}\right)\left(\frac{\partial \ell}{\partial \sigma}\right)\right] = \int_{-\infty}^{\infty} -\frac{1}{\sigma^2}\left[e^{\left(\frac{y-\mu}{\sigma}\right)} - 1\right]\left[\frac{y-\mu}{\sigma} - \left(\frac{y-\mu}{\sigma}\right)e^{\left(\frac{y-\mu}{\sigma}\right)} + 1\right]$$

$$\times\frac{1}{\sigma}\,e^{\left[\frac{y-\mu}{\sigma} - e^{\left(\frac{y-\mu}{\sigma}\right)}\right]}\,dy$$

$$= \int_0^\infty -\frac{1}{\sigma^2}(w-1)\left[\ln w - (\ln w)w + 1\right]\,e^{-w}\,dw\ ,$$

where $w = e^{\left(\frac{y-\mu}{\sigma}\right)}$

$$= -\frac{1}{\sigma^2}\left\{\int_0^\infty 2w(\ln w)e^{-w}\,dw - \int_0^\infty w^2(\ln w)e^{-w}\,dw\right.$$

$$\left. + \int_0^\infty we^{-w}\,dw - \int_0^\infty (\ln w)e^{-w}\,dw - \int_0^\infty e^{-w}\,dw\right\}$$

2^{nd} Term: $-\int_0^\infty w^2(\ln w)e^{-w}\,dw$

Let $u = w^2(\ln w)$ and $dv = e^{-w}\,dw$

$du = [2w(\ln w) + w]\,dw$ and $v = -e^{-w}$

Using integration by parts,

$$-\int_0^\infty w^2(\ln w)e^{-w}\,dw = -\left\{-w^2(\ln w)e^{-w}\Big]_0^\infty + \int_0^\infty 2w(\ln w)e^{-w}\,dw\right.$$

$$\left. + \int_0^\infty we^{-w}\,dw\right\}$$

4^{th} Term: $-\int_0^\infty (\ln w)e^{-w}\,dw = \gamma$

Therefore,

$$E\left[\left(\frac{\partial \ell}{\partial \mu}\right)\left(\frac{\partial \ell}{\partial \sigma}\right)\right] = -\frac{1}{\sigma^2}\left\{\int_0^\infty 2w(\ln w)e^{-w}\,dw + w^2(\ln w)e^{-w}\Big]_0^\infty\right.$$

$$-\int_0^\infty 2w(\ln w)e^{-w}\,dw - \int_0^\infty we^{-w}\,dw$$

$$\left.+\int_0^\infty we^{-w}\,dw + \gamma + e^{-w}\Big]_0^\infty\right\}$$

$$= -\frac{1}{\sigma^2}(\gamma - 1)$$

$$= \frac{1-\gamma}{\sigma^2}$$

$$h.\ E\left[\left(\frac{\partial \ell}{\partial \mu}\right)\left(\frac{\partial \ell}{\partial r_2}\right)\right] = \int_{-\infty}^{\infty}\frac{1}{\sigma}\left[e^{\left(\frac{y-\mu}{\sigma}\right)} - 1\right]\left\{\frac{1}{2}e^{\left[2\left(\frac{y-\mu}{\sigma}\right)\right]} - e^{\left(\frac{y-\mu}{\sigma}\right)}\right\}$$

$$\times \frac{1}{\sigma}e^{\left[\frac{y-\mu}{\sigma} - e^{\left(\frac{y-\mu}{\sigma}\right)}\right]}\,dy$$

$$= \frac{1}{\sigma}\int_0^\infty (w-1)\left(\frac{1}{2}w^2 - w\right)e^{-w}\,dw\ , \text{ where } w = e^{\left(\frac{y-\mu}{\sigma}\right)}$$

$$= \frac{1}{\sigma}\left[\int_0^\infty \frac{1}{2}w^3e^{-w}\,dw - \int_0^\infty \frac{3}{2}w^2e^{-w}\,dw + \int_0^\infty we^{-w}\,dw\right]$$

1^{st} Term: $\int_0^\infty \frac{1}{2}w^3e^{-w}\,dw$

Using integration by parts,

$$\begin{aligned}\int_0^\infty \frac{1}{2}w^3e^{-w}\,dw &= -\frac{1}{2}w^3e^{-w}\Big]_0^\infty + \int_0^\infty \frac{3}{2}w^2e^{-w}\,dw \\ &= \int_0^\infty \frac{3}{2}w^2e^{-w}\,dw\end{aligned}$$

3^{rd} Term: $\int_0^\infty we^{-w}\,dw$

Again, through integration by parts,

$$\begin{aligned}\int_0^\infty we^{-w}\,dw &= -we^{-w}\Big]_0^\infty + \int_0^\infty e^{-w}\,dw \\ &= -e^{-w}\Big]_0^\infty \\ &= 1\end{aligned}$$

Therefore,

$$E\left[\left(\frac{\partial \ell}{\partial \mu}\right)\left(\frac{\partial \ell}{\partial r_2}\right)\right] = \frac{1}{\sigma}\left(\int_0^\infty \frac{3}{2}w^2 e^{-w}\,dw - \int_0^\infty \frac{3}{2}w^2 e^{-w}\,dw + 1\right)$$

$$= \frac{1}{\sigma}$$

$$i.\ E\left[\left(\frac{\partial \ell}{\partial \sigma}\right)\left(\frac{\partial \ell}{\partial r_2}\right)\right] = \int_{-\infty}^{\infty} -\frac{1}{\sigma}\left[\frac{y-\mu}{\sigma} - \left(\frac{y-\mu}{\sigma}\right)e^{\left(\frac{y-\mu}{\sigma}\right)} + 1\right]$$

$$\times \left\{\frac{1}{2}e^{\left[2\left(\frac{y-\mu}{\sigma}\right)\right]} - e^{\left(\frac{y-\mu}{\sigma}\right)}\right\}\frac{1}{\sigma}e^{\left[\frac{y-\mu}{\sigma} - e^{\left(\frac{y-\mu}{\sigma}\right)}\right]}\,dy$$

$$= -\frac{1}{\sigma}\int_0^\infty [\ln w - (\ln w)w + 1]\left(\frac{1}{2}w^2 - w\right)e^{-w}\,dw\ ,$$

where $w = e^{\left(\frac{y-\mu}{\sigma}\right)}$

$$= -\frac{1}{\sigma}\left\{\int_0^\infty \frac{3}{2}(\ln w)w^2 e^{-w}\,dw - \int_0^\infty (\ln w)we^{-w}\,dw\right.$$

$$\left. -\int_0^\infty \frac{1}{2}(\ln w)w^3 e^{-w}\,dw + \int_0^\infty \frac{1}{2}w^2 e^{-w}\,dw - \int_0^\infty we^{-w}\,dw\right\}$$

2^{nd} Term: $-\int_0^\infty (\ln w) w e^{-w}\, dw$

Let $u = (\ln w) w$ and $dv = e^{-w}\, dw$

$du = (\ln w + 1)\, dw$ and $v = -e^{-w}$

Through integration by parts,

$$-\int_0^\infty (\ln w) w e^{-w}\, dw = -\left\{ -(\ln w) w e^{-w} \Big]_0^\infty + \int_0^\infty (\ln w) e^{-w}\, dw + \int_0^\infty e^{-w}\, dw \right\}$$

$$= \gamma - 1$$

3^{rd} Term: $-\int_0^\infty \frac{1}{2} (\ln w) w^3 e^{-w}\, dw$

Let $u = \frac{1}{2} (\ln w) w^3$ and $dv = e^{-w}\, dw$

$du = \left[\frac{1}{2} w^2 + \frac{3}{2} (\ln w) w^2 \right] dw$ and $v = -e^{-w}$

Through integration by parts,

$$-\int_0^\infty \frac{1}{2}(\ln w)w^3e^{-w}\,dw = -\left\{-\frac{1}{2}(\ln w)w^3e^{-w}\Big]_0^\infty + \int_0^\infty \frac{1}{2}w^2e^{-w}\,dw + \int_0^\infty \frac{3}{2}(\ln w)w^2e^{-w}\,dw\right\}$$

$$= -\int_0^\infty \frac{1}{2}w^2e^{-w}\,dw - \int_0^\infty \frac{3}{2}(\ln w)w^2e^{-w}\,dw$$

5^{th} Term: $-\int_0^\infty we^{-w}\,dw$

Again, using integration by parts,

$$-\int_0^\infty we^{-w}\,dw = -\left\{-we^{-w}\Big]_0^\infty + \int_0^\infty e^{-w}\,dw\right\}$$

$$= e^{-w}\Big]_0^\infty$$

$$= -1$$

Therefore,

$$
\begin{aligned}
E\left[\left(\frac{\partial \ell}{\partial \sigma}\right)\left(\frac{\partial \ell}{\partial r_2}\right)\right] &= -\frac{1}{\sigma}\left\{\int_0^\infty \frac{3}{2}(\ln w)w^2e^{-w}\,dw + \gamma - 1 - \int_0^\infty \frac{1}{2}w^2e^{-w}\,dw\right. \\
&\qquad \left. -\int_0^\infty \frac{3}{2}(\ln w)w^2e^{-w}\,dw + \int_0^\infty \frac{1}{2}w^2e^{-w}\,dw - 1\right\} \\
&= -\frac{1}{\sigma}(\gamma - 2) \\
&= \frac{2-\gamma}{\sigma}
\end{aligned}
$$

4. Information Matrix

Since $E\left[\frac{\partial \ell}{\partial \mu}\right]$, $E\left[\frac{\partial \ell}{\partial \sigma}\right]$, $E\left[\frac{\partial \ell}{\partial r_2}\right]$ are all equal to zero as shown in the previous section, the information matrix $I(\mu, \sigma, r_2 = 0)$ at $r_1 = 1$ is reduced to

$$I(\mu,\sigma,r_2=0)\Big|_{r_1=1} = \begin{bmatrix} E\left[\left(\frac{\partial \ell}{\partial \mu}\right)^2\right] & E\left[\left(\frac{\partial \ell}{\partial \mu}\right)\left(\frac{\partial \ell}{\partial \sigma}\right)\right] & E\left[\left(\frac{\partial \ell}{\partial \mu}\right)\left(\frac{\partial \ell}{\partial r_2}\right)\right] \\ & E\left[\left(\frac{\partial \ell}{\partial \sigma}\right)^2\right] & E\left[\left(\frac{\partial \ell}{\partial \sigma}\right)\left(\frac{\partial \ell}{\partial r_2}\right)\right] \\ & & E\left[\left(\frac{\partial \ell}{\partial r_2}\right)^2\right] \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sigma^2} & \frac{1-\gamma}{\sigma^2} & \frac{1}{\sigma} \\ & \frac{\frac{\pi^2}{6}+(1-\gamma)^2}{\sigma^2} & \frac{2-\gamma}{\sigma} \\ & & 2 \end{bmatrix}.$$