Statistical Methods
for
Weibull Based Random Effects Models

Chang Qing Zhu, B.Sc., M.Sc.

Submitted for the degree of Doctor of Philosophy
at
Department of Mathematics & Statistics,
The University of Surrey,
October 1998.
To my mother
and in memory of my father
Abstract

The Weibull distribution is widely used for the parametric analysis of lifetime data. Ordinary application of the Weibull model implicitly assumes that all the individuals have the same risk of failure while observed risk factors are included into a regression model as covariates. However, in reality, there may exist unobserved risk factors, which cause heterogeneity between individuals. Ignoring the existence of heterogeneity will produce incorrect estimation in survival analysis. In the case of multivariate survival analysis, association among the lifetimes of components also needs to be considered. Possible situations are the survival times of members of the same family, or different components produced in the same batch. Therefore, a generalised form of the Weibull distribution, which is able to handle heterogeneity, is of great importance.

A Weibull based random effects model, which is a mixture of Weibull distributions, is studied in this work. Conditional on a random effect, the lifetimes are independent and are Weibull distributed. By assuming the random effect has a gamma, a positive stable or an inverse Gaussian distribution, corresponding Weibull based random effects models are obtained. Following an overall review of survival analysis, the Weibull distribution and frailty models, an intensive discussion of the main features of the three Weibull based random effects models mentioned above are presented. In addition, the dependence structures of the three models are investigated and compared. Many useful association measures are proposed for describing the local and global association between the components. A two stage marginal estimation method based on the positive stable mixture of Weibull model is suggested.

In the application of survival analysis, it is of interest to know if a model based on independent Weibull distribution is adequate or if a Weibull based random
effects model is more suitable. The other major part of this work focuses on the
detection of heterogeneity in Weibull based random effects models. A score test
and a likelihood ratio test based on the positive stable mixture of Weibull model are
proposed. Their corresponding properties are investigated by asymptotic theory
and simulation studies. Some other more straightforward diagnostic approaches
and their features are also studied and compared with the score test and the
likelihood ratio test.

Throughout this thesis, three data sets are used to illustrate the proposed methods.
Acknowledgements

I would like to thank my supervisor Dr Alan Kimber for all his continual support, encouragement, guidance and invaluable advice throughout the course of this study. Without his help, this work would never have been possible.

Thanks go to the staffs in the Department of Mathematics and Statistics, for their assistance and cooperation. It is a great pleasure for me to work in such a friendly and pleasant environment. In particular, I would like to thank Dr Anthony Ledford for his assistance and supporting discussion on many occasions. I am also grateful to previous and current colleagues in our postgraduate room, Richard Steven, Susan Todd, David Gammack, David Schley and Michael Hutchinson, whose friendship, support and entertainment has enriched my life in one way or another during the three years.

I thank my mother for her continuing encouragement and unselfish sacrifice during my study. I am also indebted to Alan Woodcock, for his understanding and emotional support, as well as his patience, which are always handy in my bad and good times.

Financial support was provided by a University of Surrey Research Scholarship and a CVCP Overseas Research Studentship award. I am grateful to both of the sponsors.
Contents

Abstract ii

Acknowledgements iv

Contents v

List of Figures vi

List of Tables vii

1 Introduction 1

1.1 Background ............................................. 1

1.2 Outline of the thesis ..................................... 2

1.3 Review of the Weibull distribution ....................... 3

  1.3.1 Preliminaries on lifetime distributions ................ 3

  1.3.2 The Weibull distribution .............................. 5

  1.3.3 Some generalisations of the Weibull distribution ........ 6

  1.3.4 Models Based on the Weibull Distribution .............. 7

  1.3.5 Estimation of the Weibull parameters .................. 9

  1.3.6 Assessing the goodness of fit of the Weibull distribution . 12
1.4 Frailty Models ................................................. 13
  1.4.1 Univariate frailty models .............................. 14
  1.4.2 Multivariate frailty models .............................. 15
1.5 Examples ...................................................... 17
  1.5.1 Infant nutrition data ................................. 17
  1.5.2 Repeated endurance exercise tests data ................ 19
  1.5.3 Fibre failure strength data ........................... 20

2 Weibull based random effects models .......................... 24
  2.1 Introduction .............................................. 24
  2.2 Gamma mixture of Weibulls distribution ................. 25
    2.2.1 Univariate case ...................................... 26
    2.2.2 Multivariate case .................................... 27
  2.3 Positive stable mixture of Weibulls distribution ........ 31
    2.3.1 Univariate case ...................................... 32
    2.3.2 Multivariate case .................................... 32
  2.4 Inverse Gaussian mixture of Weibulls distribution ........ 38
    2.4.1 Univariate case ...................................... 38
    2.4.2 Multivariate case .................................... 39

3 The dependence structure of the models ......................... 44
  3.1 An expression in terms of the Archimedean distribution .... 45
  3.2 Measures of association ................................... 45
  3.3 Global association ...................................... 46
3.3.1 Correlation coefficients .................................................. 46
3.3.2 Kendall’s coefficient of concordance ..................................... 49

3.4 Local association ................................................................. 50
3.4.1 Contour plots ................................................................. 50
3.4.2 Cross ratios ................................................................. 51
3.4.3 Time-dependent correlation coefficient .................................... 55
3.4.4 Conditional expected residual life ........................................ 59
3.4.5 Conditional probability ....................................................... 61

3.5 Summary ................................................................. 62

4 Estimation methods for the PSW model ........................................ 64

4.1 Introduction ................................................................. 64
4.2 Maximum likelihood estimator and its asymptotic variance .............. 65
4.3 Two stage marginal estimator and its asymptotic properties .............. 70
  4.3.1 Two stage marginal estimation method .................................. 70
  4.3.2 Some notation and details relevant to the model ....................... 71
  4.3.3 Asymptotic variance ....................................................... 73
4.4 Simulation ................................................................. 75
  4.4.1 Bivariate models ......................................................... 77
  4.4.2 Tri-variate and four-variate models .................................... 81
4.5 Examples ................................................................. 85
  4.5.1 Infant nutrition data ...................................................... 85
  4.5.2 Fibre failure strength data ............................................... 85
6.2 The likelihood ratio test and its asymptotic properties .......... 126
6.3 Simulation study ...................................... 129
   6.3.1 Evaluation of the rate of convergence ................. 129
   6.3.2 Estimation of critical values ......................... 131
   6.3.3 Estimation of power functions ......................... 135
6.4 Extension .............................................. 138
   6.4.1 Multivariate case .................................. 138
   6.4.2 Case with covariates ................................ 139
6.5 Comparison with the Score test ................................ 139
6.6 Examples ................................................. 141
   6.6.1 Infant nutrition data .................................. 141
   6.6.2 Repeated endurance exercise tests data ................ 141
   6.6.3 Fibre failure strength data ............................ 142

7 Further diagnostic methods for detecting heterogeneity in the PSW model 143
   7.1 A Simple Test for Frailty .............................. 144
      7.1.1 Distribution of statistic S ......................... 144
      7.1.2 The case with no nuisance parameters .............. 145
      7.1.3 The case with nuisance parameters .................. 146
      7.1.4 The effect of censoring .............................. 147
   7.2 A diagnostic based on component-wise minima ............... 148
   7.3 A concordance test ..................................... 152
      7.3.1 Definition of the statistic U ....................... 152
List of Figures

1.1 Density and hazard functions for the Weibull distributions, where \( \xi = 1, \) and \( \phi \) varies corresponding to 0.5, 1.5, 2.5 and 5.0 ........................ 6

1.2 Age in months of introduction of fish and egg ............................... 18

1.3 Exercise times to angina pectoris on 3 occasions ............................ 21

1.4 Fibre failure strengths for fibre section of different lengths ................ 23

2.1 Density and hazard functions for the Burr distribution, where \( \nu = 2, \)
\( \xi = 1, \) and \( \phi \) varies corresponding to 0.5, 1.5, 2.5 and 5.0 ................ 26

2.2 Density and hazard functions for univariate PSW distribution, where
\( \nu = 0.8, \xi = 1, \) and \( \phi \) varies corresponding to 0.5, 1.5, 2.5 and 5.0 . 33

2.3 Density and hazard functions for the univariate IGW distribution,
where \( \theta = 1.2, \xi = 1, \) and \( \phi \) varies corresponding to 0.5, 1.5, 2.5
and 5.0 .................................. 39

2.4 Density contour plots for three bivariate Weibull based random effects models, with \( \xi = 1 \) and \( \phi = 2 \) ................................. 40

3.1 Correlation coefficient \( \rho(Y_1, Y_2) \) for the three bivariate Weibull based
random effects models, at \( \phi_1 = \phi_2 = 0.5, 1.0, 2.0 \) .......................... 48

3.2 Kendall’s coefficient \( \tau \) vs \( \theta \) for the IGW distribution ............ 51

3.3 Contour plots of cross ratio \( r(y_1, y_2) \) for the bivariate PSW and IGW
distributions, at \( \xi_1 = \xi_2 = \phi_1 = \phi_2 = 1, \) and \( \tau = 0.2 \) ........... 53
3.4 Plot of cross ratio $r(y_1, y_2)$ for the GW, PSW and IGW distributions, at $y_1 = y_2 = y$, $\xi_1 = \xi_2 = \phi_1 = \phi_2 = 1$, and $\tau = 0.2$ ........ 53

3.5 Plot of cross ratio $r(y_1, y_2)$ for the GW, PSW and IGW distributions, at $y_1 = \frac{1}{y_2} = y$, $\xi_1 = \xi_2 = \phi_1 = \phi_2 = 1$, and $\tau = 0.2$ ........ 54

3.6 Correlation of $(N_1(y_1), N_2(y_2))$ for three bivariate Weibull based random effects models, at $\xi_1 = \xi_2 = \phi_1 = \phi_2 = 1$ and $\tau = 0.2$ ........ 56

3.7 Correlation of $(N_1(y_1), N_2(y_2))$ for the IGW distribution, at $\xi_1 = \xi_2 = \phi_1 = \phi_2 = 1$ and $\theta = 10.0, 0.1$ ........ 57

3.8 Correlation curve for three bivariate Weibull based random effects models, at $y_1 = y_2 = y$, $\xi_1 = \xi_2 = \phi_1 = \phi_2 = 1$, and $\tau = 0.2$ ........ 58

3.9 Correlation curve for three bivariate Weibull based random effects models, at $y_1 = \frac{1}{y_2} = y$, $\xi_1 = \xi_2 = \phi_1 = \phi_2 = 1$, and $\tau = 0.2$ ........ 58

3.10 Conditional expected residual life $\psi(y_1, y_2)$ for three bivariate Weibull based random effects models, at $y_1 = y_2 = y$, $\xi_1 = \xi_2 = \phi_1 = \phi_2 = 1$, and $\tau = 0.2$ ........ 60

3.11 Conditional expected residual life $\psi(y_1, y_2)$ for three bivariate Weibull based random effects models, at $y_1 = \frac{1}{y_2} = y$, $\xi_1 = \xi_2 = \phi_1 = \phi_2 = 1$, and $\tau = 0.2$ ........ 60

4.1 Asymptotic variance of parameter estimator of $\nu$ under MLE and marginal method, as $n \to \infty$ ........ 69

4.2 Relative efficiency of estimator $\hat{\nu}$ ........ 74

4.3 Asymptotic variance of parameter estimators of $\alpha_j$ and $\beta_j$ under MLE and the marginal method, as a function of $\nu$, and where the true values of parameters $\alpha_j$ and $\beta_j$ are unity .......... 76

4.4 Relative efficiency of the marginal estimators of $\alpha_j$ and $\beta_j$ ........ 76

5.1 Estimated critical values of un-normalised score statistic $T(n)$ for various sample sizes with superimposed quadratic fit .... 104
5.2 The variation of critical values of the normalised version of \( T(n,c) \) in terms of censoring points ........................................ 105

5.3 The variation of critical values of the normalised version of \( T^*(n,c) \) in terms of censoring points ........................................ 106

5.4 Estimated power of the score statistic \( T(n) \) for different sample sizes at significance level \( P = 0.05 \) ........................................ 107

5.5 Estimated critical values of un-normalised score statistic \( T^*(n) \) for various sample sizes with superimposed quadratic fit in the case with covariate ........................................ 117

6.1 Estimated critical values of the LR test \( \Lambda(n) \) for various samples .......................... 134

6.2 Estimated critical values of the LR test \( \Lambda^*(n) \) for various samples .......................... 134

7.1 Numerical evaluation and fitted lines of integrals \( E[S_j(y_j)S(y_1, y_2)] \) and \( E[S_1(y_1)S_2(y_2)] \) ........................................ 158
## List of Tables

1.1 Age in months of introduction of fish and egg into the diets of 55 infants .................................. 19

1.2 Exercise times to angina (in seconds) on occasions before and after oral isosorbide dinitrate ........................................ 20

1.3 Fibre failure strengths for fibre section of different lengths .......... 22

1.4 Summary results of the fibre failure strengths data ................. 23

4.1 Estimation of $v$ for $n=100$ (no censoring) $p=2$ .................. 79

4.2 Estimation of $\alpha$ for $n=100$ (no censoring) $p=2$ .................. 79

4.3 Estimation of $\beta$ for $n=100$ (no censoring) $p=2$ .................. 79

4.4 Estimation for $n=50$ (no censoring) $p=2$ ........................... 80

4.5 Estimation for samples with 30% censoring ($n=100$) $p=2$ ........ 80

4.6 Estimation for samples with 30% censoring ($n=50$) $p=2$ ........ 80

4.7 Censoring point for various selections of $v$ when $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 1$, $p_c = 30\%$ ........................................ 81

4.8 Estimation for samples with a covariate ($n=100$) $p=2$ .......... 82

4.9 Estimation of $v$ for $n=100$ (no censoring) $p=3$ .................. 83

4.10 Estimation of $\alpha$ for $n=100$ (no censoring) $p=3$ .................. 83

4.11 Estimation of $\beta$ for $n=100$ (no censoring) $p=3$ .................. 83
4.12 Estimation of $\nu$ for $n=100$ (no censoring) $p = 4$ ........... 84
4.13 Estimation of $\alpha$ for $n=100$ (no censoring) $p = 4$ ........... 84
4.14 Estimation of $\beta$ for $n=100$ (no censoring) $p = 4$ ........... 84
4.15 Fit of the bivariate PSW model on the nutrition data ........ 85
4.16 Fit of the PSW model on the fibre strength data: $p = 2$ ........ 87
4.17 Fit of the PSW model on the fibre strength data: $p = 4$ ....... 87
4.18 Fit of the bivariate PSW model on the exercise tests data ....... 87

5.1 Normalised critical values of the score test $T(n)/C_n$ ........... 100
5.2 Normalised critical values of the score test $T^*_n/C_n$ ........... 101
5.3 Normalised critical values of the score test $T(n,c)/C_n$ .......... 102
5.4 Normalised critical values of the score test $T^*_n,c/C_n$ .......... 102
5.5 Estimated coefficients in the model $C_T = c_0 + c_1n + c_2n^2$ .... 103
5.6 Estimated coefficients in the model $C_T = a_0 + \frac{a_1}{c-a_2}$ ....... 106
5.7 Estimated power of score tests $T(n)$ and $T^*_n$ for different sample
sizes and $P = 0.05$ ........................................ 109
5.8 Estimated power of score tests $T(n,c)$ and $T^*_n,c$ for different sample
sizes and $P = 0.05$ ........................................ 110
5.9 Estimated power of score tests $T(n), T^*_n$ and $T^*_n,2$ when data are
from a GW model with gamma shape parameter $\nu$ and $P = 0.05$ ... 112
5.10 Normalised critical values of the score test $T(n)/C_n$ ........... 114
5.11 Estimated coefficients in the model $C_T = c_0 + c_1n + c_2n^2$ .... 118
5.12 Normalised critical values of the score test $T^*_n/C_n$ ........... 118

6.1 Estimated critical values of the likelihood ratio test statistic $\Lambda(n)$ .. 131
6.2 Estimated critical values of the likelihood ratio test statistic $\Lambda_{n}^{*}$ 132
6.3 Estimated critical values of the likelihood ratio test statistic $\Lambda_{n,c}$ 132
6.4 Estimated critical values of likelihood ratio test statistic $\Lambda_{n,c}^{*}$ 132
6.5 Estimated coefficients in the model $C_{A} = c_{0} + c_{1}n + c_{2}n^{2}$ 133
6.6 Estimated power of likelihood ratio tests $\Lambda_{(n)}$ and $\Lambda_{(n)}^{*}$ for different sample sizes and $P = 0.05$ 136
6.7 Estimated power of likelihood ratio tests $\Lambda_{(n),c}$ and $\Lambda_{(n),c}^{*}$ for different sample sizes and $P = 0.05$ 137
6.8 Estimated power of LR $\Lambda_{(n)}$ and $\Lambda_{(n)}^{*}$ when data are from a GW model and $P = 0.05$ 138
6.9 Estimated critical values of likelihood ratio test statistic $\Lambda_{n}$ 140
6.10 Estimated critical values of likelihood ratio test statistic $\Lambda_{n}^{*}$ 140

7.1 Critical values of the test $M$ and $M^{*}$ for $P = 0.05$ 149
7.2 Estimated power(%) of the tests based on $M$ and $M^{*}$ for $P = 0.05$ 150
7.3 Estimated mean and variance of statistic $U$ under $H_{0}$ 162
7.4 Estimated power(%) of the concordance test based on $U$ statistic 162
Chapter 1

Introduction

1.1 Background

The Weibull distribution is probably the most widely used distribution in the parametric modelling of lifetime data, such as the failure times in reliability, material strength data and survival data. The Weibull distribution has considerable flexibility, both in terms of the shape of the density function (positive or negative skewness) and the hazard function (monotonic increasing or decreasing). It has a closed form survivor function, which is convenient for likelihood based inference in the presence of right censoring. On the log-scale the Weibull distribution may be parameterised simply in terms of a location parameter and a scale parameter, which allows for a natural approach to incorporating covariate information. Another merit of the Weibull distribution is that it is both a proportional hazards model and an accelerated life model.

Standard methods of applying the Weibull model assume independence among individuals, while the possible risk factors on survival times may be considered as covariates in the model. However, in some situations the Weibull model might not include all the relevant risk factors, perhaps because we do not have detailed information on each individual, or we may not know that the risk factor is important or even that the factor exists. Those common but unobserved risk factors create dependence among individuals. For example, in industry, fabrication of short fibres from the same parent fibre, or machine components produced in the same batch may be inherently similar. In a medical context, patients might share
some unobserved risk factors which affect their survival times. In epidemiological studies, members of the same family share both genetic and environmental factors. The same possibility is true for animals that have been born in the same litter. The dependence among individuals in a group, which is referred to as frailty or as a random effect, induces variability over and above that anticipated from a Weibull model.

In recent years there has been considerable interest in the study of frailty models. In particular, Weibull based random effects models, which can accommodate dependence and lead to a flexible generalisation of the Weibull distribution, have been considered by many authors, see Vaupel et al (1979), Hougaard (1984, 1986a, 1986b, 1991), Crowder (1985, 1989), Clayton (1978, 1985), Kimber (1990, 1996), Oakes (1982a, 1989a, 1989b), Whitmore & Lee (1991), Shih & Louis (1995a) and Crowder & Kimber (1997). However, compared with well established statistical methods and applications of the Weibull distribution, the methodology and application of the Weibull based random effects models are less well developed. This work aims to investigate the properties of Weibull based random effects models from various aspects, and focuses on the exploration of testing for heterogeneity based on the Weibull based random effects models. Corresponding methods have been applied on some sets of real data from nutrition, medicine and reliability studies.

1.2 Outline of the thesis

In the remaining part of this chapter, some important concepts and results relating to survival analysis and the Weibull distribution are reviewed, and an outline of frailty models is given. In addition, three real data sets which are used for illustration throughout the thesis are introduced. In Chapter 2, we present three tractable Weibull based random effects models which are generated from independent Weibull distributions conditional on a random effect. Some of the features corresponding to the three models are summarised and explored. The dependence structures of the three Weibull based random effects models are investigated and compared in Chapter 3. In Chapter 4, a two stage marginal approach to estimating the positive stable mixture of Weibulls model is introduced, and compared with
the method of maximum likelihood. In Chapters 5 to 7, we focus on the detection of heterogeneity in Weibull based random effects models, particularly for the positive stable mixture of Weibulls model. A score test is derived and investigated in Chapter 5. A likelihood ratio test and its properties are discussed in Chapter 6. Some other diagnostic methods are explored in Chapter 7 and are compared with the score test and the likelihood ratio test. Finally, in Chapter 8, conclusions are given, and some possibilities for future research to extend the current work are suggested.

1.3 Review of the Weibull distribution

The Weibull distribution, see Weibull (1939, 1951), is a natural starting point in the analysis of lifetime data. Throughout this section the main concepts and results for survival analysis and the Weibull distribution are reviewed.

1.3.1 Preliminaries on lifetime distributions

Numerous books are available that present lifetime analysis in various areas. Books such as Mann et al (1974) and Nelson (1982) are oriented more towards engineering applications, while others, such as Lawless (1982) and Cox & Oakes (1984) are more in the medical context. Kalbfleisch & Prentice (1980) is an advanced text with particular emphasis on proportional hazards methods, while Martz & Waller (1982) focused on the Bayesian approaches for survival analysis. The book by Crowder et al (1991) has covered both the probability modelling and the statistical aspects of reliability/survival analysis by dealing with the statistics from a contemporary viewpoint. Ansell & Philips (1994) present some case studies on the reliability data from a practical standpoint.

The time to the occurrence of some event which is of interest is referred to as the survival time. For example, the survival time may be the lifetime of a patient or time until recurrence of some disease of the patient. In another instance, the survival time perhaps may be the time to failure of a component in a system. Commonly, lifetime or survival data includes right censored observations due to the withdrawal of experimental subjects or the termination of the experiment.
For each censored observation, we only know that the subject’s lifetime has exceeded a given value. The exact lifetime remains unknown. Censored observations should not be ignored when analysing survival data, because, among other considerations, the longer-lived subjects are generally more likely to be censored. The statistical methodology must correctly use the censored observations, as well as the uncensored observations.

Usually, a first step in the analysis of survival data is the estimation of the distribution of the failure times. The survivor function is used to describe the lifetimes of the population of interest. Assuming throughout that the survival time $Y$ is always a non-negative continuous variable, the probability that an experimental subject from the population will have a lifetime exceeding $y$ is

$$S(y) = \Pr(Y > y)$$

where $S(y)$ is the **survivor function**. The usual non-parametric estimation method on the survivor function is the product-limit method (also called the Kaplan-Meier method), see Kaplan & Meier (1958) or the life table method, see, for example, Chiang (1968).

Some other functions closely related to the survivor function are the **cumulative distribution function**, $F(y)$; the **density function**, $f(y)$; the **hazard function** (or failure rate function), $h(y)$; and the **cumulative hazard function**, $H(y)$. $F(y)$ is the probability that a lifetime time is less than or equal to $y$, that is

$$F(y) = \Pr(Y \leq y) = 1 - S(y).$$

The density $f(y)$ is defined as

$$f(y) = \frac{dF(y)}{dy} \quad \text{or} \quad f(y) = -\frac{dS(y)}{dy}.$$

The hazard function $h(y)$ specifies the instantaneous rate of failure at time $y$ given that the individual survives up to time $y$, that is

$$h(y) = \lim_{\Delta \to 0} \frac{P_r(y \leq Y < y + \Delta | Y > y)}{\Delta} = f(y)/S(y).$$

The cumulative hazard function $H(y)$ is defined as

$$H(y) = \int_0^y h(u)du.$$
The relationship between $H(y)$ and $S(y)$ is

$$H(y) = - \log S(y).$$

The likelihood function in survival analysis is more complicated. If there are no right censored data, the likelihood function is only related to the density function $f(y)$; otherwise, the likelihood function depends on both the density function and the survivor function. For example, suppose that censoring is random and that the observed survival time sample is $y = (y_1, \ldots, y_n)$. Then the likelihood of the sample is proportional to the product of the event time densities and the survivor functions.

$$L(y) = \prod_{i=1}^{n} f(y_i)^{\delta_i} S(y_i)^{1-\delta_i},$$

where $\delta_i$ is an indicator variable defined by

$$\delta_i = \begin{cases} 0 & \text{if the ith observation is censored.} \\ 1 & \text{if the ith observation is uncensored.} \end{cases}$$ (1.1)

1.3.2 The Weibull distribution

Survival time $Y$ has a Weibull distribution, if, for $y > 0$, the survivor function is

$$S(y) = \exp(-\xi y^\phi),$$ (1.2)

or its hazard function is

$$h(y) = \xi \phi y^{\phi-1},$$

where $\xi$ and $\phi$ are positive parameters, and $\phi$ is called the shape parameter. When $\phi = 1$, $Y$ has an exponential distribution with rate parameter $\xi$. One important feature of the Weibull hazard function is that it has a decreasing failure rate (DFR) for $\phi < 1$, constant failure rate for $\phi = 1$ (exponential) and an increasing failure rate (IFR) for $\phi > 1$. In particular, for $1 < \phi < 2$, the hazard function increases slower than linearly; for $\phi = 2$, the hazard function is linear; and for $\phi > 2$, the hazard increases faster than linearly, see Figure 1.1(b).

The Weibull density is

$$f(y) = \xi \phi y^{\phi-1} \exp(-\xi y^\phi),$$

for $y > 0$, and its cumulative distribution function is

$$F(y) = 1 - \exp(-\xi y^\phi).$$
Some Weibull densities are shown in Figure 1.1(a).

The mean and variance are given by

\[ E(Y) = \xi^{-1/\phi} \Gamma(\phi^{-1} + 1), \]

and

\[ Var(Y) = \xi^{-2/\phi} \{ \Gamma(2\phi^{-1} + 1) - \Gamma^2(\phi^{-1} + 1) \}, \]

where \( \Gamma \) is the gamma function

\[ \Gamma(z) = \int_0^\infty u^{z-1} e^{-u} du. \quad (1.3) \]

When \( \phi \) is large (greater than 5, say), the mean and variance are approximately \( \xi^{-1/\phi} \) and \( 1.64\xi^{-2/\phi} \phi^{-2} \) respectively.

### 1.3.3 Some generalisations of the Weibull distribution

One generalisation of the Weibull distribution is to introduce a threshold parameter \( \theta \), which means all lifetimes must exceed \( \theta \). The survivor function of this three-parameter Weibull distribution is

\[ S(y) = \exp(-\xi(y - \theta)^\phi). \quad (1.4) \]
Another generalisation of the Weibull model is the generalised extreme value distribution. This may be used as a model for \( Y \) or \( \log Y \). In the first case the survivor function for \( Y \) is

\[
S(y) = \exp\{-[1 - \zeta(y - \mu)/\sigma]^{-1/\zeta}\},
\]

(1.5)

where \( \mu(>0) \), \( \sigma(>0) \) and \( \zeta(<0) \) are parameters and \( y > (\sigma/\zeta) + \mu \). This is actually a re-parameterisation of the three-parameter Weibull distribution in (1.4), where \( \phi = -\frac{1}{\zeta}, \xi = \frac{\xi}{\sigma}, \) and \( \theta = -\frac{\mu + \sigma}{\zeta} \). The limiting case as \( \zeta \to 0 \) is the Gumbel distribution, with survivor function

\[
S(y) = \exp\{-\exp\left\{\frac{(y - \mu)}{\sigma}\right\}\}. \quad (1.6)
\]

When used as a model for \( \log Y \) the generalised extreme value distribution leads to the following survivor function for \( Y \).

\[
S(y) = \exp\{-[1 - \zeta(\log y - \mu)/\sigma]^{-1/\zeta}\},
\]

(1.7)

where \( y > 0 \), and \( \mu(>0), \sigma(>0) \) and \( \zeta(<0) \) are parameters. The limiting case as \( \zeta \to 0 \) is equivalent to the Weibull distribution, and (1.7) reduces to (1.2) with \( \mu = -\phi^{-1}\log \xi \) and \( \sigma = \phi^{-1} \).

### 1.3.4 Models Based on the Weibull Distribution

Suppose that covariates are represented quantitatively by a vector \( x \). One simple form of Weibull model uses the survivor function

\[
S(y; x) = \exp(-\xi_x y^\phi)
\]

where we let \( \log \xi_x = x^T \beta \) (log-linear model) and \( \phi \) is independent of \( x \). This is a fully parametric model with parameters \( \beta \) (regression coefficients) and \( \phi \) (Weibull shape parameter). The hazard function is \( h(y; x) = \xi_x \phi y^{\phi-1} \).

Let \( S_0(y) = \exp(-y^\phi) \) be a baseline survivor function, then

\[
S(y; x) = \exp(-\xi_x y^\phi) = \exp[-(\xi_x^{1/\phi}y)^u] = S_0(\xi_x^{1/\phi}y),
\]

thus, \( S(y; x) \) can be expressed as \( S_0(\psi_{ax}y) \) with \( \psi_{ax} = \xi_x^{1/\phi} \), an accelerated life model. Here \( \psi_{ax} \) is a positive function of \( x \). The interpretation is that, compared
with some standardised system in which $\psi_{ax} = 1$, the lifetime is divided by a factor $\psi_{ax}$, or equivalently, the decay of the system is accelerated by a factor $\psi_{ax}$. Similarly,

$$S(y; x) = \exp(-\xi_x y^\phi) = [\exp(-y^\phi)]^{\xi_x} = [S_0(y)]^{\xi_x},$$

thus, $S(y; x)$ can also be expressed as $[S_0(y)]^{\psi_{px}}$ with $\psi_{px} = \xi_x$, a proportional hazards model. Here, $\psi_{px}$ is also a positive function of $x$. Consider the ratio of the two hazard functions at time $y$ with different values of $x$,

$$\frac{h(y; x_1)}{h(y; x_2)} = \frac{\xi_{x_1}}{\xi_{x_2}},$$

which is independent of $y$. Therefore, the interpretation of this model is that the hazards at different $x$ values are in constant proportion over time.

Alternatively, the model can be presented in log form by

$$\log Y = -\frac{1}{\phi} \log \xi_x + \frac{1}{\phi} V$$

where $V$ has a Gumbel distribution with survivor function $\exp(-e^v)$ on $(-\infty, +\infty)$.

Applying standard results for the Gumbel model, the mean and variance of $\log Y$ may be obtained as

$$E(\log Y) = -\frac{1}{\phi} (\log \xi_x + \gamma),$$

and

$$Var(\log Y) = \pi^2/(6\phi^2),$$

where $\gamma$ is Euler’s constant. This representation also makes clear the role of $1/\phi$ as a scale constant for $\log Y$.

Within the framework of parametric models, there are a number of more general possibilities which could be considered. One possibility, see Crowder et al (1991), is to extend $S(y; x)$ and $\log Y$ in the form

$$S(y; x) = \exp(-\xi_x y^{\phi_x}), \quad \log Y = -\frac{1}{\phi_x} \log \xi_x + \frac{1}{\phi_x} V$$

so that both $\xi$ and $\phi$ depend on $x$. This may be appropriate when initial exploratory analysis of the data suggests that the Weibull shape parameter is not the same everywhere.
There are also possible extensions that involve the three-parameter Weibull or generalised extreme value distributions. For example, equation (1.4) may be extended to

\[ S(y; x) = \exp[-(y - \theta_x)^{\phi_x}] \]  

(1.10)

where any of \( \theta_x, \xi_x \) or \( \phi_x \) may be constant or may depend on \( x \). Similarly, but not equivalently, the log-representation in equation (1.5) may be extended to give

\[ S(y; x) = \exp[-(1 - \xi_x (\log y - \mu_x)/\sigma_x)^{-1/\xi_x}], \]  

(1.11)

in which the limiting case \( \xi_x \to 0 \) is Gumbel, but in general the distribution of \( \log Y \) is extended from Gumbel to generalised extreme value. The motivation for using equation (1.11) as opposed to equation (1.10) is partly based on the knowledge that the generalised extreme value parametrisation is more stable than the traditional three-parameter Weibull parametrisation (Smith & Naylor, 1987).

### 1.3.5 Estimation of the Weibull parameters

**Graphical method**

For the simplest Weibull model with uncensored observations, the parameters may be estimated roughly by graphical methods, see, for example, Lawless (1982) or Crowder et al (1991). Suppose that the lifetimes have been put in ascending order: \( y(1) < y(2) < ... < y(n) \). The empirical survivor function may be defined as

\[ \hat{S}(y(i)) = 1 - \frac{(i - 0.5)}{n}. \]

Since we have

\[ \log \{-\log S(y)\} = \phi \log y + \log \xi, \]

a plot of \( \log \{-\log[1 - (i - 0.5)/n]\} \) against \( \log y(i) \) should be roughly linear, if a Weibull model is appropriate for the data. Alternatively, the empirical survivor function may also be estimated in the forms such as, \( \hat{S}(y(i)) = 1 - \frac{1}{n+1} \) in practice. Furthermore, if the slope and intercept of the plot are \( a \) and \( b \) respectively, then rough estimates of \( \xi \) and \( \phi \) are \( \exp(b) \) and \( a \) respectively. These estimates may be used as starting values in an iterative scheme to obtain maximum likelihood estimates. Note that this method may be adapted in the presence of right censored observations. For a sample with \( r \) lifetimes observed and \( n - r \) right censored, the empirical survival function \( \hat{S}(y(i)) \) can be evaluated by the \( r \) ordered observed
lifetime $y(1) < y(2) < \ldots < y(r)$. The $\hat{S}(y(i))$ is undefined for the remaining $n - r$
censored observations.

**Least squares method**

For simple linear regression, $\log \xi = \beta_0 + \beta_1 x$, with uncensored data, a plot of
the observed $\log y$ against $x$ should be a straight line with slope $-\frac{\delta_1}{\phi}$ and intercept
$-\frac{\delta_0 + r}{\phi}$. The variance about this line is $\pi^2/(6\phi^2)$. Thus an ordinary least squares
fit will yield estimates of $\beta_0$, $\beta_1$ and $\phi$, though these are theoretically inefficient
compared with maximum likelihood estimates. In the multiple regression case
$\log \xi = x^T \beta$, one can similarly apply least squares methods to estimate $\beta$ and $\phi$, and to make rough checks of model assumptions. Lawless (1982) made detailed
comparisons between least squares and maximum likelihood fits for models of this
form.

**Maximum likelihood method**

More generally, the method of maximum likelihood may be applied. The log-
likelihood for a Weibull sample $y = (y_1, \ldots, y_n)$ is

$$
l(y_1, \ldots, y_n | \xi, \phi) = r \log \xi + r \log \phi + (\phi - 1) \sum_{i=1}^{n} \delta_i \log y_i - \xi \sum_{i=1}^{n} y_i^\phi,
$$

where $r = \sum_{i=1}^{n} \delta_i$, and $\delta_i$ is the indicator variable for the right-censored lifetimes
as defined in (1.1).

Thus, the first derivatives of $l$ are

$$
\frac{\partial l}{\partial \xi} = \frac{r}{\xi} - \sum_{i=1}^{n} y_i^\phi,
$$

$$
\frac{\partial l}{\partial \phi} = \frac{r}{\phi} + \sum_{i=1}^{n} \delta_i \log y_i - \xi \sum_{i=1}^{n} y_i^\phi \log y_i.
$$

The maximum likelihood estimator $\hat{\xi}$ of $\xi$ can be obtained explicitly by solving
$\frac{\partial l}{\partial \xi} = 0$ as

$$
\hat{\xi} = \frac{r}{\sum_{i=1}^{n} y_i^\phi}.
$$
Substitution into the equation \( \frac{\partial l}{\partial \phi} = 0 \) yields the form

\[
\frac{r}{\phi} + \sum_{i=1}^{n} \delta_i \log y_i - r \frac{\sum_{i=1}^{n} y_i^\phi \log y_i}{\sum_{i=1}^{n} y_i^\phi} = 0.
\]

Therefore, the maximum likelihood estimator \( \hat{\phi} \) of \( \phi \) may be obtained from the numerical solution to the equation above.

The second derivatives of \( l \) satisfy

\[
-\frac{\partial^2 l}{\partial \xi^2} = \frac{r}{\xi^2},
\]

\[
-\frac{\partial^2 l}{\partial \xi \partial \phi} = \sum_{i=1}^{n} y_i^\phi \log y_i,
\]

\[
-\frac{\partial^2 l}{\partial \phi^2} = \frac{r}{\phi^2} + \xi \sum_{i=1}^{n} y_i^\phi (\log y_i)^2.
\]

The entries of the 2x2 observed information matrix can be evaluated by substituting \((\xi, \phi)\) with their maximum likelihood estimates \((\hat{\xi}, \hat{\phi})\). Then, according to standard asymptotic theory, the variance-covariance matrix of \((\hat{\xi}, \hat{\phi})\) can be estimated from the inverse of the observed information matrix. Maximum likelihood estimators are asymptotically efficient. The method is computationally straightforward and the approximate standard errors may be evaluated easily. However, the method can yield highly biased estimates in small samples and/or with heavy censoring.

On the other hand, if it is assumed that there is no censoring, the elements of the Fisher information matrix can be derived,

\[
E\left(-\frac{\partial^2 l}{\partial \xi^2}\right) = \frac{1}{\xi^2},
\]

\[
E\left(-\frac{\partial^2 l}{\partial \xi \partial \phi}\right) = \frac{1}{\xi \phi} (1 - \gamma - \log \xi),
\]

\[
E\left(-\frac{\partial^2 l}{\partial \phi^2}\right) = -\frac{1}{\phi^2} [(1 - \gamma - \log \xi)^2 + \frac{\pi^2}{6}].
\]

The observed information matrix may also be evaluated from the Fisher information matrix by replacing \((\xi, \phi)\) in (1.15)-(1.17) by their maximum likelihood estimates \((\hat{\xi}, \hat{\phi})\).
1.3.6 Assessing the goodness of fit of the Weibull distribution

**Graphical methods**

Graphical methods are particularly valuable for checking the adequacy of the Weibull model. First, a quantile-quantile (QQ) plot can be applied. As mentioned in Section 1.3.5, a plot of \( \log \{- \log[1 - (i - 0.5)/n]\} \) against \( \log y(i) \) should be roughly linear, if a Weibull model is appropriate for the data. As mentioned before, equivalently, \( \log \{- \log[1 - (i - 0.5)/n]\} \) may be replaced by \( \log \{- \log[1 - i/(n + 1)]\} \). Secondly, the probability (PP) plot involves plotting \( 1-(i-0.5)/n \) against \( \exp(-\tilde{\xi} y(i)^{\tilde{\phi}}) \), where \((\tilde{\xi}, \tilde{\phi})\) are the maximum likelihood estimates under the Weibull model. A straight line of unit slope through the origin is indicative of a good agreement between the Weibull model and data. Finally, a generalised residual plot by first taking logarithms of the observations is presented in Crowder et al (1991, Section 4.7).

**Likelihood ratio test**

If we embed the proposed model in a more general alternative model, then the corresponding likelihood ratio test yields a goodness of fit test. In the Weibull context, we may test the adequacy of a Weibull model relative to a generalised extreme value model. The log likelihood ratio statistic \( LR = 2(l_1 - l_0) \) has an approximate \( \chi^2(n) \) null distribution, where \( l_0 \) and \( l_1 \) are the maximised log likelihood values evaluated under the Weibull and the unrestricted model, respectively.

**Test based on EDF statistics**

We may also apply a goodness of fit test based on the empirical distribution function (EDF). Suppose that the lifetimes are put in ascending order \( y(1) < y(2) < \ldots < y(n) \). The EDF is defined by

\[
\hat{F}(y(i)) = \frac{i}{n}.
\]

A statistic measuring the vertical difference between \( \hat{F}(y) \) and \( F(y) \) is called an EDF statistic. There are various commonly used tests based on EDF statistics.
For example,
\[ D^+ = \sup \{ \hat{F}(y) - F(y) \}, \]
and
\[ D^- = \sup \{ F(y) - \hat{F}(y) \}. \]

Some other well known EDF statistics are such as \( D \), introduced by Kolmogorov (1933):
\[ D = \sup |\hat{F}(y) - F(y)| = \max(D^+, D^-), \]
a closely related statistic \( V \), given by Kuiper (1960):
\[ V = D^+ + D^-, \]
and the Watson (1961) statistic \( U^2 \) defined by
\[ U^2 = n \int_{-\infty}^{\infty} \left\{ \hat{F}(y) - F(y) - \int_{-\infty}^{\infty} [\hat{F}(y) - F(y)]dF(y) \right\}^2 dF(y). \]

As stated in Section 1.3.3, the Weibull distribution becomes an extreme value (Gumbel) distribution (1.6) after the transformation \( X = \log(Y) \). All the above EDF statistics may be calculated by evaluating \( \hat{F}(x) \), and \( F(x) \), where \( F(x) \) is the EDF of the extreme value distribution, its parameters may be estimated by maximum likelihood. D’Agostino (1986, Section 4.10), gives the upper tail percentage points for all these EDF statistics for the extreme value distribution. Hence, the test of the fitness of the Weibull distribution may be carried out by comparing the values in the tables with the calculated EDF statistics.

EDF statistics are also available to handle the censored data. The versions of EDF concerning the censored cases and their corresponding tables of upper tail percentage points are given by D’Agostino (1986).

### 1.4 Frailty Models

The term ‘frailty’ is introduced by Vaupel et al (1979) to describe an unobservable random effect associated with each individual. Frailty models are effective in extending the class of survival models and including dependence in multivariate survival distributions.
Several methods of modelling the frailty in survival data have been developed during recent years. The most well developed and widely applied model is the generalisation of the proportional hazards approach, which allows for the random effect as a multiplicative adjustment to a baseline hazard function, see Clayton (1985), Hougaard (1984) and Oakes (1989a). Anderson & Louis (1995) have proposed a scale change frailty model, which is similar to the proportional hazards frailty model, but incorporates unobserved random effects into the baseline hazard function to change the time scale. Rocha (1996) has suggested an alternative model, where frailty acts additively on the hazard function.

The Weibull based random effects models which are discussed throughout this thesis refer to the generalisation of proportional hazard models with a Weibull baseline hazard function. Therefore, the following subsections focus on a review of a proportional hazards frailty model.

1.4.1 Univariate frailty models

In univariate survival analysis, the variability of lifetimes can be divided into two parts, one is observable risk factors which can be included into a model as covariates, the other is unobserved risk factors, which is frailty. Those individuals with larger frailty values are at higher risk of an event than those with lower frailty values. By introducing frailty into life table analysis, Hougaard (1984) has explained the reason that a group of patients after an operation shows decreasing hazard, though each individual may have constant hazard. Aalen (1988) has examined the impact of individual heterogeneity in univariate survival analysis in the medical context. He discusses a class of mixing distributions and extends Hougaard's model to allow for part of the population to be non-susceptible. Keiding et al (1997) present a case study and argue that accelerated failure models may be preferable in accounting for heterogeneity in univariate survival times due to 'missing' (omitted, unrecorded) covariates.

Let $W$ be a non-negative random variable with density function $g(w)$ and let $h(y)$ be the baseline hazard function. Conditionally on the frailty $W$, the hazard function for the survival time $y$ is

$$ h(y|W) = Wh(y) \exp(x^T \beta), $$

14
where \( x \) is the vector of observed covariates for each individual. The frailty \( W \) is defined as a random multiplicative factor acting on the hazard function. The model presents the population as a mixture where each individual has frailty \( W \) whereas the baseline hazard is common to all the individuals.

Considering the simplest case without covariates, the conditional survivor function of \( y \) given \( W \) is obtained from

\[
S(y|W = w) = \exp(-\int_0^y h(t|w)dt) = \exp(-w\int_0^y h(t)dt) = [B(y)]^w,
\]

where \( B(y) \) denotes the baseline survivor function, say, \( B(y) = \exp(-\int_0^y h(t)dt) \).

Therefore, if random effect \( W \) has distribution function \( G(. \) ), the unconditional survivor function of \( y \) is

\[
S(y) = \int_0^\infty [B(y)]^w dG(w) = \int_0^\infty \exp\{-w(-\log B(y))\} dG(w) = L(u)
\]

where \( L(u) \) denotes the Laplace transformation of \( u \), and \( u = -\log\{B(y)\} \), which is the baseline cumulative hazard function. It is apparent that this unconditional survivor function can be extended to accommodate the situation with covariates.

Many authors have discussed the identifiability and assumptions concerning the form of the frailty. Heckman & Singer (1984) demonstrate high sensitivity of results to alternative choices of finite mean frailty distribution. Elbers & Ridder (1982) study the conditions necessary to achieve identifiability of the frailty distribution in univariate data. Lancaster & Nickell (1980) note that ignoring frailty with finite mean would result in a bias towards zero in the parameter estimates.

### 1.4.2 Multivariate frailty models

Multivariate lifetime data arise when each study subject may experience several events or when there exists some natural or artificial grouping of subjects which induces dependence among lifetime of the same group. Examples in a medical
context are the sequence of tumour recurrences or infection episodes, the development of physical symptoms or disease in several organ systems, the occurrence of blindness in the left and right eyes. Examples in genetic studies involve the lifetimes among the family members with genetic disease. Examples in reliability studies include failure times of different types of component, repeated breakdowns of a certain type of machinery. A multivariate frailty model may be an effective way to accommodate association in correlated lifetime data.

The generalisation of frailty models from univariate survival analysis to multivariate lifetime data analysis is immediate. Let $Y_{ij}$ denote the survival time of individual $i$ in group $j$, where $i = 1, \ldots, n$, $j = 1, \ldots, p$. Suppose that, conditional on the frailty $W_i$, all individuals are independent and the hazard of $Y_{ij}$ is $W_i h_j(y_{ij})$. Furthermore, we assume that the $W_i$ are independent and identically distributed with distribution function $G(w)$. The survivor function of $Y_1, \ldots, Y_p$ conditional on $W$ is given by

$$S(y_1, \ldots, y_p | W = w) = \exp \left( - \sum_{j=1}^{p} \int_0^{y_j} h_j(t_j | w) dt_j \right) = \prod_{j=1}^{p} [B_j(y_j)]^w,$$

where $B_j(.)$ denotes the baseline survivor function for group $j$, say, $B_j(y_j) = \exp \left( - \int_0^{y_j} h_j(t_j) dt_j \right)$. This model represents that frailty $W$ varies across the individuals, but remains the same for each individual in the groups. The baseline hazard is common to all the individuals in the same group. Furthermore, the unconditional survivor function is given by

$$S(y_1, \ldots, y_p) = \int \prod_{j=1}^{p} [B(y_j)]^w dG(w) = L(u), \quad (1.19)$$

where $u = \sum_{j=1}^{p} \left\{ - \log B(y_j) \right\}$, which is the sum of $p$ cumulative hazard functions, and $L(u)$ denotes the Laplace transformation of $u$. It is shown in Chapter 2 that the Weibull based random effects models are based on this expression. As in the univariate case, this model can be easily extended to accommodate the situation with covariates.

Various published papers cover aspects of the study of multivariate frailty models. Clayton & Cuzick (1985) generalise the proportional hazards model to the problem of bivariate lifetime data with unspecified marginal distributions which are
related by a single association, where gamma frailty is assumed. Oakes (1989a) extends the class of bivariate frailty models, which is presented as a subclass of the Archimedean copula distributions, to allow for negative association, and introduces a cross-ratio function to measure the association. Hougaard (1986b) considers a multivariate survival model with positive stable frailty. Pickles & Crouchley (1995) examine the performance of conditional and mixture likelihood approaches to estimating bivariate frailty models, and conclude that the choice of a particular distribution for frailty is not critical for the estimation and testing of regression coefficients in survival models. Shih & Louis (1995) propose a graphical method for assessing the adequacy of a gamma frailty model. Again, Shih & Louis (1996) discuss the estimation of the association parameter by using non-parametric estimation of marginal survival functions. Bandeen-Roche & Liang (1996) propose a family of frailty models that accounts for multiple levels of clustering of lifetime data.

1.5 Examples

Three examples related to various application areas are used in this thesis to illustrate the application of the methods proposed in later chapters.

1.5.1 Infant nutrition data

The original data source is from a study which was conducted in Madrid, Spain. The aim of the study was to investigate the feeding practices adopted by 344 mothers whose children at the time of the study were aged between 3 and 19 months. Most of the raw data collected from the study were in the form of event times, such as the age of the infant at which a certain food was first introduced or the age of the infant at which use of a certain type of milk feeding was stopped. Details of the study design and data collection may be found in van den Boom (1994).

Since fish and egg are foods which are well known for being potentially allergenic, it is of interest to investigate the bivariate age distribution of first introduction of fish and egg. Table 1.1 lists the ages in months at which fish or egg were first
Introduction of fish

Introduction of egg

Figure 1.2: Age in months of introduction of fish and egg
given to 55 infants aged 18 or 19 months in this study. This coherent subset of the data comprises the 55 oldest children and was selected in order to illustrate the methods on uncensored data. Data were rounded to the nearest half month or month. Strictly speaking, such data ought to be taken as interval censored. However, we will ignore this aspect in our discussion. Non-parametric univariate analysis was carried out on the full data by van den Boom et al (1995). However, it is natural to consider a Weibull based approach as a starting point of a parametric analysis. Indeed van den Boom et al (1995) found that a Weibull distribution gave a good fit to the age at which breast feeding ceased conditional on breast feeding being used at all. Kimber (1996) performed a heterogeneity test on the data after postulating a particular type of Weibull based random effects model.

Initial examination of the data shows that the correlation between the log-transformed ages of introduction of fish and egg is \( \rho(\log Y_1, \log Y_2) = 0.4 \). A QQ plot, i.e. a plot of \( \log\{- \log(1 - \frac{i - 0.5}{n})\} \) against \( \log y(i) \) for the ages of introducing fish and egg are displayed in Figure 1.2, which shows reasonable straight lines for both ages. Therefore, Weibull distributions provide adequate fits to the ages of introducing fish and egg. The Weibull maximum likelihood estimates for the age of introducing fish are \( \hat{\xi}_1 = -8.817 \), and \( \hat{\phi}_1 = 4.166 \). The corresponding estimates for the age of introducing egg are \( \hat{\xi}_2 = -9.064 \), and \( \hat{\phi}_2 = 3.815 \), respectively. These estimates were obtained by considering the marginal distributions separately.
Table 1.1: Age in months of introduction of fish and egg into the diets of 55 infants

<table>
<thead>
<tr>
<th>Fish</th>
<th>Egg</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.0</td>
<td>4.0</td>
<td>1</td>
</tr>
<tr>
<td>5.0</td>
<td>10.0</td>
<td>1</td>
</tr>
<tr>
<td>5.5</td>
<td>12.0</td>
<td>1</td>
</tr>
<tr>
<td>6.0</td>
<td>7.0</td>
<td>1</td>
</tr>
<tr>
<td>6.0</td>
<td>9.0</td>
<td>3</td>
</tr>
<tr>
<td>6.0</td>
<td>11.0</td>
<td>1</td>
</tr>
<tr>
<td>6.0</td>
<td>16.0</td>
<td>1</td>
</tr>
<tr>
<td>7.0</td>
<td>7.0</td>
<td>5</td>
</tr>
<tr>
<td>7.0</td>
<td>10.0</td>
<td>2</td>
</tr>
<tr>
<td>7.0</td>
<td>18.0</td>
<td>1</td>
</tr>
<tr>
<td>8.0</td>
<td>8.0</td>
<td>6</td>
</tr>
<tr>
<td>8.0</td>
<td>10.0</td>
<td>1</td>
</tr>
<tr>
<td>9.0</td>
<td>9.0</td>
<td>2</td>
</tr>
<tr>
<td>9.0</td>
<td>12.0</td>
<td>1</td>
</tr>
<tr>
<td>9.0</td>
<td>16.0</td>
<td>1</td>
</tr>
<tr>
<td>10.0</td>
<td>10.0</td>
<td>1</td>
</tr>
<tr>
<td>10.0</td>
<td>14.0</td>
<td>1</td>
</tr>
<tr>
<td>12.0</td>
<td>12.0</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fish</th>
<th>Egg</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.0</td>
<td>6.0</td>
<td>1</td>
</tr>
<tr>
<td>5.0</td>
<td>12.0</td>
<td>1</td>
</tr>
<tr>
<td>6.0</td>
<td>6.0</td>
<td>1</td>
</tr>
<tr>
<td>6.0</td>
<td>8.0</td>
<td>2</td>
</tr>
<tr>
<td>6.0</td>
<td>10.0</td>
<td>1</td>
</tr>
<tr>
<td>6.0</td>
<td>12.0</td>
<td>2</td>
</tr>
<tr>
<td>6.5</td>
<td>9.0</td>
<td>1</td>
</tr>
<tr>
<td>7.0</td>
<td>9.0</td>
<td>3</td>
</tr>
<tr>
<td>7.0</td>
<td>11.0</td>
<td>1</td>
</tr>
<tr>
<td>8.0</td>
<td>7.0</td>
<td>1</td>
</tr>
<tr>
<td>8.0</td>
<td>9.0</td>
<td>2</td>
</tr>
<tr>
<td>8.0</td>
<td>12.0</td>
<td>1</td>
</tr>
<tr>
<td>9.0</td>
<td>11.0</td>
<td>1</td>
</tr>
<tr>
<td>9.0</td>
<td>13.0</td>
<td>1</td>
</tr>
<tr>
<td>9.0</td>
<td>9.0</td>
<td>1</td>
</tr>
<tr>
<td>10.0</td>
<td>9.0</td>
<td>1</td>
</tr>
<tr>
<td>10.0</td>
<td>12.0</td>
<td>1</td>
</tr>
<tr>
<td>12.0</td>
<td>10.0</td>
<td>1</td>
</tr>
</tbody>
</table>

1.5.2 Repeated endurance exercise tests data

The data in Table 1.2 give repeated exercise times (in seconds) to angina pectoris in patients with coronary heart disease. The data were originally presented by Danahy et al. (1977) to study the effect of high dose oral isosorbide dinitrate on exercise time until angina pectoris. 21 patients with coronary heart disease pedalled exercise bikes until they experienced angina. They were administered an oral dose of isosorbide dinitrate thereafter, and were required to return to the bike carrying on the exercise at 1 hour and 3 hours after drug treatment. One particular feature of this data is that the response times are censored for some of the cases. Table 1.2 shows that, of the 21 patients, seven did not experience angina when exercising 1 hour after treatment; four did not experience angina when exercising 3 hours after treatment. These exercise times were censored because patients became too exhausted to continue. The other feature of these data is that each patient was given a different dose of drug, which allows investigation of the effect of dose on the time to onset of angina. Therefore, the dose of the drug is a covariate that might be considered.
Table 1.2: Exercise times to angina (in seconds) on occasions before and after oral isosorbide dinitrate

<table>
<thead>
<tr>
<th>Time</th>
<th>0 hour</th>
<th>1 hour</th>
<th>3 hours</th>
<th>Dose (mm/kg)</th>
<th>Time</th>
<th>0 hour</th>
<th>1 hour</th>
<th>3 hours</th>
<th>Dose (mm/kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>136</td>
<td>(445)</td>
<td>(393)</td>
<td>0.58</td>
<td></td>
<td>250</td>
<td>306</td>
<td>206</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>215</td>
<td>232</td>
<td>258</td>
<td>0.24</td>
<td></td>
<td>235</td>
<td>248</td>
<td>298</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>129</td>
<td>121</td>
<td>110</td>
<td>0.38</td>
<td></td>
<td>425</td>
<td>580</td>
<td>613</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>441</td>
<td>(504)</td>
<td>(519)</td>
<td>0.41</td>
<td></td>
<td>208</td>
<td>264</td>
<td>210</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>154</td>
<td>110</td>
<td>123</td>
<td>0.37</td>
<td></td>
<td>89</td>
<td>145</td>
<td>172</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>230</td>
<td>264</td>
<td>0.24</td>
<td></td>
<td>147</td>
<td>403</td>
<td>290</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>231</td>
<td>(540)</td>
<td>370</td>
<td>0.49</td>
<td></td>
<td>224</td>
<td>432</td>
<td>291</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>152</td>
<td>(733)</td>
<td>492</td>
<td>0.20</td>
<td></td>
<td>417</td>
<td>(743)</td>
<td>566</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>213</td>
<td>250</td>
<td>150</td>
<td>0.38</td>
<td></td>
<td>490</td>
<td>(559)</td>
<td>(557)</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>406</td>
<td>651</td>
<td>624</td>
<td>0.51</td>
<td></td>
<td>229</td>
<td>327</td>
<td>280</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>265</td>
<td>(565)</td>
<td>(504)</td>
<td>0.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: data in the brackets are censored

Figure 1.3 shows the plots of \( \log\{-\log(1 - \frac{i-0.5}{n})\} \) against \( \log y(i) \) for the exercise times before treatment, 1 hour and 3 hours after drug treatment, respectively. The plots looks reasonably linear, suggesting that a Weibull model fits the three marginals. Initial straight lines fits were obtained based on plotting the uncensored data points, as given in Figure 1.3, which yields slopes and intercepts as \( \tilde{\phi}_1 = 2.82, \log \tilde{\xi}_1 = -15.88 \) for exercise times before treatment; \( \tilde{\phi}_2 = 2.41, \log \tilde{\xi}_2 = -14.05 \) for exercise times 1 hour after treatment; \( \tilde{\phi}_3 = 2.42, \log \tilde{\xi}_3 = -14.05 \) for exercise times 3 hours after treatment, respectively. These values are used as initial values in iterative procedures obtaining maximum likelihood estimates. The Weibull maximum likelihood estimates for each exercise time are \( \hat{\phi}_1 = 2.41, \log \hat{\xi}_1 = -13.65, \hat{\phi}_2 = 1.61, \log \hat{\xi}_2 = -10.19, \hat{\phi}_3 = 1.98, \log \hat{\xi}_3 = -11.96 \). These results ignore the covariate. A further treatment will be given in Section 5.5.2.

1.5.3 Fibre failure strength data

Table 1.3 contains a set of data on fibre failure strengths. The breaking strengths of fibre sections of length 5mm, 12mm, 30mm and 75mm, which are cut from the same fibre, are listed. The data has missing values which are indicated as zero representing any accidental breakage prior to testing. A Weibull based model for strength is a natural starting point, but there is concern about the possibility of
Figure 1.3: Exercise times to angina pectoris on 3 occasions
Table 1.3: Fibre failure strengths for fibre section of different lengths

<table>
<thead>
<tr>
<th>Fibre No.</th>
<th>Length (mm)</th>
<th>Fibre No.</th>
<th>Length (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3.30 3.32 2.39 2.08</td>
<td>5</td>
<td>4.17 3.67 2.49 2.06</td>
</tr>
<tr>
<td>12</td>
<td>4.19 4.27 3.16 2.05</td>
<td>12</td>
<td>3.64 2.41 2.20 1.80</td>
</tr>
<tr>
<td>30</td>
<td>2.73 2.24 1.91 1.68</td>
<td>30</td>
<td>4.47 4.06 2.74 2.22</td>
</tr>
<tr>
<td>75</td>
<td>3.29 3.08 2.44 2.37</td>
<td>75</td>
<td>3.55 2.35 2.38 2.37</td>
</tr>
<tr>
<td>9</td>
<td>3.03 2.26 1.64 2.03</td>
<td>9</td>
<td>6.41 5.11 2.98 2.39</td>
</tr>
<tr>
<td>11</td>
<td>5.16 4.60 2.99 2.30</td>
<td>11</td>
<td>4.92 3.03 2.80 2.30</td>
</tr>
<tr>
<td>13</td>
<td>3.01 3.17 2.41 2.07</td>
<td>13</td>
<td>4.01 2.91 2.18 1.83</td>
</tr>
<tr>
<td>15</td>
<td>5.09 3.87 2.24 2.09</td>
<td>15</td>
<td>4.65 3.82 2.59 2.48</td>
</tr>
<tr>
<td>17</td>
<td>4.57 4.07 2.40 2.22</td>
<td>17</td>
<td>3.48 2.14 2.35 2.05</td>
</tr>
<tr>
<td>19</td>
<td>3.05 2.96 1.91 2.20</td>
<td>19</td>
<td>3.60 2.92 2.42 2.09</td>
</tr>
<tr>
<td>21</td>
<td>4.60 4.28 2.86 2.13</td>
<td>21</td>
<td>4.38 3.03 2.53 2.31</td>
</tr>
<tr>
<td>23</td>
<td>3.50 3.46 2.56 2.13</td>
<td>23</td>
<td>4.43 4.26 2.63 2.16</td>
</tr>
<tr>
<td>25</td>
<td>4.58 4.61 2.75 2.17</td>
<td>25</td>
<td>4.76 3.64 2.88 2.43</td>
</tr>
<tr>
<td>27</td>
<td>4.64 3.20 2.52 2.35</td>
<td>27</td>
<td>2.65 2.01 1.87 2.12</td>
</tr>
<tr>
<td>29</td>
<td>5.03 3.85 3.12 2.53</td>
<td>29</td>
<td>5.15 3.35 2.78 2.36</td>
</tr>
<tr>
<td>31</td>
<td>3.35 2.91 2.50 2.07</td>
<td>31</td>
<td>3.62 3.31 2.50 2.08</td>
</tr>
<tr>
<td>33</td>
<td>4.04 3.35 2.41 2.37</td>
<td>33</td>
<td>3.06 2.49 2.09 2.21</td>
</tr>
<tr>
<td>35</td>
<td>4.55 2.67 2.40 2.28</td>
<td>35</td>
<td>3.23 2.27 1.92 2.12</td>
</tr>
<tr>
<td>37</td>
<td>6.20 5.10 3.47 2.24</td>
<td>37</td>
<td>3.75 2.48 2.48 2.07</td>
</tr>
<tr>
<td>39</td>
<td>3.33 2.23 2.33 2.13</td>
<td>39</td>
<td>3.47 2.51 0.0 1.76</td>
</tr>
<tr>
<td>41</td>
<td>3.70 2.31 0.0 2.06</td>
<td>41</td>
<td>3.77 2.26 0.0 2.20</td>
</tr>
<tr>
<td>43</td>
<td>0.0 2.37 0.0 0.0</td>
<td>43</td>
<td>0.0 2.39 0.0 0.0</td>
</tr>
<tr>
<td>45</td>
<td>0.0 2.41 0.0 0.0</td>
<td>45</td>
<td>0.0 2.41 0.0 0.0</td>
</tr>
</tbody>
</table>

extra variability in fibre strength.

The results of an initial examination of the data are shown in Table 1.4, where $\phi_0$ and $\log \xi_0$ are slopes and intercepts of the fitted straight lines based on the Weibull models. The maximum likelihood estimates from fitting the Weibull models separately and their corresponding maximised log-likelihoods are also listed in the table. A further check on the suitability of the Weibull models are done by plotting $\log\{-\log(1 - \frac{i-0.5}{n})\}$ against $\log y(i)$ for each fibre failure strengths as shown in Figure 1.4. Clearly, the adequacy of the Weibull models is in some doubt, particularly for the shorter fibre lengths in the lower tails of the distribution, though the variability in such plots is greatest in the tails (see Michael, 1983 and Kimber, 1985).
Table 1.4: Summary results of the fibre failure strengths data

<table>
<thead>
<tr>
<th></th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$y_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>42</td>
<td>45</td>
<td>39</td>
<td>42</td>
</tr>
<tr>
<td>mean of $y$</td>
<td>4.0503</td>
<td>3.1781</td>
<td>2.4928</td>
<td>2.1657</td>
</tr>
<tr>
<td>variance of $y$</td>
<td>0.7537</td>
<td>0.7096</td>
<td>0.1459</td>
<td>0.0344</td>
</tr>
<tr>
<td>mean of $\log(y)$</td>
<td>1.3772</td>
<td>1.1232</td>
<td>0.9017</td>
<td>0.7690</td>
</tr>
<tr>
<td>variance of $\log(y)$</td>
<td>0.0438</td>
<td>0.0667</td>
<td>0.0244</td>
<td>0.0078</td>
</tr>
<tr>
<td>$\phi_0$</td>
<td>2.5022</td>
<td>2.0269</td>
<td>3.3519</td>
<td>5.9302</td>
</tr>
<tr>
<td>$\log \xi_0$</td>
<td>-2.8687</td>
<td>-1.6993</td>
<td>-2.4453</td>
<td>-3.9830</td>
</tr>
<tr>
<td>MLE of $\phi$</td>
<td>4.8216</td>
<td>4.04034</td>
<td>6.9614</td>
<td>13.6345</td>
</tr>
<tr>
<td>MLE of $\log \xi$</td>
<td>-7.1463</td>
<td>-5.0633</td>
<td>-6.7987</td>
<td>-11.0399</td>
</tr>
<tr>
<td>log-likelihood</td>
<td>-55.0317</td>
<td>-56.3477</td>
<td>-18.8276</td>
<td>12.13726</td>
</tr>
</tbody>
</table>

Figure 1.4: Fibre failure strengths for fibre section of different lengths
Chapter 2

Weibull based random effects models

2.1 Introduction

An approach to generating Weibull based random effects models is to consider a mixture, either continuous or discrete, of the Weibull distribution as follows. In the univariate case, suppose that, given a quantity $W$, lifetime $Y$ has a Weibull distribution with survivor function $\exp(-W \xi y^\phi)$. Then, the unconditional survivor function of $Y$ is

$$S(y) = \int_0^\infty \exp(-we^{\xi y^\phi})dG(w), \quad (2.1)$$

where $G(.)$ is the distribution function of random effect $W$, $\xi(> 0)$ and $\phi(> 0)$ are Weibull parameters. Note that when $\phi = 1$, $Y$ is conditionally exponentially distributed. The expression (2.1) is the special case of the frailty model in (1.19) where the baseline survivor function $B(.)$ is taken to be of Weibull form.

The generalised form of Weibull based random effects model for the multivariate case is given similarly: conditional on a quantity $W$, suppose that random variables $Y = (Y_1, \ldots, Y_p)$ have joint survivor function

$$S(y_1, \ldots, y_p|W = w) = \exp(-ws), \quad (2.2)$$

where

$$s = \sum_{j=1}^p \xi_j y_j^{\phi_j}, \quad (2.3)$$

$\xi_j$ and $\phi_j$ are Weibull parameters, with $\xi_j > 0$, $\phi_j > 0$. That is, $Y_1, \ldots, Y_p$, conditionally on $W$, are independent Weibull random variables. If $W$ varies with
distribution function \( G(.) \) on \((0, \infty)\), then the unconditional joint survivor function of \( Y_1, \ldots, Y_p \) is

\[
S(y_1, \ldots, y_p) = \int_0^\infty \exp(-ws)dG(w),
\]

(2.4)

Again, (2.4) is a special case of the multivariate frailty model in (1.19) where the baseline survivor functions \( B_j(.) \) are taken to be of Weibull form.

Different choices of \( G \) give rise to different unconditional survivor functions for \( Y \). If \( G \) is an appropriate discrete distribution, a finite mixture distribution for \( Y \) is obtained. However, note that even a two-component mixture, the simplest possible such finite mixture, has four parameters. In the interest of parsimony we shall concentrate on simple continuous mixtures that lead to tractable unconditional survivor functions. In this chapter we exhibit and compare some of the features of Weibull based random effects models, in the cases in which \( G \) is a continuous distribution corresponding to a gamma distribution, a positive stable distribution and an inverse Gaussian distribution.

2.2 Gamma mixture of Weibulls distribution

The gamma mixture of Weibulls (GW) distribution, also called the multivariate Burr distribution, is one of the most tractable Weibull based random effects models used in lifetime analysis. The univariate Burr distribution was originally given by Burr (1942) in the context of developing systems of distributions. The derivation of the multivariate Burr (GW) as a gamma mixture of independent Weibull random variables is due to Takahasi (1965), who gives various properties for this multivariate distribution. The GW distribution is further developed by Crowder (1985), who applies the GW distribution as a standard model for repeated failure time measurements.

Assuming random effect \( W \) follows a gamma distribution with shape parameter \( \nu \) and unit scale parameter, the joint survivor function of the GW distribution is obtained as

\[
S(y_1, \ldots, y_p) = \int_0^\infty \exp(-ws)dG(w)
= \int_0^\infty e^{-w\nu} \frac{w^{\nu-1}e^{-w}}{\Gamma(\nu)} dw = (1 + s)^{-\nu},
\]

(2.5)
Figure 2.1: Density and hazard functions for the Burr distribution, where $\nu = 2$, $\xi = 1$, and $\phi$ varies corresponding to 0.5, 1.5, 2.5 and 5.0

where $s$ was defined in (2.3). There is no loss of generality to have a unit gamma scale parameter since any other value may be absorbed in the $\xi_j$ parameters.

### 2.2.1 Univariate case

In the univariate case, where $p = 1$, the survivor function of $Y$ becomes

$$S(y) = (1 + \xi y^\phi)^{-\nu},$$

which is clearly not a Weibull distribution. Instead, it is known as a Burr distribution. Its density function

$$f(y) = \nu \xi \phi y^{\phi - 1}(1 + \xi y^\phi)^{-\nu - 1}$$

is plotted in Figure 2.1(a) for various values of $\phi$. The hazard function

$$h(y) = \frac{\nu \xi \phi y^{\phi - 1}}{1 + \xi y^\phi},$$

is either decreasing($\phi \leq 1$) or upturned bathtub shaped($\phi > 1$). The parameter $\nu$ acts as a scaling parameter on the hazard function. Figure 2.1(b) shows a plot of the hazard function for various values of $\phi$. 

26
2.2.2 Multivariate case

Density

The joint density function for the \( p \)-variate GW distribution is of the form

\[
f(y_1, \ldots, y_p) = \prod_{j=1}^{p} (\nu + j - 1) \prod_{j=1}^{p} (\xi_j y_j^{\phi_j - 1}) \left(1 + \sum_{j=1}^{p} \xi_j y_j^{\phi_j}\right)^{-1+p}.
\]  

A contour plot of the density function for the bivariate case is given in Figure 2.4(a).

Hazard function

There are various definitions of multivariate failure rate in the literature. According to Brindly & Thompson (1972)'s definition of multivariate failure rate, a multivariate failure time distribution is IFR/DFR if

\[
q = P(Y > y + \Delta | Y > y)
\]

is decreasing/increasing in \( y \) for all \( \Delta > 0 \); \( 1_p \) here denotes \( (1, \ldots, 1) \).

Crowder (1985) shows that, for the \( p \)-variate GW distribution,

\[
q = \frac{S(y_1 + \Delta, \ldots, y_p + \Delta)}{S(y_1, \ldots, y_p)} = \frac{(1 + \sum_{j=1}^{p} \xi_j (y_j + \Delta y_j)^{\phi_j})^{-\nu}}{(1 + \sum_{j=1}^{p} \xi_j y_j^{\phi_j})^{-\nu}}.
\]

Thus, \( q \) has log-derivatives

\[
\frac{\partial}{\partial y_k} (\log q) = \nu \xi_k \phi_k \left\{ \frac{y_k^{\phi_k-1}}{1 + \sum_{j=1}^{p} \xi_j y_j^{\phi_j}} - \frac{(y_k + \Delta)^{\phi_k-1}}{1 + \sum_{j=1}^{p} \xi_j (y_j + \Delta)^{\phi_j}} \right\},
\]

for \( k = 1, \ldots, p \). For sufficiently large \( y_k \) this expression is positive, and tends to 0 as \( y_k \to \infty \); for small \( y_k \), and \( \phi_k > 1 \), it is negative. Hence for max(\( \phi_k \)) > 1 the \( p \)-variate GW distribution can be neither IFR nor DFR. For max(\( \phi_k \)) ≤ 1, the \( p \)-variate GW distribution is DFR.

Another form of the multivariate hazard rate defined by Johnson & Kotz (1972), is the conditional hazard function for \( Y_k \) given \( Y > y \)

\[
h_k(y_k | Y > y) = \lim_{\Delta \to 0} \Delta^{-1} P(Y_k \leq y_k + \Delta | Y > y) = -\partial[\log S(y)]/\partial y_k.
\]

27
The behaviour of the multivariate hazard rate is studied in the component variate $y_k$, and its conditional hazard.

For the $p$-variate GW distribution, the conditional hazard function is

$$h_k(y_k|Y > y) = \frac{\nu \xi_k \phi_k y_k^{\phi_k - 1}}{1 + \sum_{j=1}^{p} \xi_j y_j^{\phi_j}},$$

which is decreasing in $y_k$, if $\phi_k \leq 1$; and upturned bathtub shaped, if $\phi_k > 1$.

### Marginal and conditional distributions

Let $y_A(a \times 1)$ and $y_B(b \times 1)$ be complementary sub-vectors of $y = (y_1, \ldots, y_p)$, so that $a + b = p$. Furthermore, we denote $s_A = \sum_A \xi_j y_j^{\phi_j}$ with summation just over those $y_j$ in $y_A$. Therefore, the marginal survivor function of $Y_A$ for the $p$-variate GW distribution is obtained by setting $y_B = 0$ in (2.5)

$$S_A(y_A) = (1 + s_A)^{-\nu},$$

(2.7)

which is clearly of the GW distribution form in (2.5).

In multivariate lifetime data analysis, the situation that data are only recorded if their survival time exceeds some known period is encountered sometimes. That is to say, a case is only accepted for the study if $Y_B > y_B$. The conditional distribution of $Y_A$ given that $Y_B > y_B$ is applied for dealing with this case.

For the $p$-variate GW distribution, this conditional distribution of $Y_A$ is of the GW form

$$P(Y_A > y_A|Y_B > y_B) = \frac{S(y)}{S_B(y_B)} = (1 + \sum_A \xi_j y_j^{\phi_j})^{-\nu},$$

(2.8)

where $\xi'_j = \xi_j / (1 + \sum_B \xi_j y_j^{\phi_j})$ is a modified parameter, see Crowder (1985).

### Distributions of minima

Two types of minima are considered, see Crowder (1985). First, let $\Omega = \min\{Y_j : j = 1, \ldots, p\}$. Then $\Omega$ for the $p$-variate GW distribution has survivor function

$$S(\omega) = (1 + \sum_{j=1}^{p} \xi_j \omega^{\phi_j})^{-\nu}.$$
which is of univariate Burr form when $\phi_1 = \ldots = \phi_p$.

Secondly, suppose that $Y_1, \ldots, Y_n$ are independently and identically distributed, with $Y_i = (Y_{i1}, \ldots, Y_{ip})$. Let $\Omega' = (\Omega'_1, \ldots, \Omega'_p)$, where $\Omega'_j = \min\{Y_{ij} : i = 1, \ldots, n\}$, so that $\Omega'$ is the vector of componentwise minima. This would be relevant to a situation in which only the first event is observed for each component $y_j$, e.g. a series of competitive trials among $n$ subjects with only the winning performance recorded.

$\Omega' = (\Omega'_1, \ldots, \Omega'_p)$ for the $p$-variate GW distribution has joint survivor function

$$S(\omega'_1, \ldots, \omega'_p) = (1 + \sum_{j=1}^{p} \xi_j \omega'_j)^{-\nu},$$

which is again of the GW form with $\nu$ replaced by $n\nu$.

**Moments**

A general formula for joint moments of $(Y_1, \ldots, Y_p)$ is now given. Let $r_1, \ldots, r_p$ be non-negative integers. In terms of the definition of the survivor function of $Y = (Y_1, \ldots, Y_p)$ in (2.2), the moments of $Y_j$, $(j = 1, \ldots, p)$, conditional on $W$ are

$$E(Y_j^{r_j}|W) = \int_0^\infty y_j^{r_j} w^{r_j-1} \exp(-w \xi_j \phi_j) dy_j$$

$$= \xi_j^{-\frac{r_j}{\phi_j}} w^{-\frac{r_j}{\phi_j}} \Gamma(\frac{r_j}{\phi_j} + 1).$$

Therefore, the joint moments of $Y_1, \ldots, Y_p$ are derived from

$$E(Y_1^{r_1} \ldots Y_p^{r_p}) = E\{E(Y_1^{r_1} \ldots Y_p^{r_p}|W)\}$$

$$= E\{\prod_{j=1}^{p} E(Y_j^{r_j}|W)\}$$

$$= \{\prod_{j=1}^{p} \xi_j^{-\frac{r_j}{\phi_j}} \Gamma(\frac{r_j}{\phi_j} + 1)\} \Gamma(\nu - \sum_{j=1}^{p} \frac{r_j}{\phi_j}) \Gamma(\nu).$$ (2.9)

Since $W$ follows a gamma distribution with density function $g(w) = \frac{w^{\nu-1} \exp(-w)}{\Gamma(\nu)}$, we have

$$E(Y_1^{r_1} \ldots Y_p^{r_p}) = \{\prod_{j=1}^{p} \xi_j^{-\frac{r_j}{\phi_j}} \Gamma(\frac{r_j}{\phi_j} + 1)\} \int_0^\infty w^{-\sum_{j=1}^{p} \frac{r_j}{\phi_j}} w^{\nu-1} \exp(-w) \Gamma(\nu) \frac{w^{-\sum_{j=1}^{p} \frac{r_j}{\phi_j}} \Gamma(\nu)}{\Gamma(\nu)} dw$$

$$= \{\prod_{j=1}^{p} \xi_j^{-\frac{r_j}{\phi_j}} \Gamma(\frac{r_j}{\phi_j} + 1)\} \Gamma(\nu - \sum_{j=1}^{p} \frac{r_j}{\phi_j}) \Gamma(\nu).$$ (2.10)
However, the above expression is valid only for \( \nu > \sum_{j=1}^{p} \frac{r_j}{\phi_j} \). Specifically, the mean and variance of \( Y_j \) can be obtained based on the results in (2.10),

\[
E(Y_j) = \nu \xi_j^{-\phi_j^{-1}} \beta(\phi_j^{-1} + 1, \nu - \phi_j^{-1}),
\]

\[
Var(Y_j) = \nu \xi_j^{-2\phi_j^{-1}} \{ \beta(2\phi_j^{-1} + 1, \nu - 2\phi_j^{-1}) - \nu \beta^2(\phi_j^{-1} + 1, \nu - \phi_j^{-1}) \},
\]

where \( \beta(.,.) \) is the Beta function. Similarly, the covariance of \( Y_i \) and \( Y_j \) is given by

\[
Cov(Y_i, Y_j) = \frac{1}{2} \xi_i^{-\phi_i^{-1}} \xi_j^{-\phi_j^{-1}} \{(\nu + 1)\beta(\phi_i^{-1} + 1, \phi_j^{-1} + 1)\beta(\phi_i^{-1} + \phi_j^{-1} + 2,
\]

\[
\nu - \phi_i^{-1} - \phi_j^{-1}) - \nu \beta(\phi_i^{-1} + 1, \nu - \phi_i^{-1})\beta(\phi_j^{-1} + 1, \nu - \phi_j^{-1})\}.
\]

(2.13)

Again, note that the parameter \( \nu \) must satisfy the condition that \( \nu > 2\phi_i^{-1} \) and \( \nu > 2\phi_j^{-1} \).

Since the GW model can be interpreted in terms of the log-times, Crowder(1985) has derived the joint and marginal moments of \( (\log Y_1, \ldots, \log Y_p) \). To avoid complexity, the moments have been derived in terms of the reduced vector \( \varphi \), where \( \varphi_j = -\log \xi_j - \phi_j \log y_j \). Conditional on random effect \( W \) the \( \varphi_j \)'s are independent with distribution functions \( \exp(-We^{-\psi_i}) \) for \( -\infty < \varphi_j < \infty \). Hence their conditional joint moment generating function is

\[
M(z|w) = \prod_{j=1}^{p} E[\exp(z_j \varphi_j)|w] = \prod_{j=1}^{p} \{ w^z_j \Gamma(1 - z_j) \}
\]

(2.14)

and the unconditional joint moment generating function is

\[
M(z) = \{ \prod_{j=1}^{p} \Gamma(1 - z_j) \} E(W^{\sum_{j=1}^{p} z_j})
\]

\[
= \{ \prod_{j=1}^{p} \Gamma(1 - z_j) \} \Gamma(\nu + \sum_{j=1}^{p} z_j)/\Gamma(\nu),
\]

where \( z = (z_1, \ldots, z_p) \). The joint and marginal moments of \( (\varphi_1, \ldots, \varphi_p) \) can be obtained by differentiation of \( M(z) \) at \( z = 0 \). For \( p = 1 \), the first and second derivatives of \( M(z) \) are

\[
M'(z) = \left\{ \Gamma'(1 - z)\Gamma(\nu + z) + \Gamma(1 - z)\Gamma'(\nu + z) \right\}/\Gamma(\nu),
\]

and

\[
M''(z) = \left\{ \Gamma''(1 - z)\Gamma(\nu + z) + 2\Gamma'(1 - z)\Gamma'(\nu + z) + \Gamma(1 - z)\Gamma''(\nu + z) \right\}/\Gamma(\nu).
\]
Therefore, the first and second moments of \( \varphi_j \) are obtained as

\[
E(\varphi_j) = M'(z_j)\big|_{z_j=0} = -\gamma + \psi(\nu),
\]

and

\[
E(\varphi_j^2) = M''(z_j)\big|_{z_j=0} = \gamma^2 + \frac{\pi^2}{6} - 2\gamma\psi(\nu) + \psi^2(\nu) + \psi'(\nu).
\]

Then,

\[
Var(\varphi_j) = \frac{\pi^2}{6} + \psi'(\nu)
\]

Hence, the mean and variance of \( \log Y_j \) may be obtained from the above results. They are

\[
E(\log Y_j) = -\phi_j^{-1}(\log \xi_j - \gamma + \psi(\nu)) \tag{2.15}
\]

\[
Var(\log Y_j) = \phi_j^{-2}\left(\frac{\pi^2}{6} + \psi'(\nu)\right) \tag{2.16}
\]

where \( \psi(z) = \Gamma'(z)/\Gamma(z) \), is the digamma function. Similarly, the covariance of \( \log Y_i \) and \( \log Y_j \) is given by

\[
Cov(\log Y_i, \log Y_j) = \frac{\psi'(\nu)}{\phi_i \phi_j}. \tag{2.17}
\]

### 2.3 Positive stable mixture of Weibulls distribution

Hougaard (1986b) has suggested a positive stable mixture of Weibulls (PSW) distribution. Assuming that random effect \( W \) in (2.4) has a positive stable distribution \( G(.) \) over individuals with characteristic exponent \( \nu \) \( (0 < \nu \leq 1) \), see Feller (1971, Chapter XIII, Section 6), the joint survivor function of the \( p \)-variate PSW distribution is obtained as a Laplace transform

\[
S(y_1, \ldots, y_p) = \int_0^\infty \exp(-ws)G(w) = \exp(-s^\nu), \tag{2.18}
\]

where \( s \) is as defined in (2.3). A justification for choosing the positive stable distribution occurs when the variation of random effect \( W \) arises from many small random contributions acting additively.
2.3.1 Univariate case

For the univariate case, i.e. $p = 1$, the survivor function of $Y$ is

$$S(y) = \exp(-\xi^\nu y^{-\nu \phi}),$$

where $0 < \nu \leq 1$. Therefore, the distribution of $Y$ is another Weibull with parameters $\xi' = \xi^\nu$ and $\phi' = \nu \phi$. That implies that the Weibull distribution may be regarded as a proper mixture of Weibull distributions. However, without additional information, it is impossible to detect heterogeneity of this type in general, purely on the basis of observed data.

The hazard function is of the form

$$h(y) = \xi^\nu \nu \phi y^{\nu \phi - 1}.$$  

The hazard is a DFR for $\nu \phi < 1$; constant for $\nu \phi = 1$; and an IFR for $\nu \phi > 1$. In particular, it is a DFR when $\phi < 1$, since $0 < \nu \leq 1$. It also shows that if we start by assuming an exponential model, i.e. $\phi = 1$, the resulting random effects model is Weibull with shape parameter $\nu$. Hence a Weibull model with decreasing hazard may be regarded as a proper mixture of exponential variables. The density function of $Y$ is

$$f(y) = \xi^\nu \nu \phi y^{\nu \phi - 1} \exp(-\xi^\nu y^{-\nu \phi}).$$

Figure 2.2 shows the density and hazard functions for the univariate PSW distribution for various parameter values.

2.3.2 Multivariate case

Survivor function

Crowder (1989) has extended the multivariate PSW distribution in (2.18) into a generalised form by introducing an extra parameter $\kappa \geq 0$. Its survivor function is defined as

$$S(y_1, \ldots, y_p) = \exp[\kappa^\nu - (\kappa + s)^\nu]. \quad (2.19)$$

When $\kappa = 0$, (2.19) yields (2.18). The non-zero $\kappa$ in (2.19) significantly affects the behaviour of the distribution in some aspects. Crowder (1989) discusses this in
Density function

The density function of the generalised $p$-variate PSW distribution (2.19), is increasingly cumbersome as $p > 2$. The general form of the density is

$$f(y_1, \ldots, y_p) = (-1)^p \partial^p S(y_1, \ldots, y_p) / \partial y_1 \ldots \partial y_p = d_1 \ldots d_p \partial^p S_s(s) / \partial s^p,$$

where $d_j = \partial s / \partial y_j = \phi_j \xi_j \gamma_j^{\phi_j - 1}$ and $S_s(s) = \exp[\kappa^\nu - (\kappa + s)^\nu]$.

Using Leibnitz’s formula, a recurrence relation for the derivatives in (2.20) is obtained as

$$\partial^{p+1} S_s / \partial s^{p+1} = \frac{\partial^p}{\partial s^p} \{[-\nu(\kappa + s)^{\nu-1} S_s] \partial S_s / \partial s \}$$

$$= -\nu \sum_{r=0}^{p} \binom{p}{r} [\partial^r (\kappa + s)^{\nu-1} / \partial s^r] [\partial^{p-r} S_s / \partial s^{p-r}].$$

---

Figure 2.2: Density and hazard functions for univariate PSW distribution, where $\nu = 0.8$, $\xi = 1$, and $\phi$ varies corresponding to 0.5, 1.5, 2.5 and 5.0.

detail. He also points out that, under certain conditions, $\nu$ may take values greater than 1 for $\kappa > 0$, which allows both positive and negative association between $Y_j$.

The parameter $\kappa$ in (2.19) may be interpreted as a type of initial implicit condition in the distribution, see the following discussion in the context of the conditional distribution.
The derivatives of \((\kappa + s)^{\nu-1}\) in the summation may themselves be generated from the recursion

\[ \frac{\partial^r (\kappa + s)^{\nu-1}}{\partial s^r} = [(\nu - r)/(\kappa + s)] \frac{\partial^{r-1}(\kappa + s)^{\nu-1}}{\partial s^{r-1}}, \]  

(2.22)

Such recurrence formulae are convenient for computer programming. In particular, the density functions for \(p = 1, 2\) are respectively

\[ f(y_1) = -\partial S(y_1)/\partial y_1 = d_1 \nu(\kappa + s)^{\nu-1} S_s(s), \]

\[ f(y_1, y_2) = \frac{\partial^2 S(y_1, y_2)}{\partial y_1 \partial y_2} = d_1d_2[\nu^2(\kappa + s)^{2\nu-2} - \nu(\nu - 1)(\kappa + s)^{\nu-2}] S_s(s). \]

In the bivariate case, the density expression is guaranteed to be non-negative when \(\nu \leq 1\). A contour plot for the bivariate case with \(\nu = 0.8\) and \(\kappa = 0\) is shown in Figure 2.4(b).

**Hazard function**

The multivariate failure rate \(q\) for the generalised positive stable mixture of Weibulls distribution corresponding to Brindly & Thompson (1972)'s definition is obtained as

\[ q = P(Y > y + \Delta 1_p | Y > y) = \frac{S(y_1 + \Delta, \ldots, y_p + \Delta)}{S(y_1, \ldots, y_p)} \]

\[ = \exp[\nu - (\kappa + \sum_{j=1}^p \xi_j (y_j + \Delta)^{\phi_j})^\nu] \]

\[ = \exp[\nu - (\kappa + \sum_{j=1}^p \xi_j (y_j + \Delta)^{\phi_j})^\nu]. \]

Thus, \(q\) has log-derivatives

\[ \frac{\partial}{\partial y_k}(\log q) = \nu \xi_k \phi_k [(\kappa + \sum_{j=1}^p \xi_j y_j^{\phi_j})^{\nu-1} y_k^{\phi_k-1} - (\kappa + \sum_{j=1}^p \xi_j (y_j + \Delta)^{\phi_j})^{\nu-1} (y_k + \Delta)^{\phi_k-1}]. \]

This expression is positive for \(\phi_k \leq 1\). Therefore, \(\frac{\partial q}{\partial y_k} > 0\) for \(\phi_k \leq 1\). Hence the \(p\)-variate PSW has DFR for \(\phi_k \leq 1\). Otherwise, it can be neither IFR nor DFR.

Crowder (1989) has discussed the conditional hazard rate according to Johnson & Kotz (1972)'s definition,

\[ h_k(y_k | Y > y) = \nu(\kappa + \sum_{j=1}^p \xi_j y_j^{\phi_j})^{\nu-1} \phi_k \xi_k y_k^{\phi_k-1}, \]
which is decreasing in $y_k$, if $\phi_k \leq 1$.

**Marginal and conditional distributions**

Let $y_A(a \times 1)$ and $y_B(b \times 1)$ be complementary sub-vectors of $y = (y_1, \ldots, y_p)$, so that $a + b = p$. We denote $s_A = \sum_J \xi_j y_j^{\phi_j}$ with summation just over those $y_j$ in $y_A$. By setting $y_B = 0$ in (2.19), the marginal survivor function of $Y_A$ for the generalised $p$-variate PSW distribution is

$$S_A(y_A) = \exp[-(\kappa + s_A)^\nu]. \quad (2.23)$$

Thus $Y_A$ has the same form of the distribution as $Y$ in (2.19). In particular, the marginal distribution of the $y_j$s are each univariate Weibull when $\kappa = 0$ or $\nu = 1$, which illustrates that the distribution in (2.18) is a genuine multivariate Weibull distribution. For general $\kappa$ and $\nu$ the marginal distributions are such that $$(\kappa + s_A)^\nu - \kappa^\nu$$ is exponential with unit mean.

The conditional distribution of $Y_A$ given that $Y_B > y_B$ for the PSW distribution is

$$P(Y_A > y_A|Y_B > y_B) = \frac{S(y)}{S_B(y_B)} = \exp[(\kappa + s_B)^\nu - (\kappa + s)^\nu] \quad (2.24)$$

where $s_B = s - s_A$. This conditional distribution has the same form as the distribution (2.19) with $\kappa$ replaced by $\kappa + s_B$. Crowder (1989) gives an interpretation for $\kappa$ in terms of this conditional distribution. If we take $Y_0$ as an auxiliary component such that $(Y_0, Y)$ has the joint survivor function (2.19), i.e.

$$P(Y_0 > y_0, Y > y) = \exp[-(\xi_0 y_0^{\phi_0} + s)^\nu].$$

Then the conditional probability of $Y$ given that $Y_0 > y_0$ is just (2.19), that is

$$P(Y > y|Y_0 > y_0) = \exp[\kappa^\nu - (\kappa + s)^\nu],$$

where $\kappa = \xi_0 y_0^{\phi_0}$. Therefore, the distribution (2.19) may be regarded as a distribution for $Y = (Y_1, \ldots, Y_p)$ under an implicit initial condition. For example, $Y_j(j = 1, \ldots, p)$ might be repeated measurements from successive stages of treatment only recorded when the patient satisfies the preliminary requirement $Y_0 > y_0$. Similarly, $Y_0$ may represent the minimum quality of acceptable manufactured products.
Distributions of minima

For the generalised PSW distribution, the type one minimum $\Omega$ as described in Section 2.2.2 has survivor function

$$S(\omega) = \exp[-(\kappa + \sum_{j=1}^{p} \phi_j \omega_j)^\nu],$$

which is the univariate case of the generalised PSW distribution (2.19), when $\phi_1 = \ldots = \phi_p$.

The type two minima $\Omega'$ of the generalised PSW have joint survivor function

$$S(\omega'_1, \ldots, \omega'_p) = \exp[n\kappa' - n(\kappa + \sum_{j=1}^{p} \phi_j \omega'_j)^\nu].$$

Thus $\Omega'$ has the distribution (2.19) with $\kappa$ and $\xi_j$ replaced by $\kappa n^{1/\nu}$ and $\xi_j n^{1/\nu}$ respectively.

Moments

Following Williams (1977) and Hougaard (1986b), the moments for random variable $W$ which is from the positive stable distribution with characteristic exponent $\nu$ ($0 < \nu \leq 1$) are given, for $k < \nu$,

$$E(W^k) = \Gamma(1 - k/\nu)/\Gamma(1 - k), \quad (2.25)$$

$$\frac{\partial}{\partial k} E(W^k) = E(W^k \log W) = \frac{\Gamma(1 - k/\nu)}{\Gamma(1 - k)} \{\psi(1 - k) - \frac{\psi(1 - k/\nu)}{\nu}\},$$

where $\psi(.)$ is the digamma function.

The general form of the joint and marginal moments of $Y_1, \ldots, Y_p$ for the PSW distribution (2.18) is, as shown in (2.9),

$$E(Y_1^{r_1}, \ldots, Y_p^{r_p}) = \{\prod_{j=1}^{p} \xi_j^{r_j/\phi_j} \Gamma(r_j/\phi_j + 1)\} E(W^{-r})$$

where $r = \sum_{j=1}^{p} \frac{r_j}{\phi_j}$. Since $-r < \nu$, $E(W^{-r})$ can be obtained from (2.25). In particular, the mean and variance of $Y_j$ are given by

$$E(Y_j) = \xi_j^{-\phi_j^{-1}} \Gamma(1 + \frac{1}{\nu\phi_j}), \quad (2.26)$$
\[ \text{Var}(Y_j) = \xi_j^{-\phi_j^{-2}} \left[ \Gamma(1 + \frac{2}{\nu \phi_j}) - \Gamma^2(1 + \frac{1}{\nu \phi_j}) \right]. \] (2.27)

Similarly, the covariance of \( Y_i \) and \( Y_j \) is obtained as

\[
\text{Cov}(Y_i, Y_j) = \xi_i^{-\phi_i^{-1}} \xi_j^{-\phi_j^{-1}} \left\{ \frac{\Gamma(1 + \frac{1}{\phi_i})\Gamma(1 + \frac{1}{\phi_j})\Gamma(1 + \frac{1}{\nu \phi_i} + \frac{1}{\nu \phi_j})}{\Gamma(1 + \frac{1}{\phi_i} + \frac{1}{\phi_j})} - \Gamma(1 + \frac{1}{\nu \phi_i})\Gamma(1 + \frac{1}{\nu \phi_j}) \right\}. \] (2.28)

The general form of the joint and marginal moments of \( \log Y_1, \ldots, \log Y_p \) for the PSW distribution (2.18) may be obtained from the moment generating function in terms of the vector \( \varphi \), where \( \varphi_j = -\log \xi_j - \phi_j \log y_j \),

\[
M(z) = \left\{ \prod_{j=1}^{p} \Gamma(1 - z_j) \right\} E(W^\sum_{j=1}^{p} z_j) = \frac{\left\{ \prod_{j=1}^{p} \Gamma(1 - z_j) \right\} \Gamma(1 - \sum_{j=1}^{p} z_j)}{\Gamma(1 - \sum_{j=1}^{p} z_j)}. \]

The joint and marginal moments of \( (\varphi_1, \ldots, \varphi_p) \) can be obtained from differentiation of \( M(z) \) at \( z = 0 \). In particular, the mean and variance of \( \varphi_j \) are obtained as

\[
E(\varphi_j) = M'(z_j)|_{z_j=0} = \gamma \nu, \quad E(\varphi_j^2) = M''(z_j)|_{z_j=0} = \frac{1}{\nu^2} \left( \frac{\pi^2}{6} + \gamma^2 \right). \]

Therefore, the mean and variance of \( \log Y_j \) are

\[
E(\log Y_j) = -\frac{1}{\phi_j} (\log \xi_j + \frac{\gamma}{\nu}), \quad \text{and} \quad \text{Var}(\log Y_j) = \frac{\pi^2}{6 \phi_j^2 \nu^2}. \] (2.29)

Similarly, the covariance of \( \log Y_i \) and \( \log Y_j \) may be obtained as

\[
\text{Cov}(\log Y_i, \log Y_j) = \frac{\pi^2}{6 \phi_i \phi_j} \left( \frac{1}{\nu^2} - 1 \right). \] (2.30)
2.4 Inverse Gaussian mixture of Weibulls distribution

Suppose that random effect \( W \) in (2.4) has an inverse Gaussian distribution with positive parameters \( a \) and \( b \), that is, IG\( (\frac{1}{a}, \frac{1}{b}) \). The distribution function \( G(\cdot) \) of \( W \), see Tweedie (1957), may be expressed in the form

\[
G(w) = \Phi\left(\frac{aw - 1}{\sqrt{bw}}\right) + \exp\left(\frac{2a}{b}\right)\Phi\left(-\frac{aw + 1}{\sqrt{bw}}\right).
\]

The density function is given by

\[
g(w) = \frac{1}{\sqrt{2\pi bw^3}} \exp\left\{-\frac{(aw - 1)^2}{2bw}\right\}. \tag{2.31}
\]

The Inverse Gaussian mixture of Weibulls (IGW) distribution is introduced when the random effect \( W \) is assumed to be an inverse Gaussian distribution. The joint survivor function of the IGW distribution is

\[
S(y_1, \ldots, y_p) = \int_0^\infty \exp(-ws) dG(w) = \int_0^\infty \exp(-ws) \frac{1}{\sqrt{2\pi bw^3}} \exp\left\{-\frac{(aw - 1)^2}{2bw}\right\} dw = \exp\left\{\frac{1}{b}(a - \sqrt{a^2 + 2bs})\right\} = \exp(\theta - \sqrt{\theta^2 + s'}), \tag{2.32}
\]

where \( s' = 2bs \), is as given in (2.3), the value \( 2b \) being absorbed by the parameter \( \xi_j \). The positive parameter \( \theta \) is denoted as the ratio of the parameters \( a \) and \( b \), i.e. \( \theta = \frac{a}{b} \). This model is an extended form of the model that Whitmore & Lee (1991) have considered. The motivation for using the inverse Gaussian distribution as a random effect is discussed by Whitmore & Lee (1991). Particularly, note that, (2.32) is a special case of the generalised PSW distribution (2.19), with \( \nu = \frac{1}{2} \) and \( \kappa = \theta^2 \). In the following discussion, \( s' \) in (2.32) is replaced by \( s \).

2.4.1 Univariate case

For the univariate case, \( p = 1 \), the survivor function of \( Y \) is

\[
S(y) = \exp(\theta - \sqrt{\theta^2 + \xi y\phi}),
\]
Figure 2.3: Density and hazard functions for the univariate IGW distribution, where \( \theta = 1.2, \xi = 1, \) and \( \phi \) varies corresponding to 0.5, 1.5, 2.5 and 5.0

where \( \theta, \xi, \phi \) are positive parameters. The hazard function is given by

\[
h(y) = \frac{\xi \phi y^{\phi - 1}}{2\sqrt{\theta^2 + \xi y^\phi}}.
\]

The effect of the parameter \( \theta \) on the hazard function is more complex than that of the parameter \( \nu \) in the Burr hazard function. Also the hazard varies more than the Burr case: \( h(y) \) is decreasing (\( \phi \leq 1 \)), upturned bathtub shaped (\( 1 < \phi < 2 \)), and increasing (\( \phi \geq 2 \)).

The density function is

\[
f(y) = \frac{\xi \phi y^{\phi - 1}}{2\sqrt{\theta^2 + \xi y^\phi}} \exp(\theta - \sqrt{\theta^2 + \xi y^\phi}).
\]

Plots of the density and hazard functions for various values of the parameter \( \phi \) are shown in Figure 2.3.

2.4.2 Multivariate case

Density function

The density function of the \( p \)-variate IGW distribution is in a recursive expression

\[
f(y_1, \ldots, y_p) = \prod_{j=1}^{p} \{\xi_j \phi_j y_j^{\phi_j - 1}\} E[Z(s)^p] \exp(\theta - \sqrt{\theta^2 + s}),
\]
where $Z(s)$ is a random variable that follows an inverse Gaussian distribution with positive parameters $\sqrt{a^2 + 2bs}$ and $b$, i.e. $Z(s) \sim \text{IG}(\frac{1}{\sqrt{a^2 + 2bs}}, \frac{1}{b})$.

The $p$th moment $E[Z(s)^p]$ can be evaluated using general formulae for inverse Gaussian random variables, see the later discussion in this section. A density contour plot of the bivariate IGW distribution for $\theta = 1.2$ is plotted in Figure 2.4(c).
Hazard function

The multivariate failure rate $q$ for IGW distribution corresponding to Brindly & Thompson (1972)'s definition is obtained as

$$q = P(Y > y + \Delta_1 p | Y > y) = \frac{S(y_1 + \Delta, \ldots, y_p + \Delta)}{S(y_1, \ldots, y_p)}$$

$$= \frac{\exp(\theta - \sqrt{\theta^2 + \sum_{j=1}^{p} \xi_j (y_j + \Delta)^{\phi_j}})}{\exp(\theta - \sqrt{\theta^2 + \sum_{j=1}^{p} \xi_j y_j^{\phi_j}})}.$$ 

Then, $q$ has log-derivatives

$$\frac{\partial}{\partial y_j} (\log q) = \frac{1}{2} \xi_k \phi_k \left\{ \frac{y_j^{\phi_k - 1}}{\sqrt{\theta^2 + \sum_{j=1}^{p} \xi_j y_j^{\phi_j}}} - \frac{(y_j + \Delta)^{\phi_k - 1}}{\sqrt{\theta^2 + \sum_{j=1}^{p} \xi_j (y_j + \Delta)^{\phi_j}}} \right\}, \quad (2.33)$$

which is greater than zero for all the $\Delta > 0$, when $\phi_k \leq 1$. Therefore $\frac{\partial q}{\partial y_k} > 0$ when $\phi_k \leq 1$. Hence the $p$-variate IGW has DFR for $\phi_k \leq 1$.

The conditional hazard function for $Y_k$ given $Y > y$ in terms of Johnson & Kotz (1972) is

$$h_k(y_k | Y > y) = \frac{\xi_k \phi_k y_k^{\phi_k - 1}}{2\sqrt{\theta^2 + s}},$$

which also indicates a DFR in $y_k$ when $\phi_k \leq 1$, and non-monotone of failure rate when $\phi_k > 1$.

Marginal and conditional distributions

Let $y_A(a \times 1)$ and $y_B(b \times 1)$ be complementary sub-vectors of $y = (y_1, \ldots, y_p)$, so that $a + b = p$. We denote $s_A = \sum_A \xi_j y_j^{\phi_j}$ with summation just over those $y_j$ in $y_A$. The marginal survivor function of $Y_A$ for the $p$-variate IGW distribution is obtained by setting $y_B = 0$ in (2.32)

$$S_A(y_A) = \exp(\theta - \sqrt{\theta^2 + s_A}).$$

The conditional survivor function of $Y_A$ given $Y_B > y_B$ is

$$P(Y_A > y_A | Y_B > y_B) = \exp(\sqrt{\theta^2 + s_B} - \sqrt{\theta^2 + s_A})$$

$$= \exp(\theta' - \sqrt{\theta'^2 + s_A})$$

where $\theta' = \sqrt{\theta^2 + s_B}$.

Therefore, both of the marginal and conditional distributions have the same forms as the distribution in (2.32).
Distributions of minima

In terms of the definition of the minima in Section 2.2.2, the type one minimum \( \Omega \) of the \( p \)-variate IGW distribution has survivor function

\[
S(\omega) = \exp \left( \theta - \sqrt{\theta^2 + \sum_{j=1}^{p} \xi_j \omega_j^2} \right),
\]

which is in the form of univariate IGW survivor function, when \( \phi_1 = \ldots = \phi_p \).

Similarly, the type two minima \( \Omega' = (\Omega'_1, \ldots, \Omega'_p) \) of the multivariate IGW have joint survivor function

\[
S(\omega'_1, \ldots, \omega'_p) = \exp \left( \theta' - \sqrt{\theta'^2 + \sum_{j=1}^{p} \xi'_j \omega'_j^2} \right).
\]

Again this is of the same form as the joint survivor function in (2.32), where \( \theta' = n\theta \), and \( \xi'_j = n^2 \xi_j \).

Moments

For an inverse Gaussian distributed random variable \( W \) with density function given in (2.31), Tweedie (1957) shows the following useful formulas for positive and negative integer moments.

\[
E(W^k) = \frac{1}{a^k} \sum_{j=1}^{k-1} \frac{(k-1+j)!}{j!(k-1-j)! \left( \frac{2a}{b} \right)^j},
\]

(2.34)

and

\[
E(W^{-k}) = a^{2k+1} E(W^{k+1}).
\]

(2.35)

Furthermore, there is a recurrence relation for the moments of \( W \), which is

\[
E(W^{k+1}) = \frac{1}{a^2} \left[ b(2k-1)E(W^k) + E(W^{k-1}) \right],
\]

(2.36)

where \( E(W^0) = 1 \) and \( E(W) = 1/a \).

The general form of the joint and marginal moments of \( Y_1, \ldots, Y_p \) for the IGW is, as shown in (2.9),

\[
E(Y_1^{r_1}, \ldots, Y_p^{r_p}) = \left\{ \prod_{j=1}^{p} \xi_j^{-r_j} \Gamma \left( \frac{r_j}{\phi_j} + 1 \right) \right\} E(W^{-r})
\]
where \( W \sim IG(1/a, 1/b) \), and \( r = \sum_{j=1}^{p} \frac{r_j}{\phi_j} \). However \( E(W^{-r}) \) can be analytically evaluated from the results in (2.34) – (2.36) only if \( r \) is an integer. Otherwise, it may be evaluated numerically.

In particular, the mean and variance of \( Y_j \) can be obtained as

\[
E(Y_j) = \xi_j^{-\phi_j^{-1}} \Gamma(\phi_j^{-1} + 1) E(W^{-\phi_j^{-1}}),
\]

\[
Var(Y_j) = \xi_j^{-2\phi_j^{-1}} \left\{ \Gamma(2\phi_j^{-1} + 1) E(W^{-2\phi_j^{-1}}) - \Gamma^2(\phi_j^{-1} + 1) E^2(W^{-\phi_j^{-1}}) \right\}. \tag{2.37}
\]

The covariance of \( Y_i \) and \( Y_j \) is

\[
Cov(Y_i, Y_j) = \xi_i^{-\phi_i^{-1}} \xi_j^{-\phi_j^{-1}} \left\{ \Gamma(\phi_i^{-1} + \phi_j^{-1} + 1) E(W^{-\phi_i^{-1} + \phi_j^{-1}}) - \Gamma(\phi_i^{-1} + 1) \Gamma(\phi_j^{-1} + 1) E(W^{-\phi_i^{-1}}) E(W^{-\phi_j^{-1}}) \right\}. \tag{2.38}
\]

For the particular case, when \( \xi_j = \phi_j = 1, (j = 1, \ldots, p) \), the mean and variance of \( Y_j \) are given by Whitmore & Lee (1991). They are

\[
E(Y_j) = a + b,
\]

and

\[
Var(Y_j) = a^2 + 4ab + 5b^2.
\]

When \( \xi_i = \xi_j = \phi_i = \phi_j = 1, (i, j = 1, \ldots, p) \), the covariance of \( Y_i \) and \( Y_j \) is

\[
Cov(Y_i, Y_j) = (a + 2b)b.
\]

According to the results in (2.14), the joint moment generating function in terms of the vector \( \varphi \), where \( \varphi_j = -\log \xi_j - \phi_j \log y_j \), can be expressed as

\[
M(z) = \left\{ \prod_{j=1}^{p} \Gamma(1 - z_j) \right\} E(W^{\sum z_j}),
\]

where \( W \sim IG(1/a, 1/b) \). Since it is difficult to obtain an explicit expression for the derivative of \( M(z) \) at \( z = 0 \), the general expressions of the joint and marginal moments of \( \log Y_1, \ldots, \log Y_p \) have not been found yet. However, they may be evaluated numerically.
Chapter 3

The dependence structure of the models

In some applications of multivariate lifetime analysis, the primary interest might focus on the association between the lifetimes. For example, in genetic epidemiology studies, researchers might wish to learn the association between times to a certain disease of family members (e.g., siblings or twins). In a bone marrow transplant study, the potential concern might be about the connection between the transplant rejection time and the time of occurrence of a type of infection. Social scientists might investigate the relationship between the time of first marriage and time of first divorce. Furthermore, it is desirable to know about the association of correlated failure times, which will help researchers design studies and produce appropriate standard errors for parameter estimates.

The Weibull based random effects models provide a way of modelling association. Different models may lead to quite different dependence structures. In this chapter we investigate and compare the dependence structures of the three Weibull based random effects models discussed earlier by using various association measures. Note that only bivariate models are considered here.
3.1 An expression in terms of the Archimedean distribution

The three bivariate distributions featured here are a subset of the Archimedean copula distributions studied by Genest & Mackey (1986). The generalised form is

\[ S(y_1, y_2) = R \{ R^{-1}[S_1(y_1)] + R^{-1}[S_2(y_2)] \}, \tag{3.1} \]

where \( R(u) \) is any nonnegative decreasing function with \( R(0) = 1 \) and nonnegative second derivative. \( R^{-1}(u) \) is the inverse function of \( R(u) \), and \( S_j(y_j) \), \( (j = 1, 2) \), are the marginal survivor functions of \( Y_1 \) and \( Y_2 \). The three Weibull based random effects models we discuss can be expressed in the form of the Archimedean copula model. For the GW, PSW and IGW distributions, \( R(u) \) is \((1 + u)^{-\nu}, \exp(-u^\nu), \) and \( \exp(\theta - \sqrt{\theta^2 + u}) \), respectively. For example, the marginal survivor function of the PSW distribution is

\[ S_j(y_j) = \exp(-\xi_j y_j^\phi_j)^\nu, \quad j = 1, 2. \]

Therefore, the bivariate joint survivor function of the PSW takes the form

\[ S(y_1, y_2) = \exp[-(\xi_1 y_1^{\phi_1} + \xi_2 y_2^{\phi_2})^\nu] \]
\[ = \exp[- \left\{ [- \log S_1(y_1)]^{\frac{1}{\nu}} + [- \log S_2(y_2)]^{\frac{1}{\nu}} \right\}^\nu], \]

where \( R(u) = \exp(-u^\nu) \) and \( R^{-1}(u) = (-\log u)^{\frac{1}{\nu}} \).

3.2 Measures of association

Generally speaking, a desirable association measure should have the following properties. First, if a pair of lifetimes \((Y_1, Y_2)\) is exchangeable, the association measure should be symmetric with respect to each lifetime. Secondly, to simplify interpretation, it is desirable that the association measure is free from the marginal parameters. For example, the correlation coefficient \( \rho \) is a proper measure summarising the dependence structure of a bivariate normal distribution because it is free from the location and scale parameters of the marginal distributions and determines the correlation structure uniquely.

Two types of association measure are investigated here to describe the degree of association between correlated lifetimes. One type is a global association measure.
that summarises the overall relationship between the variates, such as the corre-
lation coefficient $\rho$ and Kendall's coefficient of concordance $\tau$. The other type
is a local measure that shows the changes of strength of association at the local
level. It is sometimes of prime interest to learn the time-dependent association
structure. In particular, the time of maximum association may be of major con-
cern. For example, in the study of some monozygotic (MZ) and dizygotic (DZ)
twin data (Anderson et al, 1992), researchers wish to learn about the effect of
genetic factors that influence life span between MZ and DZ twins. They believe
that an important genetic impact may exist only in old age. Therefore association
measures indexed by age or time may provide a means of detecting such an effect.
Several time-dependent association measures are discussed later.

3.3 Global association

3.3.1 Correlation coefficients

The correlation coefficient $\rho$ is a commonly used statistic for measuring the linear
association between two variates. The numerical value of $\rho$ lies between $-1$ and $1$.
Values closer to $-1$ or $1$ indicate a high degree of dependence while values closer
to $0$ indicate little linear dependence.

By definition

$$\rho(Y_1, Y_2) = \frac{Cov(Y_1, Y_2)}{\sqrt{Var(Y_1)Var(Y_2)}}. \quad (3.2)$$

Using (2.12) and (2.13), the correlation coefficient $\rho$ for the GW distribution has
the form

$$\rho(Y_1, Y_2) = \frac{(\nu+1)\beta(\phi_1^{-1}+1,\phi_2^{-1}+1)\beta(\phi_1^{-1}+\phi_2^{-1}+2,\nu-\phi_1^{-1}-\phi_2^{-1})}{\prod_{i=1}^{2}\beta(\phi_i^{-1}+\nu-\phi_i^{-1})} \frac{\beta(\phi_1^{-1}+1,\nu-\phi_1^{-1})}{\beta(\phi_2^{-1}+1,\nu-\phi_2^{-1})}^{\frac{1}{2}},$$

where $\beta(.,.)$ is the Beta function. Note that, this expression is applicable only
when the parameter $\nu$ satisfies the condition that $\nu > 2\phi_1^{-1}$ and $\nu > 2\phi_2^{-1}$.

For the special case in which $\phi_1 = \phi_2 = \phi$, and $\nu > 2\phi^{-1}$, $\rho$ becomes

$$\rho(Y_1, Y_2) = \frac{\Gamma(\phi^{-1}+1) - m}{\Gamma(2\phi^{-1}+1) - m},$$
where \( m = \frac{\Gamma^2(\phi^{-1}+1)\Gamma^2(\nu-\phi^{-1})}{\Gamma(\nu-2\phi^{-1})\Gamma(\nu)} \). Figure 3.1(a) shows the relationship between \( \rho(Y_1, Y_2) \) and \( \nu \) in terms of some \( \phi \) values (\( \phi = 0.5, \phi = 1.0, \) and \( \phi = 2.0 \)) for the GW distribution. Furthermore, if \( \phi_1 = \phi_2 = 1 \), the correlation coefficient \( \rho \) becomes \( \rho = \frac{1}{\nu} \). Note that, \( \rho \to 0 \) as \( \nu \to \infty \), as expected since this describes the situation in which \( Y_1 \) and \( Y_2 \) are independent.

According to (3.2) together with (2.27) and (2.28), the correlation coefficient \( \rho \) for the PSW distribution has the form

\[
\rho(Y_1, Y_2) = \frac{\Gamma(\phi^{-1}+1)\Gamma(\phi^{-1}_{2}+1)\Gamma(\phi^{-1}_{2}+\phi^{-1}_1+1)+\Gamma(\phi^{-1}_{2}+\phi^{-1}_1+1)-\Gamma(\phi^{-1}_{1}+\phi^{-1}_1+1)\Gamma(\phi^{-1}_{2}+\phi^{-1}_2+1)}{\sqrt{[\Gamma(\phi^{-1}_{1}+\phi^{-1}_1+1)+\Gamma(\phi^{-1}_{2}+\phi^{-1}_2+1)]\Gamma(\phi^{-1}_{2}+\phi^{-1}_2+1)}}.
\]

For the special case in which \( \phi_1 = \phi_2 = \phi \), \( \rho \) becomes

\[
\rho(Y_1, Y_2) = \frac{\Gamma^2(\phi^{-1}+1)\Gamma(2\phi^{-1}+1)+\Gamma(2\phi^{-1}+1)-\Gamma^2(\phi^{-1}+1)}{\Gamma(2\phi^{-1}+1)-\Gamma^2(\phi^{-1}+1)}.
\]

Figure 3.1(b) shows the relationship between \( \rho(Y_1, Y_2) \) and \( \nu \) for the PSW distribution in the cases in which \( \phi = 0.5, \phi = 1.0, \) and \( \phi = 2.0 \) respectively.

For the more special case in which \( \phi_1 = \phi_2 = 1 \), \( \rho \) becomes

\[
\rho(Y_1, Y_2) = \frac{\Gamma(2\nu^{-1}+1)-2\Gamma^2(\nu^{-1}+1)}{2[\Gamma(2\nu^{-1}+1)-\Gamma^2(\nu^{-1}+1)]}.
\]

\( \rho \) lies between 0 and 1 when \( 0 < \nu < 1 \); \( \rho = 0 \) when \( \nu = 1 \), which indicates independence; \( \rho = 1 \) when \( \nu = 0 \), which indicates complete dependence.

The form of the correlation coefficient \( \rho \) of \( Y_1 \) and \( Y_2 \) for the IGW distribution is intractable in general, although it may be evaluated numerically according to the general expression (2.9) and the moments of the inverse Gaussian distribution.

Assume that \( \phi_1 = \phi_2 = \phi \), the correlation coefficient \( \rho \) of \( Y_1 \) and \( Y_2 \) for the IGW distribution is obtained based on the results from (2.37) and (2.38):

\[
\rho(Y_1, Y_2) = \frac{E(V^2) - E^2(V)}{mE(V^2) - E^2(V)},
\]

where \( V = W^{-\phi^{-1}}, W \sim IG(1/a, 1/b) \), and \( m = \frac{\Gamma(2\phi^{-1}+1)}{\Gamma(\phi^{-1}+1)} \). Figure 3.1(c) shows the relationship between \( \rho(Y_1, Y_2) \) and \( \theta \) for the IGW distribution in the cases in which \( \phi = 0.5, \phi = 1.0 \) and \( \phi = 2.0 \) respectively.

More specifically, when \( \phi_1 = \phi_2 = 1 \),

\[
\rho(Y_1, Y_2) = \frac{\theta + 2}{\theta^2 + 4\theta + 5}.
\]
Figure 3.1: Correlation coefficient $\rho(Y_1, Y_2)$ for the three bivariate Weibull based random effects models, at $\phi_1 = \phi_2 = 0.5, 1.0, 2.0$

where, as defined in Chapter 2, $\theta = \frac{\phi}{\phi'}$. In this case, $\rho$ lies between 0 and $\frac{2}{5}$. Independence ($\rho = 0$) occurs when $\theta \to \infty$; the maximum dependence is $\rho = \frac{2}{5}$ when $\theta = 0$. This result is given by Whitmore & Lee (1991).

The correlation coefficient of the lifetimes $Y_1$ and $Y_2$ have demonstrated the positive dependence structure of the three models. Note that in each case $\rho(Y_1, Y_2)$ involves the Weibull shape parameters. After a log-transformation, the Weibull shape parameter becomes a scale parameter. Since the correlation coefficient is invariant to scale changes, it follows that $\rho(\log Y_1, \log Y_2)$ will not depend on $\phi_1$ and $\phi_2$.

Crowder (1985) gives the correlation coefficient for the GW distribution in terms of log-transformed lifetimes

$$
\rho(\log Y_1, \log Y_2) = \left(1 + \frac{\pi^2}{6\psi'(\nu)}\right)^{-1}.
$$
It increases from 0 to 1 as $\nu$ decreases from $\infty$ to 0.

The correlation between $\log Y_1$ and $\log Y_2$ for the PSW model (2.19) where $\kappa = 0$ is

$$\rho(\log Y_1, \log Y_2) = 1 - \nu^2.$$  

The correlation increases from 0 to 1 as $\nu$ decreases from 1 to 0.

The correlation between $\log Y_1$ and $\log Y_2$ is uniquely determined by the parameter $\nu$ in the GW and PSW cases, which plays the role of an association parameter. Therefore, $\rho(\log Y_1, \log Y_2)$ is a better choice to describe the association structure of the Weibull based random effects models rather than $\rho(Y_1, Y_2)$.

Although the correlation coefficient $\rho$ of $\log Y_1$ and $\log Y_2$ for the IGW distribution may be evaluated numerically, its general expression is intractable because of the reason which is stated in Section 2.4.2.

### 3.3.2 Kendall’s coefficient of concordance

Kendall’s coefficient of concordance, $\tau$, see Kendall (1938), is another useful global association measure. To avoid the association information being dominated by the extreme values of the pairs, Kendall’s $\tau$ uses the contribution from each pair equally by using the rank information instead of the magnitude of the variates. It is invariant to monotone transformations, see Hoeffding (1948). The coefficient of concordance $\tau$ is bounded between $-1$ and 1 and has a similar interpretation to the correlation coefficient.

Genest & Mackay (1986) show that there is a relationship between the Archimedean copula models and Kendall’s $\tau$. $\tau$ is determined by a simple function

$$\tau = 4 \int_0^\infty uR(u)R''(u)du - 1,$$  \hspace{1cm} (3.3)

where $R(u)$ is as in (3.1).

Kendall’s $\tau$ for the corresponding Weibull based random effects models can be calculated easily using (3.3). For the GW distribution, where $R(u) = (1 + u)^{-\nu}$,

$$\tau = 4 \int_0^\infty uR(u)R''(u)du - 1$$
\[
\begin{align*}
\tau &= 4 \int_0^\infty u(1 + u)^{-\nu} \nu(\nu + 1)(1 + u)^{-\nu-2} du - 1 \\
&= \frac{1}{1 + 2\nu}.
\end{align*}
\]

It increases from 0 to 1 as \( \nu \) decreases from \( \infty \) to 0; \( \tau \to 1 \) as \( \nu \to 0 \), which indicates complete dependence between \( Y_1 \) and \( Y_2 \); \( \tau \to 0 \) as \( \nu \to \infty \), which indicates independence of \( Y_1 \) and \( Y_2 \).

For the PSW distribution, where \( R(u) = \exp(-u\nu) \),
\[
\tau = 4 \int_0^\infty uR(u)R'(u)du - 1 \\
= 4 \int_0^\infty ue^{-u\nu}e^{-u\nu} [\nu^2 u^{2\nu-2} - \nu(\nu - 1) u^{\nu-2}] du - 1 \\
= 1 - \nu.
\]

Clearly \( \tau = 0 \) if \( \nu = 1 \), which corresponds to independence of \( Y_1 \) and \( Y_2 \); and \( \tau \to 1 \) if \( \nu \to 0 \), which corresponds to complete dependence.

For the IGW distribution, where \( R(u) = \exp(\theta - \sqrt{\theta^2 + u}) \),
\[
\tau = 4 \int_0^\infty uR(u)R''(u)du - 1 \\
= \int_0^\infty ue^{2\theta} e^{-2\sqrt{\theta^2 + u}} \frac{1}{\theta^2 + u} \left(1 + \frac{1}{\theta^2 + u}\right) du - 1 \\
= \frac{1}{2} - \theta + 2\theta^2 e^{2\theta} E_1(2\theta).
\]

where \( E_1(t) = \int_t^\infty u^{-1} e^{-u} du \). Hence \( \tau \) can be evaluated numerically. As shown in Figure 3.2, it is a decreasing function of \( \theta \) with range from \( \frac{1}{2} \) to 0, as \( \theta \) varies from 0 to \( \infty \). When \( \theta \to 0 \), the IGW model is a special case of the PSW model with \( \nu = 0.5 \). In this case \( \tau = 0.5 \) may be obtained from the both expressions.

### 3.4 Local association

#### 3.4.1 Contour plots

A contour plot of the joint density of \( Y_1 \) and \( Y_2 \) is a useful tool for explaining how the relationship between \( Y_1 \) and \( Y_2 \) changes as their values vary. Density contour plots for the GW, PSW and IGW models are shown in Figure 3.4, where, for illustration, the parameter \( \nu \) or \( \theta \) is chosen to correspond to \( \tau = 0.2 \). They are
Figure 3.2: Kendall’s coefficient $\tau$ vs $\theta$ for the IGW distribution

$\nu = 2.0$ for the GW model, $\nu = 0.8$ for the PSW model, and $\theta = 1.2$ for the IGW model, respectively.

The contour plots show that, when $\tau = 0.2$, the GW and IGW models exhibit similar shapes. The strength of the association between $Y_1$ and $Y_2$ increases with $Y_1$ and $Y_2$. For the PSW model, high dependence occurs in the small values of $Y_1$ and $Y_2$.

### 3.4.2 Cross ratios

Oakes (1989a) introduces a time-dependent association measure which is called the cross-ratio function:

$$r(y_1, y_2) = \frac{SS_{12}}{S_1S_2}$$

where $S = S(y_1, y_2), S_{12} = \partial^2 S(y_1, y_2) / \partial y_1 \partial y_2, \text{ and } S_j = \partial S(y_1, y_2) / \partial y_j$ for $j = 1, 2$. The cross ratio equals 1 if and only if $Y_1$ and $Y_2$ are independent. If $Y_1$ and $Y_2$ are positively associated, then $r(y_1, y_2) > 1$. If $Y_1$ and $Y_2$ are negatively associated, then $0 < r(y_1, y_2) < 1$.

This function may be interpreted as the ratio of the hazard rate of the conditional distribution of $Y_1$ given that $Y_2 = y_2$, to the conditional distribution of $Y_1$ given
that \( Y_2 > y_2 \), because the definition of \( r(y_1, y_2) \) can be represented as

\[
r(y_1, y_2) = \frac{f(y_1|y_2)/S(y_1|y_2)}{f(y_1|Y_2 > y_2)/S(y_1|Y_2 > y_2)} = \frac{h(y_1|Y_2 = y_2)}{h(y_1|Y_2 > y_2)}.
\]

For \( \delta > 0 \), the odds ratio is defined by

\[
OR(y_1, y_2, \delta) = \frac{\text{odds}(Y_2 > y_2 + \delta|Y_1 > y_1 + \delta)}{\text{odds}(Y_2 > y_2 + \delta|Y_1 < y_1 + \delta)}
= \frac{P(Y_2 > y_2 + \delta|Y_1 > y_1 + \delta)}{P(Y_2 > y_2 + \delta|Y_1 < y_1 + \delta)}.
\]

Anderson et al (1992) point out that \( r(y_1, y_2) \) can be interpreted as the instantaneous odds ratio at \( (y_1, y_2) \), that is

\[
\lim_{\delta \to 0} OR(y_1, y_2, \delta) = r(y_1, y_2).
\]

Furthermore, Oakes shows the unique relationship between the bivariate survivor function in Archimedean copula families and \( r(y_1, y_2) \). Cross ratio \( r(y_1, y_2) \) depends on \( (y_1, y_2) \) only through \( S(y_1, y_2) \). That is to say, \( r(y_1, y_2) = r^*(S(y_1, y_2)) \) characterises the Weibull based random effects models and determines the function \( R(u) \) in (3.1) uniquely, up to a scale factor. The cross ratio for the GW distribution is

\[
r(y_1, y_2) = 1 + \frac{1}{\nu}.
\]

It is a constant that does not depend on the times \( (y_1, y_2) \). For the PSW distribution,

\[
r(y_1, y_2) = 1 - s^{-\nu} + \nu^{-1}s^{-\nu}
= 1 + (1 - \nu)/(-\nu \log(S(y_1, y_2)));
\]

where \( s = \xi_1 y_1^{\phi_1} + \xi_2 y_2^{\phi_2} \). For the IGW distribution,

\[
r(y_1, y_2) = 1 + \frac{1}{\sqrt{\theta^2 + s}}
= 1 + \frac{1}{\theta - \log S(y_1, y_2)},
\]

where \( s = \xi_1 y_1^{\phi_1} + \xi_2 y_2^{\phi_2} \). Therefore the cross ratio is a useful measure to represent local association.
Figure 3.3: Contour plots of cross ratio $r(y_1, y_2)$ for the bivariate PSW and IGW distributions, at $\xi_1 = \xi_2 = \phi_1 = \phi_2 = 1$, and $\tau = 0.2$.

Figure 3.4: Plot of cross ratio $r(y_1, y_2)$ for the GW, PSW and IGW distributions, at $y_1 = y_2 = y$, $\xi_1 = \xi_2 = \phi_1 = \phi_2 = 1$, and $\tau = 0.2$. 

53
Figure 3.3 presents contour plots of the cross ratio $r(y_1, y_2)$ for the PSW and IGW distributions, where the Weibull parameters $\xi_1 = \xi_2 = \phi_1 = \phi_2 = 1$, and $\nu$ or $\theta$ in the models is chosen to correspond to Kendall’s $\tau$ equal to 0.2. It shows that, for the PSW distribution, the high dependence exists in smaller values of $y_1$ and $y_2$, while the dependence drops down quickly when either $y_1$ or $y_2$ increases. The contour plot of the cross ratio for the IGW distribution is much flatter than that of the PSW distribution. The cross ratio for the IGW distribution decreases as both $y_1$ and $y_2$ increase. However, the rate of decrease is much slower than that in the PSW case.

The cross ratio plot for the three distributions when $y_1 = y_2 = y$ is shown in Figure 3.4. It gives a profile of the changes of the cross ratio for the three distributions as $y_1$ and $y_2$ move together. The cross ratio is independent of the $y$ values for the GW model, and is almost a constant as $y$ varies for the IGW model. However, for the PSW model, the cross ratio gives a very high positive dependence when $y$ is very small whereas it becomes very close to 1 as $y$ get larger. Figure 3.5 shows a cross ratio plot for the three distributions when $y_1 = \frac{1}{y_2} = y$, which gives an insight into the change of the cross ratio when the components $y_1$ and $y_2$ change in opposite directions. For the PSW and IGW models, the cross ratio increases when one component $y_1$ increases and the other component $y_2$ decreases, reaches a
maximum at \( y_1 = 1.0 \), and then decreases slowly. Note that in this case the cross ratio for the PSW model does not rise much above 1.0.

### 3.4.3 Time-dependent correlation coefficient

In order to separate aspects of the shapes of the marginal distributions of \( Y_1 \) and \( Y_2 \) from the association measure, Prentice & Cai (1992) have considered expressing association by means of covariance between the cumulative hazard variate of \( Y_1 \) and \( Y_2 \). By denoting

\[
N_j(y_j) = \begin{cases} 
-\Lambda_j(y_j) & \text{if } Y_j > y_j \\
1 - \Lambda_j(y_j) & \text{if } Y_j \leq y_j 
\end{cases} \quad (3.4)
\]

where \( \Lambda_j(y_j) = -\log S_j(y_j) \), \( (j = 1, 2) \), is the marginal cumulative hazard function, a covariance function for \( Y_1 \) and \( Y_2 \) can be expressed as

\[
\text{Cov}(N_1(y_1), N_2(y_2)) = S(y_1, y_2) - 1 + \int_0^{y_1} S(t_1, y_2) \Lambda_1(dt_1) + \int_0^{y_2} S(y_1, t_2) \Lambda_2(dt_2) + \int_0^{y_1} \int_0^{y_2} S(t_1, t_2) \Lambda_1(dt_1) \Lambda_2(dt_2).
\]

Therefore, a time-dependent correlation coefficient can be specified as, see Shih & Louis (1995b),

\[
\lambda(y_1, y_2) = \frac{\text{Cov}(N_1(y_1), N_2(y_2))}{\sqrt{\text{Var}(N_1(y_1))\text{Var}(N_2(y_2))}}.
\]

The correlation coefficient \( \lambda(y_1, y_2) \) is an appealing association measure because it captures the nature of the dependence between \( Y_1 \) and \( Y_2 \), given their corresponding marginal distribution. It is bounded between -1 and 1 so that it gives an explicit interpretation of the strength of association. In terms of the definition of \( N_j(y_j) \) given in (3.4), the variance of \( N_j(y_j) \) is

\[
\text{Var}(N_j(y_j)) = 1 - S_j(y_j).
\]

To display the different association structure of the three Weibull based random effects models, the contour plots of \( \lambda(y_1, y_2) \) are shown in Figure 3.6, where the Weibull parameters \( \xi_1 = \xi_2 = \phi_1 = \phi_2 = 1 \) and \( \nu \) or \( \theta \) are chosen to correspond to Kendall’s tau equal to 0.2. It shows that, for both the GW and the IGW distributions, \( \lambda(y_1, y_2) \) increases when both \( y_1 \) and \( y_2 \) increase, and the rate of increase decreases as \( y_1 \) and \( y_2 \) increase. The increase of the correlation according
Figure 3.6: Correlation of \((N_1(y_1), N_2(y_2))\) for three bivariate Weibull based random effects models, at \(\xi_1 = \xi_2 = \phi_1 = \phi_2 = 1\) and \(\tau = 0.2\)
Figure 3.7: Correlation of \((N_1(y_1), N_2(y_2))\) for the IGW distribution, at \(\xi_1 = \xi_2 = \phi_1 = \phi_2 = 1\) and \(\theta = 10.0, 0.1\)

to \(y_1\) and \(y_2\) is slightly quicker in the GW distribution than that in the IGW distribution. Meanwhile, for the PSW distribution, a high correlation only occurs at small values of \(y_1\) and \(y_2\), and \(\lambda(y_1, y_2)\) slowly decreases as \(y_1\) and \(y_2\) increase.

Further exploration of the contour plots of \(\lambda(y_1, y_2)\) indicates that qualitatively the structure of \(\lambda(y_1, y_2)\) remains similar to that shown in Figure 3.6 for the GW and the PSW distribution when the parameter \(\nu\) changes. However, \(\lambda(y_1, y_2)\) exhibits quite different patterns as parameter \(\theta\) varies in the IGW case. The correlation increases as both \(y_1\) and \(y_2\) increase at larger and smaller values of \(y_1\) and \(y_2\) in the case that \(\theta\) has a larger value, see Figure 3.7(a). The correlation decreases as both \(y_1\) and \(y_2\) increase for one of the \((y_1, y_2)\) has larger values and the other has smaller values in the case that \(\theta\) is smaller (close to 0), see Figure 3.7(b). When \(\theta\) tends to 0, the contour plot tends to the pattern of the PSW case.

Figure 3.8 plots the correlation curve \(\lambda(y_1, y_2)\) when \(y_1 = y_2 = y\). It characterises the distinctive correlation feature in each of the three distributions, although it does not exhibit the full association information that the contour plot does. For example, the plot shows that the correlation for the PSW model decreases as the survival times of the both components increase whereas the correlations for the GW and IGW model increase as the survival times of the both components increase. The rate of the increase is higher for the GW model than that of the IGW model at the beginning of the time, and then becomes similar as time increases. The correlation for the PSW model is higher than those for the GW and IGW
Figure 3.8: Correlation curve for three bivariate Weibull based random effects models, at $y_1 = y_2 = y, \xi_1 = \xi_2 = \phi_1 = \phi_2 = 1$, and $\tau = 0.2$

Figure 3.9: Correlation curve for three bivariate Weibull based random effects models, at $y_1 = \frac{1}{y_2} = y, \xi_1 = \xi_2 = \phi_1 = \phi_2 = 1$, and $\tau = 0.2$
models when \( y \) is smaller (\(< 0.6\)); it is between the GW and IGW models when \( y \) is within (0.6, 2.0); it is less than those for the GW and IGW models when \( y \) is larger (> 2).

Figure 3.9 displays the correlation curve \( \lambda(y_1, y_2) \) when \( y_1 = \frac{1}{y_2} = y \). It shows that, for the three models, the changes of the correlation \( \lambda(y_1, y_2) \) with respect to the time \( y_1 \) and \( y_2 \) follow similar pattern when \( y_1 \) increases and \( y_2 \) decreases. However, the correlation for the GW model is higher than those of the PSW and IGW models while the correlation for the IGW model is lower than those of the GW and the PSW models when \( y_1 = \frac{1}{y_2} \).

### 3.4.4 Conditional expected residual life

The conditional expected residual life also describes association between two lifetime variables. Anderson et al (1992) define this measure by

\[
\psi(y_1, y_2) = \frac{E(Y_1|Y_1 > y_1, Y_2 > y_2) - y_1}{E(Y_1|Y_1 > y_1) - y_1}.
\]

The numerator is the expected residual life of \( Y_1 \), given \( Y_1 > y_1 \) and \( Y_2 > y_2 \), that is, the survival time expectancy for \( Y_1 \) beyond \( y_1 \) given that \( Y_1 > y_1 \) and \( Y_2 > y_2 \). The denominator is the expected residual life for \( Y_1 \) given that \( Y_1 > y_1 \). This measure interprets how the information about \( Y_2 > y_2 \) influences the expectation of \( Y_1 \). For example, in a twin study or a study of lifetimes of a parent and a child, the conditional expected residual life provides an appropriate summary of association for demographic and actuarial analysis designed to predict longevity for individuals and small groups.

The conditional expected residual life \( \psi(y_1, y_2) \) is a time-dependent association measure. However, it is not symmetric in \( y_1 \) and \( y_2 \) in general. Values of \( \psi \) very different from 1 indicate strong influence from \( Y_2 \), and therefore, strong association between \( Y_1 \) and \( Y_2 \). If \( Y_1 \) and \( Y_2 \) are positively associated, then \( \psi(y_1, y_2) \) increases as \( y_2 \) increases.

The conditional expected residual life \( \psi(y_1, y_2) \) for the three Weibull based random effects models can be calculated numerically from the expression

\[
\psi(y_1, y_2) = \frac{\int_{y_1}^{\infty} \int_{y_2}^{\infty} S(y_2|w)S(u|w)dG(w)du/S(y_1, y_2)}{\int_{y_1}^{\infty} \int_{y_2}^{\infty} S(u|w)dG(w)du/S_1(y_1)}
\]
Figure 3.10: Conditional expected residual life $\psi(y_1, y_2)$ for three bivariate Weibull based random effects models, at $y_1 = y_2 = y$, $\xi_1 = \xi_2 = \phi_1 = \phi_2 = 1$, and $\tau = 0.2$

Figure 3.11: Conditional expected residual life $\psi(y_1, y_2)$ for three bivariate Weibull based random effects models, at $y_1 = \frac{1}{y_2} = y$, $\xi_1 = \xi_2 = \phi_1 = \phi_2 = 1$, and $\tau = 0.2$
Figure 3.10 gives a plot of the conditional expected residual life \( \psi(y_1, y_2) \) for the three Weibull based random effect models in the case that \( y_1 = y_2 = y \). The conditional expected residual life \( \psi(y_1, y_2) \) increases when both of the components \( y_1 \) and \( y_2 \) increase. The rate of the increase is much faster for the GW model than those of the PSW and IGW models. The change with respect to the time \( y \) is similar for the PSW and IGW models. In contrast, Figure 3.11 shows the conditional expected residual life \( \psi(y_1, y_2) \) decreases when one of the components \( y_1 \) increases and the other component \( y_2 \) decreases. Furthermore, the rates of decrease are very fast when \( y_1 < 1 \) for both the PSW and IGW models, and then stabilise. However, for the GW model, the \( \psi(y_1, y_2) \) is nearly a constant when \( y_1 \) and \( y_2 \) change in the opposite directions.

### 3.4.5 Conditional probability

A generalised form of the definition of dependence for bivariate variables given by Lehmann (1966) is another measure to describe the time-dependent association between two lifetimes. This measure is defined as

\[
 c(y_1, y_2) = \frac{S(y_1, y_2)}{S_1(y_1)S_2(y_2)},
\]

where \( S_j(y_j), j = 1, 2, \) are the marginal distributions. It is the ratio of the conditional probability of \( Y_1 > y_1 \) given that \( Y_2 > y_2 \) to the unconditional probability of \( Y_1 > y_1 \). It is a symmetric function. That is

\[
 c(y_1, y_2) = \frac{P(Y_1 > y_1|Y_2 > y_2)}{P(Y_1 > y_1)} = \frac{P(Y_2 > y_2|Y_1 > y_1)}{P(Y_2 > y_2)}
\]

It is said that \( Y_1 \) and \( Y_2 \) are positively (negatively) dependent if the ratio is greater (less) than unity. If the ratio is 1, then \( y_1 \) and \( y_2 \) are independent. An example of using conditional probability as an association parameter can be the study of breast cancer in women. The conditional probability describes the dependence between age at breast cancer diagnosis and age at birth of the first child. This dependence summary could assist in family planning decisions.
In terms of the above definition, it may be shown that there is positive dependence for the variates from the GW distribution. That is because

\[
c(y_1, y_2) = \left\{ \frac{1 + \sum_{j=1}^{2} \xi_j y_j^{\rho_j}}{\prod_{j=1}^{2} (1 + \xi_j y_j^{\rho_j})} \right\}^{-\nu} > 1, \quad \text{for all } \nu > 0.
\]

Crowder (1989) shows that, for the generalised PSW distribution (2.19), there is positive dependence for \( \nu < 1 \), negative dependence for \( \nu > 1 \), and independence for \( \nu = 1 \). The ratio is obtained from

\[
c(y_1, y_2) = \exp\left[ \sum_{j=1}^{2} (\kappa + \xi_j y_j^{\rho_j})^\nu - (\kappa + \sum_{j=1}^{2} \xi_j y_j^{\rho_j})^\nu - \kappa^\nu \right].
\]

When \( \kappa = 0 \), the above expression becomes

\[
c(y_1, y_2) = \exp\left[ \sum_{j=1}^{2} \xi_j y_j^{\rho_j}^\nu - (\sum_{j=1}^{2} \xi_j y_j^{\rho_j})^\nu \right] > 1, \quad \text{for all } 0 < \nu < 1,
\]

which is the case of the PSW model. Hence, \( Y_1 \) and \( Y_2 \) are positively dependent for the PSW model.

The ratio for the IGW distribution is

\[
c(y_1, y_2) = \exp\left\{ \sum_{j=1}^{2} \sqrt{\theta^2 + \xi_j y_j^{\rho_j}} - \sqrt{\theta^2 + \sum_{j=1}^{2} \xi_j y_j^{\rho_j}} \right\} > 1, \quad \text{for all } \theta > 0.
\]

Thus, there is positive dependence for all \( \theta > 0 \). The ratio tends to 1 as \( \theta \to \infty \), which is the case of independence for the IGW distribution.

It is clear that the results from the conditional probability measure agree with all the previous measures discussed.

### 3.5 Summary

Correlation coefficient \( \rho \) and Kendall’s coefficient of concordance \( \tau \) are useful global association measures summarising the dependence structures in the three Weibull based random effects models. The correlation coefficient between log-transformed lifetimes, i.e. \( \rho(\log Y_1, \log Y_2) \), is preferred to the correlation coefficient between the lifetimes, i.e. \( \rho(Y_1, Y_2) \), because it only depends on the association parameter \( \nu \) in the GW and PSW models. The relationship between Kendall’s \( \tau \) and the association parameters \( \nu \) or \( \theta \) are explicit, as shown in Section 3.3.2.
Various local association measures on the three Weibull based random effects models are discussed, which include contour plots, cross ratio, time-dependent correlation coefficient, conditional expected residual life and conditional probability. Each of them has its own specific explanation on the changes of strength of association with regards to the changes of lifetimes. For the GW model, the cross ratio does not change with the times $y_1$ and $y_2$. Its time dependent correlation coefficient increases when both $y_1$ and $y_2$ increase, so does its conditional expected residual life. High dependence exists in bigger values of $y_1$ and $y_2$. However, the cross ratio, time dependent correlation coefficient and conditional expected residual life for the PSW model show that high dependence in the model exists in smaller values of $y_1$ and $y_2$. The local dependence exhibits different structures as parameter $\theta$ varies in the IGW model. It is the same as that of the PSW model when $\theta = 0$; it is similar to that of the GW model when $\theta = 1.2$.

All these local measures confirm the positive dependence in the GW, PSW and IGW models.
Chapter 4

Estimation methods for the PSW model

In this chapter, a two stage marginal approach to estimating the parameters of the PSW model is introduced. Asymptotic properties of the method are investigated and compared with those of the corresponding maximum likelihood estimators. The finite sample performance of the estimators obtained by these methods are then studied by simulation. Throughout this chapter, we only consider the bivariate model in detail. The methodology in principle is applicable to the more general case where \( p > 2 \), and is discussed briefly. The method is illustrated on three data sets.

4.1 Introduction

According to the definition in Chapter 2 (2.18), the bivariate PSW model has joint survivor function

\[
S(y_1, y_2) = \exp[-(\xi_1 y_1^{\phi_1} + \xi_2 y_2^{\phi_2})^\nu]
\]  

(4.1)

One of the most attractive properties of the model is that the marginal distribution of this model is Weibull, however with the parameters \((\xi_j, \phi_j)\) in the conditional model replaced by new parameters \((\xi_j^\nu, \nu \phi_j)\), where \( j = 1, 2 \). Therefore, the parameter \( \nu \), which plays an important role in the association of the variables, is involved in the marginal distribution in this parameterisation. To ensure that the marginal distributions do not involve the association parameter \( \nu \), we re-parameterise the model (4.1) by letting \( \alpha_j = \xi_j^\nu, \beta_j = \nu \phi_j, \; j = 1, 2 \), so that the
re-parametrised model becomes

\[ S(y_1, y_2) = \exp[-(\alpha_1^\nu y_1^\nu + \alpha_2^\nu y_2^\nu)^{\nu}] \]  \hspace{1cm} (4.2)

The log-likelihood function of model (4.2) for a sample with \( n \) observations is

\[
l_n = -n \sum_{i=1}^{n} s_i^\nu + n \frac{\nu}{\nu} (\log \alpha_1 + \log \alpha_2) + n (\log \beta_1 + \log \beta_2 - \log \nu) + \left( \frac{\beta_1}{\nu} - 1 \right) \sum_{i=1}^{n} \log y_{i1} + \left( \frac{\beta_2}{\nu} - 1 \right) \sum_{i=1}^{n} \log y_{i2} + (\nu - 2) \sum_{i=1}^{n} \log s_i + \sum_{i=1}^{n} \log \{ \nu s_i^\nu - \nu + 1 \} \tag{4.3}
\]

where \( s_i = \frac{1}{\alpha_1^\nu y_{i1}^\nu} + \frac{1}{\alpha_2^\nu y_{i2}^\nu} \).

More generally, the log-likelihood function is, when there is right censoring

\[
l_n = -n \sum_{i=1}^{n} s_i^\nu + n \sum_{i=1}^{n} \delta_{ij} \left[ \frac{1}{\nu} \log \alpha_j + \log \beta_j + \left( \frac{\beta_j}{\nu} - 1 \right) \log y_{ij} \right] + \sum_{i=1}^{n} (\delta_{i1} + \delta_{i2})(\nu - 1) \log s_i + \sum_{i=1}^{n} \delta_{i1} \delta_{i2} [\log(\nu s_i^\nu - \nu + 1) - \nu \log s_i - \log \nu], \tag{4.4}
\]

where \( \delta_{ij} \) is an indicator variable, such that \( \delta_{ij} = 1 \) when \( y_{ij} \) is uncensored, and \( \delta_{ij} = 0 \) when \( y_{ij} \) is censored, \( i = 1, \ldots, n, j = 1, 2 \).

The advantage of this representation is that the margins do not depend on the dependency parameter. In other words, we can model and estimate the dependency and the margins separately. Meanwhile, with this representation, explicit formulae for the elements of the Fisher information matrix may be derived, see Oakes & Manatunga (1992).

### 4.2 Maximum likelihood estimator and its asymptotic variance

In this section we illustrate the method of maximum likelihood estimation for model (4.2) and present the corresponding asymptotic variance results. The non-regular behaviour of the maximum likelihood estimator of \( \nu \) at \( \nu = 1 \) is discussed.
in Chapter 6. At present, we restrict our discussion to the model with $0 < \nu < 1$. In addition, the following discussion is in the context of the no censoring case only.

The standard approach to estimating the parameters of the model is the method of maximum likelihood estimation (MLE), that is, maximising the full log-likelihood function in (4.3) or (4.4) with respect to all the parameters $\alpha_1, \beta_1, \alpha_2, \beta_2, \nu$.

Assuming that there is no censoring, the first derivatives of the log-likelihood for a sample of $n$ observations with respect to the parameters $\alpha_1, \beta_1, \alpha_2, \beta_2, \nu$ are

\[
\frac{\partial l_n}{\partial \alpha_j} = \alpha_j^{\alpha_j - 1} \left\{ \sum_{i=1}^{n} \frac{\beta_j}{y_{ij}^{\beta_j}} s_i^{\nu s_i^{\nu} - \nu + 1} - s_i^{\nu} + 1 - \frac{2}{\nu} \frac{\nu - \alpha_j}{\alpha_j} \right\},
\]

\[
\frac{\partial l_n}{\partial \beta_j} = \sum_{i=1}^{n} \alpha_j^{\frac{1}{\beta_j}} y_{ij}^{\frac{1}{\beta_j}} \log y_{ij} s_i^{\nu s_i^{\nu} - \nu + 1} - s_i^{\nu} + 1 - \frac{2}{\nu},
\]

\[
\frac{\partial l_n}{\partial \nu} = \sum_{i=1}^{n} (1 - s_i^{\nu}) \log s_i - \frac{1}{\nu^2} \sum_{j=1}^{n} (n \log \alpha_j - \beta_j \sum_{i=1}^{n} \log y_{ij})
\]

\[
+ \sum_{i=1}^{n} (\nu s_i^{\nu} - \nu + 1)^{-1} (\nu s_i^{\nu} \log s_i + s_i^{\nu} - 1) - \frac{n}{\nu},
\]

where $j = 1, 2$. The maximum likelihood estimators of $\alpha_1, \beta_1, \alpha_2, \beta_2, \nu$, i.e. $(\hat{\alpha}_1, \hat{\beta}_1, \hat{\alpha}_2, \hat{\beta}_2, \hat{\nu})$, can be obtained by simultaneously solving $\frac{\partial l_n}{\partial \alpha_j} = 0$, $\frac{\partial l_n}{\partial \beta_j} = 0$, and $\frac{\partial l_n}{\partial \nu} = 0$. Some iterative scheme, such as the Newton or quasi-Newton algorithms, must be employed to obtain numerical results.

Let $l$ be the log-likelihood function for a single observation, the Fisher information matrix for a single observation is defined as

\[
I = \begin{pmatrix}
I_{11} & I_{12} & I_{13} \\
I_{21} & I_{22} & I_{23} \\
I_{31} & I_{32} & I_{33}
\end{pmatrix}
\]

where

\[
I_{ij} = \begin{pmatrix}
-E(\frac{\partial^2 l}{\partial \alpha_i \partial \alpha_j}) & -E(\frac{\partial^2 l}{\partial \alpha_i \partial \beta_j}) \\
-E(\frac{\partial^2 l}{\partial \beta_i \partial \alpha_j}) & -E(\frac{\partial^2 l}{\partial \beta_i \partial \beta_j})
\end{pmatrix}, \quad i, j = 1, 2
\]

\[
I_{i3}^T = I_{3i} = -E(\frac{\partial^2 l}{\partial \alpha_i \partial \nu}), -E(\frac{\partial^2 l}{\partial \beta_i \partial \nu}), \quad i = 1, 2
\]

and

\[
I_{33} = -E(\frac{\partial^2 l}{\partial \nu^2}).
\]
The Fisher information matrix $I$ for our model may be obtained on the basis of the results of Oakes & Manatunga (1992). The elements of the negative of the Fisher information matrix for the five parameters $\alpha_1, \beta_1, \alpha_2, \beta_2$, and $\nu$ are as follows

$$
\begin{align*}
E\left(\frac{\partial^2 l}{\partial \alpha_i^2}\right) &= \frac{1}{\alpha_i^2} M_{11}, \\
E\left(\frac{\partial^2 l}{\partial \alpha_i \partial \beta_i}\right) &= \frac{1}{\alpha_i \beta_i} (M_{12} - \log \alpha_i M_{11}), \\
E\left(\frac{\partial^2 l}{\partial \beta_i^2}\right) &= \frac{1}{\beta_i^2} (M_{22} - 2 \log \alpha_i M_{12} + \log^2 \alpha_i M_{11}), \\
E\left(\frac{\partial^2 l}{\partial \alpha_i \partial \nu}\right) &= M_{15}, \\
E\left(\frac{\partial^2 l}{\partial \beta_i \partial \nu}\right) &= \frac{1}{\beta_i} (M_{25} - \log \alpha_i M_{15}), \\
E\left(\frac{\partial^2 l}{\partial \alpha_1 \partial \alpha_2}\right) &= \frac{1}{\alpha_1 \alpha_2} M_{13}, \\
E\left(\frac{\partial^2 l}{\partial \alpha_1 \partial \beta_2}\right) &= \frac{1}{\alpha_1 \beta_2} (M_{14} - \log \alpha_2 M_{13}), \\
E\left(\frac{\partial^2 l}{\partial \alpha_2 \partial \beta_1}\right) &= \frac{1}{\alpha_2 \beta_1} (M_{23} - \log \alpha_1 M_{13}), \\
E\left(\frac{\partial^2 l}{\partial \beta_1 \partial \beta_2}\right) &= \frac{1}{\beta_1 \beta_2} (\log \alpha_1 \log \alpha_2 M_{13} - \log (\alpha_1 \alpha_2) M_{23} + M_{24}), \\
E\left(\frac{\partial^2 l}{\partial \nu^2}\right) &= M_{55},
\end{align*}
$$

where the $M_{ij}$ are the entries of the $5 \times 5$ matrix $M$, $(i, j = 1, \ldots, 5)$, satisfying

$$
\begin{align*}
M_{11} &= M_{33} = -\frac{1}{3}\{\Phi^2 + 2\nu + (\Phi - 1)^2 \nu e^{\Phi - 1} E_1(\Phi - 1)\}, \\
M_{12} &= M_{34} = \frac{1}{3}\gamma \Phi^2 - \frac{2}{9} \Phi - \frac{1}{3} - \frac{2}{3} \gamma (1 - \gamma) + \frac{2}{9} \nu^2 + \frac{1}{9} (1 - \nu)^2 e^{\Phi - 1} E_1(\Phi - 1) \\
&\quad - \frac{1}{3} (1 - \nu)^2 J_1(\nu), \\
M_{13} &= \frac{1}{3} \Phi^2 - \frac{1}{3} \nu - \frac{1}{6} \nu (\Phi - 1)^2 e^{(\Phi - 1)} E_1(\Phi - 1), \\
M_{14} &= M_{23} = -\frac{1}{3} \gamma \Phi^2 + \frac{2}{9} \Phi - \frac{1}{3} - \frac{1}{3} (1 - \gamma) \nu + \frac{5}{18} \nu^2 + \frac{5}{36} (1 - \nu)^2 e^{\Phi - 1} E_1(\Phi - 1) \\
&\quad - \frac{1}{6} (1 - \nu)^2 J_1(\nu), \\
M_{15} &= M_{35} = -\frac{1}{4} \Phi \{2\nu^2 - 2 + (\Phi - 1)(\nu - \nu^2 + 2) e^{\Phi - 1} E_1(\Phi - 1)\}, \\
M_{22} &= M_{44} = -\frac{4}{27} \nu^3 + \frac{4}{9} \nu^2 (1 - \gamma) - \frac{2}{27} \nu (1 - \nu)^2 e^{\Phi - 1} E_1(\Phi - 1) \\
&\quad - \frac{2}{9} \nu (3\eta + 3\gamma^2 - 6\gamma + 2) + \left(\frac{2}{3} \gamma - \frac{23}{27}\right) + \frac{4}{9} \Phi \gamma - \frac{1}{3} \Phi^2 (\gamma^2 + \eta) \\
&\quad + \frac{2}{9} \nu (1 - \nu)^2 J_1(\nu) - \frac{1}{3} (1 - \nu)^2 J_2(\nu),
\end{align*}
$$
\[
M_{24} = \frac{1}{3} \Phi^2(\eta + \gamma^2) - \frac{4}{9} \Phi \gamma + (\gamma - \frac{\eta}{3} + \frac{5}{27}) + \nu\{-\frac{1}{3}(\eta + \gamma^2) + \frac{2}{3}\gamma - \frac{1}{18}\} \\
+ \frac{5}{9} \nu^2(-\gamma + 1) + \nu^2(\frac{\eta}{3} - \frac{37}{54}) - \nu(1 - \nu^2)(\frac{37}{108} - \frac{\eta}{6})e^{\Phi - 1}E_1(\Phi - 1) \\
+ \frac{5}{18} \nu(1 - \nu)^2J_1(\nu) - \frac{1}{6}(1 - \nu)^2J_2(\nu),
\]

\[
M_{25} = M_{45} = \frac{1}{3} \Phi(\eta + \frac{1}{2}) - \frac{1}{2} + \nu(\frac{\gamma}{2} - \frac{1}{2}) + \frac{1}{3} \nu^2(\frac{5}{2} - \eta) - \frac{1}{2} \Phi \gamma \\
+(1 - \nu)e^{\Phi - 1}E_1(\Phi - 1)\{\frac{1}{4} \Phi + \frac{1}{6}(1 - \nu)(\frac{5}{2} - \eta)\} \\
- \frac{1}{4}(\Phi - 1)(2 + \nu - \nu^2)J_1(\nu),
\]

\[
M_{55} = \Phi^2(-\frac{2}{3} + \frac{\pi^2}{9}) + \Phi - \frac{1}{3}\Phi(5 - 2\eta) - \{\Phi^3 + \Phi^2 - \frac{\Phi}{3}(\eta + \frac{1}{2})\nu \\
- \frac{1}{3}(5 - 2\eta) + \frac{5 - 2\eta}{6\Phi}\}E_1(\Phi - 1)e^{\Phi - 1},
\]

where, \( \Phi = \frac{1}{\nu} \), and \( \gamma = -\psi(1) \) is Euler’s constant, \( \eta = \psi'(1) = \pi^2/6 \), \( E_1(x) = \int_x^\infty u^{-1}e^{-u}du \),

\[
J_k(\nu) = \int_0^\infty e^{-x}(\log x)^k/(1 - \nu + \nu x)dx \quad \text{for} \quad k = 1, 2
\]

For the case in which \( \alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 1 \), the elements of the Fisher information matrix are equivalent to \( M_{ij} \), \( (i, j = 1, \ldots, 5) \), i.e. \( I = -M \).

For the case in which \( \nu = 1 \), which corresponds to independence of \( Y_1 \) and \( Y_2 \), each item of the negative of the Fisher information matrix for a single observation is

\[
E(\frac{\partial^2 l}{\partial \alpha_i^2}) = -\frac{1}{\alpha_i^2}, \quad (4.5)
\]

\[
E(\frac{\partial^2 l}{\partial \alpha_i \partial \beta_i}) = -\frac{1}{\alpha_i \beta_i}(1 - \gamma - \log \alpha_i), \quad (4.6)
\]

\[
E(\frac{\partial^2 l}{\partial \beta_i^2}) = \frac{1}{\beta_i^2}[(1 - \gamma - \log \alpha_i)^2 + \frac{\pi^2}{6}], \quad (4.7)
\]

for \( i = 1, 2 \), and

\[
E(\frac{\partial^2 l}{\partial \alpha_i \partial \alpha_j}) = E(\frac{\partial^2 l}{\partial \beta_i \partial \beta_j}) = E(\frac{\partial^2 l}{\partial \alpha_i \partial \beta_j}) = E(\frac{\partial^2 l}{\partial \alpha_i \partial \nu}) = E(\frac{\partial^2 l}{\partial \beta_i \partial \nu}) = 0,
\]

for \( i, j = 1, 2 \), and \( i \neq j \). Furthermore, we have

\[
E(\frac{\partial^2 l}{\partial \nu^2}) = -\infty,
\]

68
Figure 4.1: Asymptotic variance of parameter estimator of $\nu$ under MLE and marginal method, as $n \to \infty$

which is relevant to the score test that we discuss in the next chapter. Expressions in (4.5) - (4.7) form the sub-matrix

$$I' = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix},$$

which agrees with the Weibull Fisher information matrix.

For the case $0 < \nu < 1$, standard regularity conditions (Cox and Hinkley, 1982) hold and hence the MLEs have the usual asymptotic properties and are asymptotically efficient. The asymptotic variance-covariance matrix of the maximum likelihood estimators can be easily obtained by inverting the Fisher information matrix, i.e. $n^{-1}I^{-1}$, where $n$ is the sample size. By choosing $\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 1$, we have plotted the asymptotic variance of the MLE of $\nu$ for $0 < \nu < 1$, and $n \to \infty$, see Figure 4.1.

Maximisation of the five parameter log-likelihood function (4.3) is computationally non-trivial. One particular problem which might occur in the maximisation is that a good choice of the initial values is needed. A poor choice of initial values may lead to non-convergence of standard iterative methods. Also, care is needed when the parameter $\nu$ is near the boundary $\nu = 1$.
4.3 Two stage marginal estimator and its asymptotic properties

4.3.1 Two stage marginal estimation method

In terms of the property that the margins are independent of the association parameter \( \nu \) in our re-parameterised model (4.2), we consider another possible estimation method that turns out to be nearly as efficient as maximum likelihood. Although the method for the case with censored observations is also investigated in the simulation study, at present, we only consider the case in which there are no censored observations. An advantage of this method is that the computational work is greatly simplified because the marginal parameters and the association parameter are estimated separately. The method is called the two stage marginal method here. As the margins of the model (4.2) have Weibull distributions, with parameters independent of \( \nu \), at the first stage, we are able to obtain parameter estimators \( \hat{\alpha}_1 \) and \( \hat{\beta}_1 \) from the first sample, and corresponding estimators \( \hat{\alpha}_2 \) and \( \hat{\beta}_2 \) from the second sample separately.

At the second stage, by replacing the true values of \( \alpha_j \) and \( \beta_j \) in the log-likelihood function (4.3) with the maximum likelihood estimators \( \hat{\alpha}_j \) and \( \hat{\beta}_j \), the log-likelihood function of model (4.2) becomes

\[
\tilde{l}_n = - \sum_{i=1}^{n} \tilde{s}_i + \frac{n}{\nu} (\log \hat{\alpha}_1 + \log \hat{\alpha}_2) + n (\log \hat{\beta}_1 + \log \hat{\beta}_2 - \log \nu) \\
+ \left( \frac{\hat{\beta}_1}{\nu} - 1 \right) \sum_{i=1}^{n} \log y_{i1} + \left( \frac{\hat{\beta}_2}{\nu} - 1 \right) \sum_{i=1}^{n} \log y_{i2} \\
+ (\nu - 2) \sum_{i=1}^{n} \log \tilde{s}_i + \sum_{i=1}^{n} \log \{ \nu \tilde{s}_i^\nu - \nu + 1 \}, \quad (4.8)
\]

where \( \tilde{s}_i = \frac{1}{\hat{\alpha}_1} \hat{\beta}_1 y_{i1}^\nu + \frac{1}{\hat{\alpha}_2} \hat{\beta}_2 y_{i2}^\nu \).

Maximising this one-parameter pseudo likelihood with respect to \( \nu \) yields an estimator \( \hat{\nu} \) of \( \nu \). This is straightforward numerically because the maximisation is one-dimensional.
4.3.2 Some notation and details relevant to the model

This section provides some of the intermediate steps leading to the representation and results in the following section.

Consider the random variables $Z_1 = \alpha_1 Y_1^{\beta_1}$ and $Z_2 = \alpha_2 Y_2^{\beta_2}$. Their joint survivor function is given by

$$S(z_1, z_2) = \exp[-(z_1^{\frac{1}{\alpha_1}} + z_2^{\frac{1}{\alpha_2}})^{\nu}].$$

Lee (1979) shows that $Z_1$ and $Z_2$ can be represented in terms of two independent random variables $U$ and $V$, defined by

$$U = \frac{Z_1^{\frac{1}{\alpha_1}}}{Z_1^{\frac{1}{\alpha_1}} + Z_2^{\frac{1}{\alpha_2}},}$$

$$V = (Z_1^{\frac{1}{\alpha_1}} + Z_2^{\frac{1}{\alpha_2}})^{\nu},$$

where $U$ is uniformly distributed on $[0,1]$ and $V$ is a mixture of gamma random variables with density

$$g(v) = e^{-v}(1 - \nu + \nu v) \quad v > 0.$$  

The representation of $U$ and $V$ helps us to calculate the joint moments of $Z_1$ and $Z_2$, which are, for any non-negative $p$ and $q$,

$$E(Z_1^p Z_2^q) = E[(U^p V)^p((1 - U)^q V)^q]$$

$$= E[U^{vp}(1 - U)^{\nu q}] E[V^{p+q}]$$

$$= \frac{\Gamma(\nu p + 1)\Gamma(\nu q + 1)\Gamma(p+q+1)}{\Gamma(\nu p + \nu q + 1)}. \quad (4.9)$$

Expectations involving log $Z_1$ and log $Z_2$ can be obtained by differentiating $E(Z_1^p Z_2^q)$ with respect to $p$ and $q$. That is

$$E[Z_1^p Z_2^q (\log Z_1)^m (\log Z_2)^n] = \frac{\partial^m}{\partial q^n} \frac{\partial^n}{\partial q^n} [E(Z_1^p Z_2^q)], \quad (4.10)$$

where $m \geq 0$, and $n \geq 0$.

Let $\theta$ be the column vector of parameters $\alpha_1, \beta_1, \alpha_2, \beta_2$. Let $l(\theta, \nu)$ be the log likelihood for a single (uncensored) observation, and $l^*(\theta) = l(\theta, 1)$ be the log likelihood for $\theta$ that arises from assuming that $Y_1$ and $Y_2$ are independent. The marginal estimator $\hat{\theta} = (\hat{\alpha_1}, \hat{\beta_1}, \hat{\alpha_2}, \hat{\beta_2})^T$ is the maximum likelihood estimator found by
maximising \( l^* (\theta) \), and \( \nu \) is the maximum likelihood estimator found by maximising \( l(\theta, \nu) \) with respect to \( \nu \).

Let

\[
I^* = \begin{pmatrix}
I^*_{11} & I^*_{12} \\
I^*_{21} & I^*_{22}
\end{pmatrix},
\]

where

\[
I^*_{ij} = \begin{pmatrix}
-\text{E}(\partial^2 l^* / \partial \alpha_i \partial \alpha_j) & -\text{E}(\partial^2 l^* / \partial \alpha_i \partial \beta_j) \\
-\text{E}(\partial^2 l^* / \partial \beta_i \partial \alpha_j) & -\text{E}(\partial^2 l^* / \partial \beta_i \partial \beta_j)
\end{pmatrix}_{i, j = 1, 2}.
\]

It is obvious that \( I^*_{jj} (j = 1, 2) \), can be evaluated as the negative of the corresponding part of the Fisher information matrix \( I_{jj} \) for the case in which \( \nu = 1 \), see (4.5) - (4.7). To evaluate the other elements of the matrix \( I^* \), we consider the following relation between the first derivative of \( l^* \) with respect to \( \alpha_j \) and \( \beta_j \) and the random variables \( Z_1 \) and \( Z_2 \). Each element of the vector \( \partial l^*/\partial \theta = (\partial l^*/\partial \alpha_1, \partial l^*/\partial \beta_1, \partial l^*/\partial \alpha_2, \partial l^*/\partial \beta_2)^T \) can be expressed in terms of random variables \( Z_1 \) and \( Z_2 \), that is

\[
\frac{\partial l^*}{\partial \alpha_j} = -y_j + \frac{1}{\alpha_j} = \frac{1}{\alpha_j} (1 - Z_j)
\]

\[
\frac{\partial l^*}{\partial \beta_j} = -\alpha_j y_j \log y_j + \frac{1}{\beta_j} + \log y_j
\]

\[
= \frac{1}{\beta_j} [1 + (1 - Z_j)(\log Z_j - \log \alpha_j)]
\]

where \( j = 1, 2 \). Therefore,

\[
E(\frac{\partial l^*}{\partial \alpha_1} \frac{\partial l^*}{\partial \alpha_2}) = \frac{1}{\alpha_1 \alpha_2} E \{(1 - Z_1)(1 - Z_2)\} \tag{4.11}
\]

\[
E(\frac{\partial l^*}{\partial \beta_1} \frac{\partial l^*}{\partial \beta_2}) = \frac{1}{\beta_1 \beta_2} E[1 + (1 - Z_1)(\log Z_1 - \log \alpha_1)][1 + (1 - Z_2)(\log Z_2 - \log \alpha_2)] \tag{4.12}
\]

\[
E(\frac{\partial l^*}{\partial \alpha_1} \frac{\partial l^*}{\partial \beta_2}) = \frac{1}{\alpha_1 \beta_2} E \{(1 - Z_1)[1 + (1 - Z_2)(\log Z_2 - \log \alpha_2)]\} \tag{4.13}
\]

\[
E(\frac{\partial l^*}{\partial \alpha_2} \frac{\partial l^*}{\partial \beta_2}) = \frac{1}{\alpha_2 \beta_1} E \{(1 - Z_2)[1 + (1 - Z_1)(\log Z_1 - \log \alpha_1)]\} \tag{4.14}
\]

The expressions (4.11)-(4.14) involve various combinations of joint moments of \( Z_1 \), \( Z_2 \), \( \log Z_1 \), and \( \log Z_2 \), which can be evaluated using (4.9) and (4.10). In terms of the following relationship

\[
E(\frac{\partial l^*}{\partial \alpha_i} \frac{\partial l^*}{\partial \alpha_j}) = -E(\frac{\partial^2 l^*}{\partial \alpha_i \partial \alpha_j}) \tag{4.15}
\]
\[
E(\frac{\partial l^*}{\partial \alpha_j \partial \beta_j}) = -E(\frac{\partial^2 l^*}{\partial \alpha_j \partial \beta_j}) \tag{4.16}
\]
\[
E(\frac{\partial l^*}{\partial \beta_i \partial \beta_j}) = -E(\frac{\partial^2 l^*}{\partial \beta_i \partial \beta_j}) \tag{4.17}
\]

all the entries of matrix \( I^* \) may be evaluated in this way.

### 4.3.3 Asymptotic variance

Shih & Louis (1995b) investigated the two stage marginal estimation method for the Archimedean copula models. The variance of the association parameter is derived assuming that the functional forms of the margins are known and have a finite number of unknown parameters. As stated in Section 3.1, the bivariate PSW model we discuss here is a special case of Shih & Louis (1995b)'s work. Applying their general results to our specific model that the margins are two-parameter Weibull distributions, we obtain the asymptotic variance of the two stage marginal estimator \( \tilde{\nu} \)

\[
Var(\tilde{\nu}) = \frac{1}{I_{33}} + \frac{1}{I_{33}} (I_{31} I_{11}^{-1} I_{13} + I_{32} I_{22}^{-1} I_{23} + I_{31} I_{11}^{-1} I_{11}^{-1} I_{13}). \tag{4.18}
\]

This result also agrees with that obtained by Shi et al (1992), who discussed the two stage marginal estimation for the bivariate Logistic model where the margins are Gumbel distributed. The joint distribution function of the bivariate Logistic model is

\[
F(y_1, y_2) = \Pr(Y_1 < y_1, Y_2 < y_2) = \exp\left\{ -e^{-\frac{(y_1-\mu_1)}{\sigma_1}} + e^{-\frac{(y_2-\mu_2)}{\sigma_2}} \right\}. \tag{4.19}
\]

Let

\[
X_j = e^{-\frac{(y_j-\mu_j)}{\sigma_j}}, \quad j = 1, 2
\]

The survivor function of \( X_1 \) and \( X_2 \) is then

\[
S(x_1, x_2) = \Pr(X_1 > x_1, X_2 > x_2) = \Pr(Y_1 < y_1, Y_2 < y_2) = \exp\left\{ -(\frac{1}{\sigma_1} x_1^{\frac{1}{\sigma_1}} + \frac{1}{\sigma_2} x_2^{\frac{1}{\sigma_2}}) \right\}. \tag{4.20}
\]
which is in the form of the PSW model with $\xi_j = 1$ and $\phi_j = 1$.

Evaluation of each term in the expression (4.18) is explicit. The second term in $\text{Var}(\hat{\nu})$ accounts for the loss in asymptotic efficiency in estimating $\nu$ due to the lack of information on $(\alpha_1, \beta_1, \alpha_2, \beta_2)$. We have $I_{13} = I_{23} = 0$ as $Y_1$ and $Y_2$ are independent. In this case, the asymptotic variance of $\hat{\nu}$ agrees with that of $\hat{\nu}$ when the true value of $\nu$ is $\nu = 1$.

To compare the asymptotic variance of estimator $\hat{\nu}$ from the two stage marginal method with that of the maximum likelihood estimator $\hat{\nu}$, we calculate and plot $\text{Var}(\hat{\nu})$, see Figure 4.1, as $0 < \nu < 1$, and $\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 1$. In that case, $\text{Var}(\hat{\nu})$ simplifies to

$$\text{Var}(\hat{\nu}) = \frac{1}{I_{33}} + \frac{2}{I_{33}^2} [I_{31}(I_{11}^{-1} + I_{11}^{-1}I_{12}I_{11}^{-1})I_{13}].$$

Figure 4.1 indicates that, for a wide range of $\nu$ values, the difference between the variance of the estimator of $\nu$ from MLE and the marginal method is negligible.

The relative efficiency of the estimation method is measured by the ratio of the variance of $\hat{\nu}$ to the variance of $\hat{\nu}$. A plot of the relative efficiency as a function of the parameter $\nu$ is given in Figure 4.2, which shows that on the whole the maximum likelihood estimator of $\nu$ is slightly better than the two stage marginal estimator of $\nu$. However, note that the relative efficiency of $\hat{\nu}$ is greater than 99% for $0 < \nu < 1$. The relative efficiency is comparatively smaller in the middle of
the curve where the dependence is moderate; at the two ends of the region, that is when the components are nearly independent or nearly completely dependent, the asymptotic variances of the two estimation methods are virtually the same. However, $\nu$ is essentially fully efficient for all values of $\nu$ for all practical purposes.

For the marginal estimators, standard regularity conditions hold and therefore the estimators are asymptotically normal and unbiased. The asymptotic variance of the marginal estimators $\hat{\theta} = (\hat{\alpha}_1, \hat{\beta}_1, \hat{\alpha}_2, \hat{\beta}_2)$ can be easily evaluated from the standard method, which is the inverse of $I_{ii}$ $(i = 1, 2)$ given in (4.5) – (4.7).

Similarly, we plot the asymptotic variance of the maximum likelihood estimators of $\alpha_j$ and $\beta_j$, as well as the asymptotic variance of the marginal estimators of $\alpha_j$ and $\beta_j$. Corresponding plots are shown in Figure 4.3. Plots of the ratio, i.e. the relative efficiency of the estimators of $\alpha_j$ and $\beta_j$, are given in Figure 4.4. It can be seen that, as $\nu$ varies within $(0, 1)$, the marginal estimators of $\alpha_j$ and $\beta_j$ from the two stage marginal method perform as well as those from the MLE method. The biggest difference in variance between the methods occurs near $\nu = 0.7$ and $\nu = 0.5$ for the estimators of $\alpha_j$ and $\beta_j$ respectively. The corresponding relative efficiency of $\hat{\alpha}_j$ and $\hat{\beta}_j$ is greater than 98% and 93% respectively for all $\nu$. Again the loss of efficiency in using the two stage marginal method is very small.

4.4 Simulation

A simulation study was performed to investigate the efficiency of the two stage marginal estimators compared with the corresponding MLEs for finite samples. As an extension to the discussion on the bivariate model estimation, the simulation study includes the bivariate, tri-variate and four-variate cases. Programs were written in Fortran 77 and NAG (1995) library routines were called for generating random variables and performing maximum likelihood estimations, see Appendix A. The random variables $y_{1i}$ and $y_{2i}$ based on the PSW model were obtained from $y_{ij} = x_{ij}/w_i$, where $x_{i1}$ and $x_{i2}$ were generated from two independent Weibull distribution by calling NAG (1995) subroutine G05DPF, $w_i$ were generated from a positive stable distribution with characteristic exponent $\nu$, and $i = 1, \ldots, n$, $j = 1, 2$. The simulation method of generating positive stable random variables $w_i$ is based on the algorithm of Chambers et al (1976).
4.1 Bivariate models

...with correlations... Variables (alpha and beta)...

Figure 4.3: Asymptotic variance of parameter estimators of $\alpha_j$ and $\beta_j$ under MLE and the marginal method, as a function of $\nu$, and where the true values of parameters $\alpha_j$ and $\beta_j$ are unity.

Figure 4.4: Relative efficiency of the marginal estimators of $\alpha_j$ and $\beta_j$. 

76
4.4.1 Bivariate models

Case without covariates

First we chose sample size \( n = 100 \). Without loss of generality, for each of various values of \( \nu \) between 0.1 and 0.9, we chose parameters \( \alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 1 \), and generated 1000 simulated samples. The parameters \((\alpha_1, \beta_1, \alpha_2, \beta_2, \nu)\) for each sample were estimated using both MLE and the two stage marginal method, and then the means and variances of the estimators were calculated. The estimators and their relative efficiencies are listed in Tables 4.1, 4.2, and 4.3. Note that the estimation of \( \alpha \) and \( \beta \) given in Tables 4.2, 4.3 and later tables refer to the estimation of the marginal parameters from the first component. The results show that the efficiencies of the marginal method are good everywhere. The biases for both methods are small.

A similar simulation was conducted for \( n = 50 \). In this case, \( \nu \) was chosen as 0.3, 0.5 and 0.7. Similar results are obtained which are presented in Table 4.4. The performance of the parameter estimators for the smaller sample size is similar to that for the larger sample size.

We also consider the situation in which the sample has censored observations on either or both components. Assume that \( Y_1 \) and \( Y_2 \) are censored at the same fixed time \( c \). For the case in which there is censoring and the marginal parameters are \( \alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 1 \), the survivor function is

\[
S(y_1, y_2) = \exp\left[-\left(y_1^{\frac{1}{\nu}} + y_2^{\frac{1}{\nu}}\right)\right].
\]

Therefore, the population proportion of censored observations \( p_c \) is related to the censoring time \( c \) as follows,

\[
p_c = \Pr(Y_1 > c, Y_2 > c) - \Pr(Y_1 > c, Y_2 > c) = 2 \exp(-c) - \exp[-(2^\nu c)].
\]

Using the above expression, the censoring points \( c \) in terms of various selections of \( \nu \) were obtained so that \( p_c = 0.3 \). The results are given in Table 4.7.

Samples with 30\% censoring on average are generated by fixing the censoring points for various \( \nu \) values. Tables 4.5 and 4.6 shows the mean and variance results.
estimated from 30% censored samples with sample size \( n = 100 \) and \( n = 50 \). The performance of the estimators is almost as good as that of uncensored cases.

**Comparison with an estimator of \( \nu \) based on a sample correlation coefficient**

As we discussed in Chapter 3, the association parameter \( \nu \) in the PSW model is related to the correlation coefficient \( \rho \), such that

\[
\rho(\log Y_1, \log Y_2) = 1 - \nu^2.
\]

In the case in which there is no censoring and no covariates, an easier way to estimate the parameter \( \nu \) may be obtained via the estimate of correlation coefficient \( \rho \). Assume that \( \hat{\rho} \) is a sample correlation coefficient. Then, the estimate of parameter \( \nu \), \( \hat{\nu}^* \), is

\[
\hat{\nu}^* = \sqrt{1 - \hat{\rho}}.
\]

The asymptotic properties of \( \hat{\nu}^* \) are difficult to obtain. However, we performed a simulation experiment to examine the relative efficiency of \( \hat{\nu}^* \), which is the ratio of the variance of \( \hat{\nu} \) to the variance of \( \hat{\nu}^* \). The simulation results are shown in the last two columns in Table 4.1, which indicate that \( \hat{\nu}^* \) has much lower relative efficiency than that of the two stage estimator of \( \nu \). Therefore, we conclude that, although the parameter \( \nu \) may be easily estimated on the basis of the sample correlation coefficient, the precision of the estimation is not as satisfactory as that from the two stage estimation method. The estimate of \( \nu \) based on a sample correlation coefficient might be a good choice as the initial value for the maximum likelihood estimation in the two stage marginal estimation method.

**Case with covariates**

In practice, some covariates might be associated with survival times. Therefore, these covariates should be considered in our model. Simulation was carried out to investigate whether the presence of covariates affects the performance of the estimators.

In our simulation, it is assumed that only one covariate is included in the PSW model, which is a binary variable, with \( x_i = 1 \) for half of the observations and
Table 4.1: Estimation of $\nu$ for $n=100$ (no censoring) $p = 2$

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>MLE Method Mean(Variance)</th>
<th>Marginal Method Mean(Variance)</th>
<th>Estimated Efficiency</th>
<th>Correlation Method Mean(Variance)</th>
<th>Estimated Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.10078(0.00013)</td>
<td>0.10167(0.00013)</td>
<td>0.97612</td>
<td>0.10145(0.00018)</td>
<td>0.69235</td>
</tr>
<tr>
<td>0.2</td>
<td>0.20144(0.00048)</td>
<td>0.20285(0.00050)</td>
<td>0.96741</td>
<td>0.20055(0.00067)</td>
<td>0.73035</td>
</tr>
<tr>
<td>0.3</td>
<td>0.30177(0.00100)</td>
<td>0.30338(0.00101)</td>
<td>0.98921</td>
<td>0.30203(0.00154)</td>
<td>0.69036</td>
</tr>
<tr>
<td>0.4</td>
<td>0.40119(0.00164)</td>
<td>0.40298(0.00167)</td>
<td>0.98419</td>
<td>0.40103(0.00238)</td>
<td>0.62335</td>
</tr>
<tr>
<td>0.5</td>
<td>0.50141(0.00242)</td>
<td>0.50278(0.00242)</td>
<td>1.00004</td>
<td>0.50287(0.00381)</td>
<td>0.68856</td>
</tr>
<tr>
<td>0.6</td>
<td>0.60196(0.00309)</td>
<td>0.60330(0.00314)</td>
<td>0.98545</td>
<td>0.60295(0.00449)</td>
<td>0.68895</td>
</tr>
<tr>
<td>0.7</td>
<td>0.70267(0.00322)</td>
<td>0.70385(0.00322)</td>
<td>1.0018</td>
<td>0.69790(0.00536)</td>
<td>0.66452</td>
</tr>
<tr>
<td>0.8</td>
<td>0.79815(0.00404)</td>
<td>0.79890(0.00401)</td>
<td>1.00663</td>
<td>0.80035(0.00508)</td>
<td>0.76503</td>
</tr>
<tr>
<td>0.9</td>
<td>0.89968(0.00363)</td>
<td>0.90027(0.00358)</td>
<td>1.01140</td>
<td>0.90110(0.00479)</td>
<td>0.78475</td>
</tr>
</tbody>
</table>

Table 4.2: Estimation of $\alpha$ for $n=100$ (no censoring) $p = 2$

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>MLE Method Mean(Variance)</th>
<th>Marginal Method Mean(Variance)</th>
<th>Estimated Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1.01292(0.01185)</td>
<td>1.01294(0.01185)</td>
<td>1.00018</td>
</tr>
<tr>
<td>0.2</td>
<td>1.00785(0.01100)</td>
<td>1.00826(0.01112)</td>
<td>0.98925</td>
</tr>
<tr>
<td>0.3</td>
<td>1.00433(0.01133)</td>
<td>1.00492(0.01144)</td>
<td>0.99027</td>
</tr>
<tr>
<td>0.4</td>
<td>1.00321(0.01123)</td>
<td>1.00406(0.01128)</td>
<td>0.99503</td>
</tr>
<tr>
<td>0.5</td>
<td>1.00698(0.01098)</td>
<td>1.00706(0.01108)</td>
<td>0.99107</td>
</tr>
<tr>
<td>0.6</td>
<td>1.00455(0.01180)</td>
<td>1.00513(0.01195)</td>
<td>0.98696</td>
</tr>
<tr>
<td>0.7</td>
<td>1.00835(0.01112)</td>
<td>1.00900(0.01120)</td>
<td>0.99263</td>
</tr>
<tr>
<td>0.8</td>
<td>1.00320(0.01090)</td>
<td>1.00360(0.01111)</td>
<td>0.98141</td>
</tr>
<tr>
<td>0.9</td>
<td>1.00857(0.01195)</td>
<td>1.00865(0.01203)</td>
<td>0.99297</td>
</tr>
</tbody>
</table>

Table 4.3: Estimation of $\beta$ for $n=100$ (no censoring) $p = 2$

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>MLE Method Mean(Variance)</th>
<th>Marginal Method Mean(Variance)</th>
<th>Estimated Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1.01292(0.01185)</td>
<td>1.01294(0.01185)</td>
<td>1.00018</td>
</tr>
<tr>
<td>0.2</td>
<td>1.01694(0.00641)</td>
<td>1.01705(0.00648)</td>
<td>0.98954</td>
</tr>
<tr>
<td>0.3</td>
<td>1.01656(0.00596)</td>
<td>1.01652(0.00617)</td>
<td>0.96617</td>
</tr>
<tr>
<td>0.4</td>
<td>1.01530(0.00657)</td>
<td>1.01577(0.00686)</td>
<td>0.95904</td>
</tr>
<tr>
<td>0.5</td>
<td>1.01767(0.00617)</td>
<td>1.01707(0.00670)</td>
<td>0.92153</td>
</tr>
<tr>
<td>0.6</td>
<td>1.01576(0.00623)</td>
<td>1.01505(0.00670)</td>
<td>0.93000</td>
</tr>
<tr>
<td>0.7</td>
<td>1.01767(0.00617)</td>
<td>1.01707(0.00670)</td>
<td>0.92153</td>
</tr>
<tr>
<td>0.8</td>
<td>1.01389(0.00647)</td>
<td>1.01367(0.00659)</td>
<td>0.98296</td>
</tr>
<tr>
<td>0.9</td>
<td>1.01303(0.00640)</td>
<td>1.01333(0.00650)</td>
<td>0.98397</td>
</tr>
</tbody>
</table>

79
Table 4.4: Estimation for \( n = 50 \) (no censoring) \( p = 2 \)

<table>
<thead>
<tr>
<th>Estimate</th>
<th>( \nu )</th>
<th>MLE Method Mean( (\text{Variance}) )</th>
<th>Marginal Method Mean( (\text{Variance}) )</th>
<th>Estimated Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \nu )</td>
<td>0.3</td>
<td>0.30164(0.00133)</td>
<td>0.30422(0.00135)</td>
<td>0.98585</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.50009(0.00256)</td>
<td>0.50211(0.00259)</td>
<td>0.98853</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.70208(0.00378)</td>
<td>0.70352(0.00377)</td>
<td>1.00194</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.3</td>
<td>1.00102(0.01341)</td>
<td>1.00224(0.01382)</td>
<td>0.97048</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.00026(0.01307)</td>
<td>1.00042(0.01338)</td>
<td>0.97700</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>1.00297(0.01223)</td>
<td>1.00282(0.01237)</td>
<td>0.98913</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.3</td>
<td>1.01383(0.00970)</td>
<td>1.01618(0.01068)</td>
<td>0.90882</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.01162(0.00836)</td>
<td>1.01198(0.00916)</td>
<td>0.91223</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>1.01929(0.00924)</td>
<td>1.02048(0.00970)</td>
<td>0.95258</td>
</tr>
</tbody>
</table>

Table 4.5: Estimation for samples with 30% censoring \( (n = 100) \) \( p = 2 \)

<table>
<thead>
<tr>
<th>Estimate</th>
<th>( \nu )</th>
<th>MLE Method Mean( (\text{Variance}) )</th>
<th>Marginal Method Mean( (\text{Variance}) )</th>
<th>Estimated Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \nu )</td>
<td>0.3</td>
<td>0.30164(0.00133)</td>
<td>0.30422(0.00135)</td>
<td>0.98585</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.50009(0.00256)</td>
<td>0.50211(0.00259)</td>
<td>0.98853</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.70208(0.00378)</td>
<td>0.70352(0.00377)</td>
<td>1.00194</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.3</td>
<td>1.00102(0.01341)</td>
<td>1.00224(0.01382)</td>
<td>0.97048</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.00026(0.01307)</td>
<td>1.00042(0.01338)</td>
<td>0.97700</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>1.00297(0.01223)</td>
<td>1.00282(0.01237)</td>
<td>0.98913</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.3</td>
<td>1.01383(0.00970)</td>
<td>1.01618(0.01068)</td>
<td>0.90882</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.01162(0.00836)</td>
<td>1.01198(0.00916)</td>
<td>0.91223</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>1.01929(0.00924)</td>
<td>1.02048(0.00970)</td>
<td>0.95258</td>
</tr>
</tbody>
</table>

Table 4.6: Estimation for samples with 30% censoring \( (n = 50) \) \( p = 2 \)

<table>
<thead>
<tr>
<th>Estimate</th>
<th>( \nu )</th>
<th>MLE Method Mean( (\text{Variance}) )</th>
<th>Marginal Method Mean( (\text{Variance}) )</th>
<th>Estimated Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \nu )</td>
<td>0.3</td>
<td>0.29997(0.00254)</td>
<td>0.30464(0.00255)</td>
<td>0.99353</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.49790(0.00549)</td>
<td>0.50223(0.00549)</td>
<td>0.99952</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.69536(0.00780)</td>
<td>0.69811(0.00772)</td>
<td>1.01034</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.3</td>
<td>1.01044(0.02804)</td>
<td>1.01176(0.02891)</td>
<td>0.96974</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.01223(0.02603)</td>
<td>1.01267(0.02654)</td>
<td>0.98068</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>1.01525(0.02812)</td>
<td>1.01593(0.02835)</td>
<td>0.99207</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.3</td>
<td>1.01869(0.01950)</td>
<td>1.02116(0.02110)</td>
<td>0.92396</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.02579(0.01963)</td>
<td>1.02818(0.02163)</td>
<td>0.90746</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>1.02152(0.01652)</td>
<td>1.02441(0.01758)</td>
<td>0.93976</td>
</tr>
</tbody>
</table>
Table 4.7: Censoring point for various selections of $\nu$ when $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 1$, $p_c = 30\%$

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>1.30</td>
<td>1.40</td>
<td>1.45</td>
<td>1.50</td>
<td>1.60</td>
<td>1.65</td>
<td>1.70</td>
<td>1.75</td>
<td>1.80</td>
</tr>
</tbody>
</table>

$x_i = -1$ for the other half of the observations, $i = 1, \ldots, n$. For example, $x$ may represent the sex, male or female. It may be an indicator of smoking habit, yes or no. Simulated pairs of samples were generated from the PSW model, with parameters $\beta_1 = \beta_2 = 1$ and $\log \alpha_j = \nu(b_{0j} + b_{1j}x_i), j = 1, 2$. The regression coefficients were chosen as $b_{01} = b_{02} = 1, b_{11} = 1$ and $b_{12} = -1$. For sample size $n = 100$, the mean and variance of all estimated parameters and their corresponding relative efficiencies are listed in Table 4.8. The true value of $\nu$ was selected as $0.3, 0.5, 0.7$.

The results indicate that the performance of the two stage marginal method in the case with covariates is comparable to that in the case without covariates.

### 4.4.2 Tri-variate and four-variate models

Simulation was also carried out to compare the two estimation methods for tri-variate and four-variate models. Following the same procedures as mentioned above, the parameters of the tri-variate and four-variate models are estimated by using the MLE and the two stage marginal method. Simulated samples have sample size $n = 100$. The means and variances of the estimated parameters and their corresponding efficiencies are listed in Tables 4.9–4.14 in terms of selected true $\nu$ values, between 0.2 and 0.8. The results indicate that the good efficiency of the two stage marginal method remains in higher dimensions, at least when $p = 3$ and $p = 4$. Generally speaking, the estimation from the two stage marginal method is almost as good as that from the MLE method everywhere. The worst efficiencies for the parameter estimators are those for the parameter $\beta$ in both the $p = 3$ and $p = 4$ cases, where the lowest efficiencies occur with values around 0.85 when $\nu$ is close to 0.5. However, even with these values, the two stage marginal method is still worth considering either as an alternative to ML or as a method of obtaining good starting values for an iterative ML scheme.
Table 4.8: Estimation for samples with a covariate \((n = 100)\) \(p = 2\)

<table>
<thead>
<tr>
<th>Estimate</th>
<th>(\nu)</th>
<th>MLE Method Mean(Variance)</th>
<th>Marginal Method Mean(Variance)</th>
<th>Estimated Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\nu)</td>
<td>0.3</td>
<td>0.30127(0.00106)</td>
<td>0.30385(0.00107)</td>
<td>0.99690</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.50182(0.00238)</td>
<td>0.50463(0.00237)</td>
<td>1.00300</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.70174(0.00353)</td>
<td>0.70368(0.00351)</td>
<td>1.00733</td>
</tr>
<tr>
<td>(b_{01})</td>
<td>0.3</td>
<td>1.02603(0.01267)</td>
<td>1.02762(0.01287)</td>
<td>0.98456</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.02080(0.01266)</td>
<td>1.02180(0.01319)</td>
<td>0.95959</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>1.02604(0.01277)</td>
<td>1.02585(0.01289)</td>
<td>0.99113</td>
</tr>
<tr>
<td>(b_{11})</td>
<td>0.3</td>
<td>1.01452(0.01787)</td>
<td>1.01559(0.01891)</td>
<td>0.94530</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.02414(0.01760)</td>
<td>1.02321(0.01845)</td>
<td>0.95366</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>1.02433(0.01765)</td>
<td>1.02397(0.01841)</td>
<td>0.95882</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>0.3</td>
<td>1.01948(0.00651)</td>
<td>1.02050(0.00680)</td>
<td>0.95798</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.01895(0.00570)</td>
<td>1.01890(0.00612)</td>
<td>0.93060</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>1.02313(0.00640)</td>
<td>1.02230(0.00677)</td>
<td>0.94572</td>
</tr>
<tr>
<td>(b_{02})</td>
<td>0.3</td>
<td>1.02344(0.01266)</td>
<td>1.02308(0.01326)</td>
<td>0.95519</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.01948(0.01315)</td>
<td>1.02010(0.01355)</td>
<td>0.97020</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>1.02468(0.01337)</td>
<td>1.02493(0.01345)</td>
<td>0.99449</td>
</tr>
<tr>
<td>(b_{12})</td>
<td>0.3</td>
<td>-1.02266(0.01656)</td>
<td>-1.02162(0.01704)</td>
<td>0.97209</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>-1.01687(0.01457)</td>
<td>-1.01649(0.01554)</td>
<td>0.93763</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>-1.01817(0.01588)</td>
<td>-1.01729(0.01643)</td>
<td>0.96652</td>
</tr>
<tr>
<td>(\beta_2)</td>
<td>0.3</td>
<td>1.01713(0.00638)</td>
<td>1.01629(0.00663)</td>
<td>0.96314</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.02040(0.00613)</td>
<td>1.02037(0.00660)</td>
<td>0.92833</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>1.02164(0.00613)</td>
<td>1.02144(0.00642)</td>
<td>0.95557</td>
</tr>
</tbody>
</table>
Table 4.9: Estimation of $\nu$ for $n=100$ (no censoring) $p = 3$

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>MLE Method Mean (Variance)</th>
<th>Marginal Method Mean (Variance)</th>
<th>Estimated Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.20055 (0.00032)</td>
<td>0.20249 (0.00033)</td>
<td>0.97280</td>
</tr>
<tr>
<td>0.3</td>
<td>0.30143 (0.00075)</td>
<td>0.30371 (0.00077)</td>
<td>0.96505</td>
</tr>
<tr>
<td>0.4</td>
<td>0.40131 (0.00110)</td>
<td>0.40347 (0.00113)</td>
<td>0.97306</td>
</tr>
<tr>
<td>0.5</td>
<td>0.49991 (0.00148)</td>
<td>0.50186 (0.00149)</td>
<td>0.98865</td>
</tr>
<tr>
<td>0.6</td>
<td>0.60381 (0.00189)</td>
<td>0.60537 (0.00190)</td>
<td>0.99542</td>
</tr>
<tr>
<td>0.7</td>
<td>0.69917 (0.00183)</td>
<td>0.70054 (0.00184)</td>
<td>0.99288</td>
</tr>
<tr>
<td>0.8</td>
<td>0.80115 (0.00202)</td>
<td>0.80240 (0.00201)</td>
<td>1.00506</td>
</tr>
</tbody>
</table>

Table 4.10: Estimation of $\alpha$ for $n=100$ (no censoring) $p = 3$

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>MLE Method Mean (Variance)</th>
<th>Marginal Method Mean (Variance)</th>
<th>Estimated Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>1.00498 (0.01076)</td>
<td>1.00506 (0.01089)</td>
<td>0.98768</td>
</tr>
<tr>
<td>0.3</td>
<td>1.00750 (0.01153)</td>
<td>1.00727 (0.01169)</td>
<td>0.98598</td>
</tr>
<tr>
<td>0.4</td>
<td>1.00500 (0.01030)</td>
<td>1.00501 (0.01061)</td>
<td>0.97055</td>
</tr>
<tr>
<td>0.5</td>
<td>1.00920 (0.01121)</td>
<td>1.01041 (0.01178)</td>
<td>0.95218</td>
</tr>
<tr>
<td>0.6</td>
<td>1.00130 (0.01079)</td>
<td>1.00306 (0.01103)</td>
<td>0.97841</td>
</tr>
<tr>
<td>0.7</td>
<td>1.00076 (0.01123)</td>
<td>1.00051 (0.01177)</td>
<td>0.95356</td>
</tr>
<tr>
<td>0.8</td>
<td>1.00562 (0.01147)</td>
<td>1.00662 (0.01173)</td>
<td>0.97767</td>
</tr>
</tbody>
</table>

Table 4.11: Estimation of $\beta$ for $n=100$ (no censoring) $p = 3$

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>MLE Method Mean (Variance)</th>
<th>Marginal Method Mean (Variance)</th>
<th>Estimated Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>1.01384 (0.00570)</td>
<td>1.01444 (0.00609)</td>
<td>0.93715</td>
</tr>
<tr>
<td>0.3</td>
<td>1.01426 (0.00618)</td>
<td>1.01433 (0.00696)</td>
<td>0.88336</td>
</tr>
<tr>
<td>0.4</td>
<td>1.01657 (0.00600)</td>
<td>1.01672 (0.00683)</td>
<td>0.87841</td>
</tr>
<tr>
<td>0.5</td>
<td>1.01236 (0.00552)</td>
<td>1.01169 (0.00614)</td>
<td>0.89805</td>
</tr>
<tr>
<td>0.6</td>
<td>1.01773 (0.00574)</td>
<td>1.01653 (0.00623)</td>
<td>0.92131</td>
</tr>
<tr>
<td>0.7</td>
<td>1.01580 (0.00576)</td>
<td>1.01737 (0.00624)</td>
<td>0.92325</td>
</tr>
<tr>
<td>0.8</td>
<td>1.01380 (0.00603)</td>
<td>1.01374 (0.00632)</td>
<td>0.95391</td>
</tr>
</tbody>
</table>
Table 4.12: Estimation of $\nu$ for $n=100$ (no censoring) $p = 4$

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>MLE Method Mean(Variance)</th>
<th>Marginal Method Mean(Variance)</th>
<th>Estimated Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.19863(0.00033)</td>
<td>0.20149(0.00034)</td>
<td>0.98238</td>
</tr>
<tr>
<td>0.2</td>
<td>0.20049(0.00031)</td>
<td>0.20260(0.00031)</td>
<td>0.99286</td>
</tr>
<tr>
<td>0.3</td>
<td>0.30020(0.00061)</td>
<td>0.30276(0.00062)</td>
<td>0.97903</td>
</tr>
<tr>
<td>0.4</td>
<td>0.40248(0.00097)</td>
<td>0.40495(0.00100)</td>
<td>0.97033</td>
</tr>
<tr>
<td>0.5</td>
<td>0.50229(0.00126)</td>
<td>0.50449(0.00129)</td>
<td>0.97999</td>
</tr>
<tr>
<td>0.6</td>
<td>0.60050(0.00146)</td>
<td>0.60242(0.00151)</td>
<td>0.96473</td>
</tr>
<tr>
<td>0.7</td>
<td>0.70049(0.00151)</td>
<td>0.70189(0.00160)</td>
<td>0.94398</td>
</tr>
<tr>
<td>0.8</td>
<td>0.79960(0.00157)</td>
<td>0.80088(0.00159)</td>
<td>0.99181</td>
</tr>
</tbody>
</table>

Table 4.13: Estimation of $\alpha$ for $n=100$ (no censoring) $p = 4$

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>MLE Method Mean(Variance)</th>
<th>Marginal Method Mean(Variance)</th>
<th>Estimated Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>1.00765(0.01220)</td>
<td>1.00821(0.01246)</td>
<td>0.97938</td>
</tr>
<tr>
<td>0.3</td>
<td>1.00594(0.01046)</td>
<td>1.00603(0.01075)</td>
<td>0.97304</td>
</tr>
<tr>
<td>0.4</td>
<td>1.00149(0.01102)</td>
<td>1.00225(0.01135)</td>
<td>0.97088</td>
</tr>
<tr>
<td>0.5</td>
<td>1.00236(0.01241)</td>
<td>1.00386(0.01306)</td>
<td>0.95016</td>
</tr>
<tr>
<td>0.6</td>
<td>1.00445(0.01129)</td>
<td>1.00432(0.01175)</td>
<td>0.96097</td>
</tr>
<tr>
<td>0.7</td>
<td>1.00299(0.01111)</td>
<td>1.00494(0.01180)</td>
<td>0.94199</td>
</tr>
<tr>
<td>0.8</td>
<td>1.00481(0.01131)</td>
<td>1.00469(0.01181)</td>
<td>0.95792</td>
</tr>
</tbody>
</table>

Table 4.14: Estimation of $\beta$ for $n=100$ (no censoring) $p = 4$

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>MLE Method Mean(Variance)</th>
<th>Marginal Method Mean(Variance)</th>
<th>Estimated Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>1.01571(0.00604)</td>
<td>1.01512(0.00641)</td>
<td>0.94267</td>
</tr>
<tr>
<td>0.3</td>
<td>1.01128(0.00577)</td>
<td>1.01098(0.00631)</td>
<td>0.91431</td>
</tr>
<tr>
<td>0.4</td>
<td>1.01775(0.00562)</td>
<td>1.01728(0.00630)</td>
<td>0.89309</td>
</tr>
<tr>
<td>0.5</td>
<td>1.01617(0.00574)</td>
<td>1.01363(0.00677)</td>
<td>0.84800</td>
</tr>
<tr>
<td>0.6</td>
<td>1.01162(0.00544)</td>
<td>1.01154(0.00632)</td>
<td>0.86145</td>
</tr>
<tr>
<td>0.7</td>
<td>1.01435(0.00538)</td>
<td>1.01300(0.00617)</td>
<td>0.87187</td>
</tr>
<tr>
<td>0.8</td>
<td>1.01063(0.00567)</td>
<td>1.01209(0.00627)</td>
<td>0.90485</td>
</tr>
</tbody>
</table>
4.5 Examples

4.5.1 Infant nutrition data

The infant nutrition data in Table 1.1 gives the ages in months of the first introduction of fish and egg into the diets of 55 infants. There are no censored observations in this data set.

We use MLE and the two stage marginal method to estimate the parameters of the bivariate PSW model for these data. The estimated parameters and the corresponding log likelihood are given in Table 4.15. The results show that the estimates from the two methods are very close.

4.5.2 Fibre failure strength data

The fibre failure strengths data is listed in Table 1.3. Initial examination of the marginal distributions shows that the Weibull model would be a reasonable starting point for fitting the strength data. The correlation of log transformed pairs of variables are calculated, which are 0.768, 0.784, 0.768, 0.774, 0.410 and 0.490 for the pairs \((y_1, y_2), (y_1, y_3), (y_1, y_4), (y_2, y_3), (y_2, y_4), (y_3, y_4)\), respectively. Concerning the possibility of extra variability in fibre strength, the PSW model is chosen to fit the data, although there is some doubt about its suitability given the unequal correlation.

First, for illustration purpose we consider the six bivariate PSW models. The parameter estimates for the two methods are given in Table 4.16. Clearly the two methods give similar results. Next, the two methods were applied to the full four-variate data. The resulting estimates are given in Table 4.17. Once again the estimates are similar for the two methods.

---

### Table 4.15: Fit of the bivariate PSW model on the nutrition data

<table>
<thead>
<tr>
<th>Method</th>
<th>(\alpha_1)</th>
<th>(\beta_1)</th>
<th>(\alpha_2)</th>
<th>(\beta_2)</th>
<th>(\nu)</th>
<th>log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>-8.640</td>
<td>4.074</td>
<td>-8.724</td>
<td>3.685</td>
<td>0.681</td>
<td>-241.837</td>
</tr>
<tr>
<td>Marginal</td>
<td>-8.817</td>
<td>4.166</td>
<td>-9.064</td>
<td>3.815</td>
<td>0.698</td>
<td>-241.965</td>
</tr>
</tbody>
</table>
It is necessary to point out that, practically speaking, the selection of initial values for the maximum likelihood estimation actually highly rely on the estimation from the two stage marginal method. We usually choose the estimation from two stage marginal method as initial values to obtain quick convergence of maximisation.

4.5.3 Repeated endurance exercise tests data

The exercise times 1 hour and 3 hour after the drug treatment in the repeated endurance exercise tests in Table 1.2 were fitted by a bivariate PSW model. Parameters in the model were estimated by both MLE and the two stage marginal methods. There are censored observations in the data set. An indicator variable \( I_i \) was used in calculating log-likelihood function, where \( I_i = 1 \) indicates the \( i \)th subject is observed; \( I_i = 0 \) indicates the \( i \)th subject is censored. The covariate, dose of the drug, was concerned in the PSW model. That is \( \log \alpha_{ij} = b_{0j} + b_{1j} \text{dose}(i) \), where \( i = 1, \ldots, n, j = 1, 2 \).

The estimated parameters are listed in Table 4.18, which indicate that the two estimation methods give very similar results.

4.6 Discussion

We have investigated the asymptotic properties of the MLE and the two stage marginal estimation method for the PSW model. The conclusion is that it is satisfactory to use the two stage marginal method for estimation, in the sense that the asymptotic variances of estimators from the method are very similar to those from the MLE method. The advantage of applying the two stage marginal method would become more convincing when the model has more than two variates. Messy computation in the maximisation of a high dimensional function can be avoided. At the very least, the two stage marginal method helps to set up the initial values for the maximum likelihood estimation.

Extension of the two stage marginal method to accommodate covariate information in the model is also briefly explored. As one might expect, the simulation results provide similar relative efficiencies to those obtained in the case without covariates.
Table 4.16: Fit of the PSW model on the fibre strength data: $p = 2$

<table>
<thead>
<tr>
<th>Pair</th>
<th>Method</th>
<th>$\alpha_1$</th>
<th>$\beta_1$</th>
<th>$\alpha_2$</th>
<th>$\beta_2$</th>
<th>$\nu$</th>
<th>log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y1, y2)</td>
<td>MLE</td>
<td>-6.949</td>
<td>4.725</td>
<td>-4.888</td>
<td>3.914</td>
<td>0.400</td>
<td>-91.031</td>
</tr>
<tr>
<td></td>
<td>Marginal</td>
<td>-7.146</td>
<td>4.822</td>
<td>-5.063</td>
<td>4.040</td>
<td>0.409</td>
<td>-91.136</td>
</tr>
<tr>
<td>(y1, y3)</td>
<td>MLE</td>
<td>-7.036</td>
<td>4.756</td>
<td>-6.545</td>
<td>6.725</td>
<td>0.377</td>
<td>-53.497</td>
</tr>
<tr>
<td></td>
<td>Marginal</td>
<td>-7.146</td>
<td>4.822</td>
<td>-6.798</td>
<td>6.961</td>
<td>0.384</td>
<td>-53.561</td>
</tr>
<tr>
<td>(y1, y4)</td>
<td>MLE</td>
<td>-7.080</td>
<td>4.783</td>
<td>-10.348</td>
<td>12.806</td>
<td>0.610</td>
<td>-36.432</td>
</tr>
<tr>
<td></td>
<td>Marginal</td>
<td>-7.146</td>
<td>4.822</td>
<td>-11.040</td>
<td>13.635</td>
<td>0.626</td>
<td>-36.582</td>
</tr>
<tr>
<td>(y2, y3)</td>
<td>MLE</td>
<td>-4.974</td>
<td>3.980</td>
<td>-6.259</td>
<td>6.523</td>
<td>0.392</td>
<td>-56.348</td>
</tr>
<tr>
<td></td>
<td>Marginal</td>
<td>-5.063</td>
<td>4.040</td>
<td>-6.799</td>
<td>6.961</td>
<td>0.413</td>
<td>-56.989</td>
</tr>
<tr>
<td>(y2, y4)</td>
<td>MLE</td>
<td>-5.026</td>
<td>4.013</td>
<td>-10.441</td>
<td>12.943</td>
<td>0.738</td>
<td>-41.275</td>
</tr>
<tr>
<td></td>
<td>Marginal</td>
<td>-5.063</td>
<td>4.040</td>
<td>-11.040</td>
<td>13.635</td>
<td>0.753</td>
<td>-41.384</td>
</tr>
<tr>
<td>(y3, y4)</td>
<td>MLE</td>
<td>-6.664</td>
<td>6.858</td>
<td>-10.447</td>
<td>12.928</td>
<td>0.654</td>
<td>-1.945</td>
</tr>
<tr>
<td></td>
<td>Marginal</td>
<td>-6.799</td>
<td>6.961</td>
<td>-11.040</td>
<td>13.635</td>
<td>0.671</td>
<td>-2.050</td>
</tr>
</tbody>
</table>

Table 4.17: Fit of the PSW model on the fibre strength data: $p = 4$

<table>
<thead>
<tr>
<th>Method</th>
<th>$\alpha_1$</th>
<th>$\beta_1$</th>
<th>$\alpha_2$</th>
<th>$\beta_2$</th>
<th>$\nu$</th>
<th>log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>-6.968</td>
<td>4.822</td>
<td>-4.898</td>
<td>4.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marginal</td>
<td>-7.146</td>
<td>4.822</td>
<td>-5.063</td>
<td>4.040</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>$\alpha_3$</th>
<th>$\beta_3$</th>
<th>$\alpha_4$</th>
<th>$\beta_4$</th>
<th>$\nu$</th>
<th>log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>-6.378</td>
<td>6.754</td>
<td>-9.732</td>
<td>11.982</td>
<td>0.535</td>
<td>-100.687</td>
</tr>
</tbody>
</table>

Table 4.18: Fit of the bivariate PSW model on the exercise tests data

<table>
<thead>
<tr>
<th>Method</th>
<th>$b_{01}$</th>
<th>$b_{11}$</th>
<th>$\beta_1$</th>
<th>$b_{02}$</th>
<th>$b_{12}$</th>
<th>$\beta_2$</th>
<th>$\nu$</th>
<th>log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>-10.025</td>
<td>-0.642</td>
<td>1.626</td>
<td>-11.319</td>
<td>-1.128</td>
<td>1.934</td>
<td>0.254</td>
<td>-199.008</td>
</tr>
<tr>
<td>Marginal</td>
<td>-10.136</td>
<td>-0.117</td>
<td>1.613</td>
<td>-11.523</td>
<td>-1.223</td>
<td>1.976</td>
<td>0.259</td>
<td>-199.226</td>
</tr>
</tbody>
</table>
Chapter 5

Score test for heterogeneity based on the PSW model

From the practical point of view, it is of interest to know if there is frailty in a set of multivariate survival data. In other words, it is desirable to detect whether a model based on independent Weibull random variables is adequate or if a Weibull based random effects model is more appropriate. Crowder & Kimber (1997) develop a score test for testing the heterogeneity in the GW model. They also indicate that the score test is of more general applicability: it is in fact a score test against a class of Weibull mixtures that have a one parameter mixing distribution with finite variance. However, Crowder & Kimber's test is not a score test for testing the heterogeneity in the PSW model, since the positive stable mixing distribution has infinite variance.

In the following three chapters, we explore several diagnostic methods for testing the independent Weibull case against a PSW alternative. In this chapter, a score test for detecting heterogeneity in the PSW model is derived. Its asymptotic properties are investigated. Simulations are carried out to give further insight into the features of the score statistic, including the cases with and without nuisance parameters and the possibility of censoring. Examples are given to illustrate the application of the method.

Apart from Section 5.4, where some generalised results are given, we only consider the bivariate case, $p = 2$, in this chapter. That is the model in the form

$$S(y_1, y_2) = \exp(-s^\nu),$$

(5.1)
where \( s = \xi_1 y_{11}^{\phi_1} + \xi_2 y_{12}^{\phi_2} \). Under the model, independence of the components \( Y_1 \) and \( Y_2 \) occurs when the parameter \( \nu \) tends to 1, in which case, the model degenerates to the independent Weibulls model. Hence, the null hypothesis of the heterogeneity test is \( H_0 : \nu = 1 \), and the alternative hypothesis is \( H_1 : 0 < \nu < 1 \).

### 5.1 The score test and its properties in the uncensored case

#### 5.1.1 Uncensored case without nuisance parameters

First, we consider the case in which there is no censoring, and we assume that the parameters \( \xi_j, \phi_j (j = 1, 2) \) are known. Suppose that we observe a bivariate sample of \((Y_{i1}, Y_{i2}), \ldots, (Y_{n1}, Y_{n2})\), the log-likelihood of the sample is

\[
\ell_n(\nu) = -\sum_{i=1}^{n} s_i^{\nu} + \sum_{i=1}^{n} \sum_{j=1}^{2} \log(\xi_j \phi_j y_{ij}^{\phi_j-1}) + \sum_{i=1}^{n} \log(\nu^2 s_i^{2\nu-2} - \nu(\nu - 1) s_i^{2\nu-2}),
\]

where \( s_i = \xi_1 y_{i1}^{\phi_1} + \xi_2 y_{i2}^{\phi_2} \). The first derivative of the log-likelihood is

\[
\frac{\partial \ell_n}{\partial \nu} = \sum_{i=1}^{n} \left(-s_i^{\nu} \log s_i + \frac{1}{\nu} + \log s_i + \frac{s_i^{\nu} + \nu s_i^{\nu} \log s_i - 1}{\nu s_i^{\nu} - \nu + 1}\right).
\]

Thus the corresponding score statistic is

\[
T(n) = \frac{\partial \ell_n}{\partial \nu} \bigg|_{\nu=1} = \sum_{i=1}^{n} (2 - s_i \log s_i + 2 \log s_i - \frac{1}{s_i}). \tag{5.2}
\]

The score statistic \( T(n) \) depends on the observations \((y_{i1}, y_{i2}) (i = 1, \ldots, n)\), only through the variables \( s_i \) which all have the same distribution even though the parameters \( \xi_j \) and \( \phi_j \) may vary over \( i \). Let

\[
T_i = 2 - s_i \log s_i + 2 \log s_i - \frac{1}{s_i},
\]

where \( i = 1, \ldots, n \). Thus \( T(n) \) is the sum of independent, identically distributed variates \( T_i \). This property is useful for calculating the asymptotic distribution of \( T(n) \). On the other hand, the mean and variance of \( T(n) \) can be expressed as

\[
E[T(n)] = \sum_{i=1}^{n} E(T_i), \tag{5.3}
\]

89
and

\[ \text{Var}[T(n)] = \sum_{i=1}^{n} E(T_i^2) - \sum_{i=1}^{n} E^2(T_i), \]  

(5.4)

where

\[ E(T_i) = E(2 - s_i \log s_i + 2 \log s_i - \frac{1}{s_i}) \]

\[ = 2 - E(s_i \log s_i) + 2E(\log s_i) - E\left(\frac{1}{s_i}\right), \]  

(5.5)

and

\[ E(T_i^2) = E[(2 - s_i \log s_i + 2 \log s_i - \frac{1}{s_i})^2] \]

\[ = 4 + 4E(\log s_i)^2 + E(s_i \log s_i)^2 + 10E(\log s_i) - 4E(s_i \log s_i) \]

\[ - 4E(1/s_i) - 4E(s_i^2 \log s_i) - 4E(\log s_i/s_i) + E(1/s_i^2). \]  

(5.6)

The null case

Under the null hypothesis \( H_0 : \nu = 1 \), and the assumption that \( \xi_j \) and \( \phi_j, (j = 1, 2) \), are known, the components \( y_{ij}, (j = 1, 2) \) of the \( i \)th observation are independent and Weibull distributed, so that \( \xi_j y_{ij}^\phi \) are independent unit exponential variables. Thus, \( s_i \) has a gamma distribution with shape parameter \( 2 \) and scale parameter \( 1 \). Its density function is

\[ f_0(s_i) = s_i e^{-s_i}. \]

Some expectations involving \( s_i \) and \( \log s_i \) under the null hypothesis are obtained. They are

\[ E(s_i^{-1}) = \int_0^\infty s_i^{-1} f_0(s_i) ds_i = \int_0^\infty e^{-s_i} ds_i = 1, \]

\[ E(\log s_i) = \int_0^\infty \log s_i f_0(s_i) ds_i = \Gamma'(2) = 1 - \gamma, \]

\[ E(s_i \log s_i) = \int_0^\infty s_i \log s_i f_0(s_i) ds_i = \Gamma'(3) = 3 - 2\gamma, \]

\[ E(\log^2 s_i) = \int_0^\infty \log^2 s_i f_0(s_i) ds_i = \Gamma''(2), \]

\[ E(s_i \log^2 s_i) = \int_0^\infty s_i \log^2 s_i f_0(s_i) ds_i = \Gamma''(3), \]

\[ E(s_i^2 \log^2 s_i) = \int_0^\infty s_i^2 \log^2 s_i f_0(s_i) ds_i = \Gamma''(4), \]

\[ E(s_i^{-1} \log s_i) = \int_0^\infty s_i^{-1} \log s_i f_0(s_i) ds_i = \Gamma'(1) = -\gamma \]

\[ E(s_i^{-2}) = \int_0^\infty s_i^{-2} f_0(s_i) ds_i = \int_0^\infty s_i^{-1} e^{-s_i} ds_i = \infty, \]

90
where $\Gamma(.)$ is the gamma function and $\gamma$ is Euler’s constant.

Therefore, under the null hypothesis,

\[
E[T_i] = 2 - E[s_i \log s_i] + 2E[\log s_i] - E[\frac{1}{s_i}]
\]

\[
= 2 - (3 - 2\gamma) + 2(1 - \gamma) - 1
\]

\[
= 0
\]

It follows that the expected value of the score statistic $T(n)$ is zero under the null hypothesis, i.e.

\[
E[T(n)] = 0.
\]

Each of the terms in the expression (5.6) has a finite value except for $E[1/s_i^2]$, which is infinite. Hence we have, under the null hypothesis,

\[
Var[T(n)] = \infty.
\]

Thus, the usual central limit theorem argument can not be applied to $T(n)$.

To deal with the non-regular behaviour of the score statistic, the central limit theorem for infinite variance (Feller, 1971) has been applied to obtain a standard normal test statistic.

**Lemma 1 (Central Limit Theorem for Infinite Variance)** Let $X_1, \ldots, X_n$ be a set of iid random variables, each with mean zero. Then

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = N(0,1),
\]

if, and only if, for all $\varepsilon > 0$, a sequence $C_n$ exists, with $C_n \to \infty$ and

\[
\lim_{n \to \infty} \frac{n}{C_n^2} Var(X_i; \varepsilon C_n) = 1.
\]

The notation $Var(X; \omega)$ specifies the variance of $X$ truncated at $\omega$, which is defined as

\[
Var(X; \omega) = \int_{|x|<\omega} x^2 dF(x) - \left( \int_{|x|<\omega} xdF(x) \right)^2,
\]

where $F$ is the distribution function of $X$.

The score statistic $T(n)$ is the sum of $T_1, T_2, \ldots, T_n$. Since all the terms in $T_i$ except for $\frac{1}{s_i}$ have finite variance, it is the term $\frac{1}{s_i}$ that exhibits the non-regularity
and leads to the required norming. Therefore, for $\omega \to \infty$, the leading term of $\text{Var}(T_i; \omega)$, is $\int_{|u|<\omega} u^2 dF(u_i)$, where $u_i = \frac{1}{s_i}$ with density function $f(u_i) = \frac{1}{u_i e^{-u_i}}$. Hence, we have

$$\text{Var}(T_i; \omega) \sim \int_{|u|<\omega} u^2 dF(u) = \int_0^\omega \frac{1}{u} e^{-\frac{1}{u}} du = \int_1^\infty \frac{1}{s} e^{-s} ds = \log \omega - \gamma - \sum_{k=1}^{\infty} \frac{(-1)^k}{k k!} \left(\frac{1}{\omega}\right)^k,$$

see Abramowitz et al (1972).

Let the sequence $C_n$ be defined by $C_n = \sqrt{\frac{1}{2} n \log n}$. Then, clearly $C_n \to \infty$ as $n \to \infty$. Also let $\varepsilon > 0$, then

$$\lim_{n \to \infty} \frac{n}{C_n} \text{Var}(T_i, \varepsilon C_n) = \lim_{n \to \infty} \frac{n \log \varepsilon + \frac{1}{2} \log \left(\frac{1}{2} n \log n\right) + \mathcal{O}(\log n)}{\frac{1}{2} n \log n} = 1.$$

Therefore, by Lemma 1 (central limit theorem for infinite variance), we have asymptotically

$$\sum_{i=1}^{n} \frac{T_i}{C_n} = \frac{T(n)}{\sqrt{\frac{1}{2} n \log n}} \to N(0,1), \text{ as } n \to \infty,$$

when independence of components $Y_1, Y_2$ occurs. The term $(\log n)^{\frac{1}{2}}$ in $C_n$ is slowly varying in $n$, but provides the extra scaling relative to the regular case which is necessary to obtain convergence to a normal limit. In fact, the rate of convergence to a normal limit is very slow which can be verified by simulation. This slow convergence property is also noted by Tawn (1988) and Ledford (1996). In their work, the approaches to modelling the dependence structure between extreme values are studied. The logistic model

$$S(y_1, y_2) = \exp \left\{-(y_1^r + y_2^r)^{\frac{1}{r}}\right\}, \quad r \geq 1 \quad (5.8)$$

is one of the models in their discussion. Independence occurs when $r = 1$. Let $\nu = \frac{1}{r}$. (5.8) becomes

$$S(y_1, y_2) = \exp \left\{-(y_1^{\frac{1}{\nu}} + y_2^{\frac{1}{\nu}})^{\nu}\right\}, \quad \text{for} \quad 0 \leq \nu \leq 1$$

which is the special case of the PSW model (5.1) when $\xi_1 = \xi_2 = 1$ and $\phi_1 = \phi_2 = \frac{1}{\nu}$. 

92
The non-null case

Under the alternative hypothesis $H_1 : 0 < \nu < 1$, the density function of variable $s_i$ is

$$f_1(s_i) = \exp(-s_i^\nu)\nu s_i^{\nu-1}((\nu s_i^\nu - \nu + 1). \quad (5.9)$$

The derivation of (5.9) is given in Section 7.1.1. Some expectations are obtained:

$$E\left(\frac{1}{s_i}\right) = \int_0^\infty s_i^{-1}f_1(s_i)ds_i = \nu \int_0^\infty t^{1 - \frac{1}{\nu}}\exp(-t)dt - (\nu - 1) \int_0^\infty t^{-\frac{1}{\nu}}\exp(-t)dt = -\infty,$n

$$E(s_i \log s_i) = \int_0^\infty s_i \log s_i f_1(s_i)ds_i = \Gamma\left(\frac{1}{\nu} + 1\right)\left[\frac{2}{\nu}\psi(1 + \frac{1}{\nu}) + 1\right],$$

$$E(\log s_i) = \int_0^\infty \log s_i f_1(s_i)ds_i = 1 - \frac{\gamma}{\nu},$$

where $\psi(.) = \Gamma'(.)/\Gamma(.)$ is the digamma function. Therefore, the mean of $T_{(n)}$ under $H_1$ is

$$E[T_{(n)}] = E[T_i] = -\infty.$$n

Furthermore,

$$E\left(\frac{1}{s_i^2}\right) = \int_0^\infty s_i^{-2}f_1(s_i)ds_i = \nu \int_0^\infty t^{1 - \frac{2}{\nu}}\exp(-t)dt - (\nu - 1) \int_0^\infty t^{-\frac{2}{\nu}}\exp(-t)dt = \infty.$$

We conclude that, under the alternative hypothesis $H_1 : 0 < \nu < 1$, the mean and variance of $T_{(n)}$ are both infinitive as a consequence of the $\frac{1}{s_i}$ terms. We evaluate the power of the score test $T_{(n)}$ based on a simulation study, see Section 5.3.3.

5.1.2 Uncensored case with nuisance parameters

Since $\xi_j, \phi_j$ are usually unknown in practice, a more useful test statistic is $T_{(n)}^*$, which is of the same form as $T_{(n)}$ but with $\xi_j, \phi_j$ replaced by their null maximum likelihood estimators. That is

$$T_{(n)}^* = \sum_{i=1}^n (-\hat{s}_i \log \hat{s}_i + 2 \log \hat{s}_i - \frac{1}{\hat{s}_i} + 2), \quad (5.10)$$
where \( s_i = \hat{\xi}_1 y_{i1}^{\hat{\phi}_1} + \hat{\xi}_2 y_{i2}^{\hat{\phi}_2} \) and \( \hat{\xi}_j, \hat{\phi}_j \) are the maximum likelihood estimators of \( \xi_j, \phi_j \) under \( H_0 \).

Since maximum likelihood estimation of the Weibull parameters is regular, one might anticipate that the asymptotic null distribution of \( T^*_n \) would be similar to that of \( T_n \). Crowder & Kimber (1997) have discussed the asymptotic properties for their score statistic in the case with nuisance parameters. By applying the theory of Pierce (1982), they show that, under the null hypothesis, the asymptotic distributions of \( T^*_n \) and \( T_n \) are normal with mean zero. However, the asymptotic variances are different. The variance of \( T^*_n \) is reduced when the nuisance parameters are estimated. Therefore, failure to take account of parameter estimation would lead to extremely conservative tests. The asymptotic properties of our version of \( T^*_n \) is not pursued further here. However, the null and non-null properties of the statistic \( T^*_n \) are investigated further by simulation in Section 5.3.

\[ 5.2 \quad \text{The score test and its properties in the censored case} \]

\[ 5.2.1 \quad \text{Censored case without nuisance parameters} \]

Consider now the case in which \( Y_j \ (j = 1, 2) \) might be right-censored, and the Weibull parameters \( \xi_j, \phi_j \ (j = 1, 2) \) are known. Assume that \( Y_1 \) and \( Y_2 \) are censored at fixed times \( c_1 \) and \( c_2 \), respectively. There are four possibilities for each individual observation. For a particular subject with index \( i \),

1) The likelihood contribution from \( (y_{i1}, y_{i2}) \), where \( y_{i1} \) and \( y_{i2} \) are both uncensored, is

\[ L_{i1} = \prod_{j=1}^{2} \xi_j^{\phi_j} y_{ij}^{\phi_j - 1} \left[ \nu^\phi \lambda_{ij}^{2\nu-2} - \nu(\nu - 1)s_i^\nu - 2 \right] \exp(-s_i^\nu), \]

where \( s_i = \xi_1 y_{i1}^{\phi_1} + \xi_2 y_{i2}^{\phi_2} \). The corresponding contribution to the score function under \( H_0 \) is

\[ \frac{\partial \log L_{i1}}{\partial \nu} \bigg|_{\nu=1} = 2 + 2 \log s_i - s_i \log s_i - \frac{1}{s_i}. \]

2) The likelihood contribution from \( (y_{i1}, y_{i2}) \), where \( y_{i1} \) is observed, and \( y_{i2} \) is
censored, i.e. \((y_{1i}, y_{i2}) = (y_{1i}, c_2)\), is

\[ L_{i2} = \xi_1 y_{1i}^{\phi_1 - 1} \nu s_i^{\nu-1} \exp(-s_i^\nu), \]

where \(s_i = \xi_1 y_{1i}^{\phi_1} + \xi_2 c_2^{\phi_2}\). The corresponding contribution to the score function under \(H_0\) is

\[ \frac{\partial \log L_{i2}}{\partial \nu} |_{\nu=1} = 1 + \log s_i - s_i \log s_i. \]

3) The likelihood contribution from \((y_{1i}, y_{i2})\), where \(y_{1i}\) is censored, and \(y_{i2}\) is observed, i.e. \((y_{1i}, y_{i2}) = (c_1, y_{i2})\), is

\[ L_{i3} = \xi_2 y_{i2}^{\phi_2 - 1} \nu s_i^{\nu-1} \exp(-s_i^\nu), \]

where \(s_i = \xi_1 c_1^{\phi_1} + \xi_2 y_{i2}^{\phi_2}\). The corresponding contribution to the score function under \(H_0\) is

\[ \frac{\partial \log L_{i3}}{\partial \nu} |_{\nu=1} = 1 + \log s_i - s_i \log s_i. \]

4) The likelihood contribution from \((y_{1i}, y_{i2})\), where both \(y_{1i}\) and \(y_{i2}\) are censored, i.e. \((y_{1i}, y_{i2}) = (c_1, c_2)\) is

\[ L_{i4} = \exp(-s_i^\nu), \]

where \(s_i = \xi_1 c_1^{\phi_1} + \xi_2 c_2^{\phi_2}\). The corresponding contribution to the score function under \(H_0\) is

\[ \frac{\partial \log L_{i4}}{\partial \nu} |_{\nu=1} = -s_i \log s_i. \]

Summarising the above information, we obtain the score statistic for a sample with \(n\) observations

\[ T(n), c = \frac{\partial l_n}{\partial \nu} |_{\nu=1} = \sum_{i=1}^{n} \left\{ I_i (1 + \log s_i) - \frac{I_i (I_i - 1)}{2s_i} - s_i \log s_i \right\} \quad (5.11) \]

where \(s_i = \xi_1 y_{1i}^{\phi_1} + \xi_2 y_{i2}^{\phi_2}\), and \(I_i (i = 1, \ldots, n)\) is an indicator variable: \(I_i = 0\) if \(y_{1i}\) and \(y_{i2}\) are both censored; \(I_i = 1\) if exactly one of \(y_{1i}\) and \(y_{i2}\) is censored; \(I_i = 2\) if \(y_{1i}\) and \(y_{i2}\) are both uncensored. The score test \(T(n)\) described in Section 5.1 is the special case of \(T(n), c\), where \(I_i = 2\) for all the observations. Note that, in the situation with right-censoring, \(s_i\) is bounded by \(\xi_1 y_{1i}^{\phi_1} + \xi_2 y_{i2}^{\phi_2}\).

In order to calculate the mean of \(T(n), c\), we specify the expression of \(T(n), c\) as the sum of a function \(g_i(y_{1i}, y_{i2})\). That is

\[ T(n), c = \sum_{i=1}^{n} g_i(y_{1i}, y_{i2}), \]

95
where

\[ g_i(y_{i1}, y_{i2}) = I_i(1 + \log s_i) - \frac{I_i(I_i - 1)}{2s_i} - s_i \log s_i. \]

Furthermore, because of the dependence through \( s_i \) and the identity of the distribution of \( s_i \), we omit the subscript \( i \) of \( g_i(y_{i1}, y_{i2}) \) in the discussion below. Since the component \( y_j \) is right-censored at \( c_j \), the expected value of \( g(y_1, y_2) \) with respect to a joint density \( f(y_1, y_2) \) is

\[
E[g(y_1, y_2)] = \int_0^{c_1} \int_0^{c_2} g(y_1, y_2) f(y_1, y_2) dy_1 dy_2 + \int_0^{c_1} \int_{c_2}^{\infty} g(y_1, c_2) f(y_1, y_2) dy_1 dy_2
\]

\[
+ \int_{c_1}^{\infty} \int_0^{c_2} g(c_1, y_2) f(y_1, y_2) dy_1 dy_2 + g(c_1, c_2) S(c_1, c_2).
\]

where \( S(y_1, y_2) \) is the joint survivor function of \((y_1, y_2)\).

**The null case**

Using the notation that \( z_j = \xi_j y_j^{\phi_j} \), and \( d_j = \xi_j c_j^{\phi_j} \), where \( j = 1, 2 \), we have that, under \( H_0 : \nu = 1 \), \( z_1 \) and \( z_2 \) are independent unit exponential variables, with density function

\[ f(z_j) = \exp(-z_j), \quad \text{for } z_j \leq d_j, \ j = 1, 2 \]

Then, each part of \( E[g(y_1, y_2)] \) may calculated as below:

\[
EG_1 = \int_0^{c_1} \int_0^{c_2} g(y_1, y_2) f(y_1, y_2) dy_1 dy_2
\]

\[
= \int_0^{d_1} \int_0^{d_2} \{2(1 + \log(z_1 + z_2)) - \frac{1}{z_1 + z_2} - (z_1 + z_2) \log(z_1 + z_2) \} e^{-(z_1 + z_2)} dz_1 dz_2
\]

\[
= -(d_1 + d_2) \log(d_1 + d_2) e^{-(d_1 + d_2)} + d_1 \log d_1 e^{-d_1} + d_2 \log d_2 e^{-d_2}
\]

\[
EG_2 = \int_0^{c_1} \int_{c_2}^{\infty} g(y_1, c_2) f(y_1, y_2) dy_1 dy_2
\]

\[
= \int_0^{d_1} \int_{d_2}^{\infty} \{(1 + \log(z_1 + d_2) - (z_1 + d_2) \log(z_1 + d_2)) \} e^{-(z_1 + d_2)} dz_1 dz_2
\]

\[
= (d_1 + d_2) \log(d_1 + d_2) e^{-(d_1 + d_2)} - d_2 \log d_2 e^{-d_2},
\]

\[
EG_3 = \int_{c_1}^{\infty} \int_0^{c_2} g(c_1, y_2) f(y_1, y_2) dy_1 dy_2
\]

\[
= \int_{d_1}^{\infty} \int_0^{d_2} \{(1 + \log(d_1 + z_2) - (d_1 + z_2) \log(d_1 + z_2)) \} e^{-(z_1 + z_2)} dz_1 dz_2
\]
\[ \begin{align*}
E G_4 &= g(c_1, c_2)S(c_1, c_2) = (d_1 + d_2) \log(d_1 + d_2)e^{-(d_1 + d_2)}.
\end{align*} \]

Therefore, under the null hypothesis \( H_0 \), the expected value of \( T(n,c) \) is zero, because
\[ E[T(n,c)] = \sum_{i=1}^{n} E[g(y_{i1}, y_{i2})] = n\{E G_1 + E G_2 + E G_3 + E G_4\} = 0. \]

The variance of \( T(n,c) \) may be obtained from
\[ \begin{align*}
V \text{ar}[T(n,c)] &= \sum_{i=1}^{n} E[g^2(y_{i1}, y_{i2})] \\
&= n \left\{ \int_0^{c_1} \int_0^{c_2} g^2(y_1, y_2)f(y_1, y_2)dy_1dy_2 + \int_{c_1}^{\infty} \int_0^{c_2} g^2(y_1, c_2)f(y_1, y_2)dy_1dy_2 \\
&\quad + \int_0^{c_1} \int_{c_2}^{\infty} g^2(c_1, y_2)f(y_1, y_2)dy_1dy_2 + g^2(c_1, c_2)S(c_1, c_2) \right\}.
\end{align*} \]

In detail, each part of the above may be obtained from
\[ \begin{align*}
E H_1 &= \int_0^{c_1} \int_0^{c_2} g^2(y_1, y_2)f(y_1, y_2)dy_1dy_2 \\
&= \int_0^{d_1} \int_0^{d_2} \left\{ 4 \left[ 1 + \log(z_1 + z_2) \right]^2 + \frac{1}{(z_1 + z_2)^2} \\
&\quad + (z_1 + z_2)^2 \log^2(z_1 + z_2) - 4 \left[ \frac{1}{(z_1 + z_2)} + \frac{\log(z_1 + z_2)}{z_1 + z_2} \right] \\
&\quad - 4 \left[ 1 + \log(z_1 + z_2) \right] (z_1 + z_2) \log(z_1 + z_2) - 2 \log(z_1 + z_2) \right\}e^{-(z_1 + z_2)}dz_1dz_2,
\end{align*} \]
\[ \begin{align*}
E H_2 &= \int_0^{c_1} \int_{c_2}^{\infty} g^2(y_1, c_2)f(y_1, y_2)dy_1dy_2 \\
&= \int_0^{d_1} \int_{d_2}^{\infty} \left\{ 1 + \log(z_1 + d_2) \right\}^2 + (z_1 + d_2)^2 \log^2(z_1 + d_2) \\
&\quad - 2 \left[ 1 + \log(z_1 + d_2) \right] (z_1 + d_2) \log(z_1 + d_2) \right\}e^{-(z_1 + z_2)}dz_1dz_2,
\end{align*} \]
\[ \begin{align*}
E H_3 &= \int_{d_1}^{\infty} \int_0^{d_2} g^2(c_1, y_2)f(y_1, y_2)dy_1dy_2 \\
&= \int_{d_1}^{\infty} \int_0^{d_2} \left\{ 1 + \log(d_1 + z_2) \right\}^2 + (d_1 + z_2)^2 \log^2(d_1 + z_2) \\
&\quad - 2 \left[ 1 + \log(d_1 + z_2) \right] (d_1 + z_2) \log(d_1 + z_2) \right\}e^{-(z_1 + z_2)}dz_1dz_2,
\end{align*} \]
\[EH_4 = g^2(c_1, c_2)S(c_1, c_2) = (d_1 + d_2)^2 \log^2(d_1 + d_2)e^{-(d_1+d_2)}.\]

All the above expressions are finite except for the term \[\int_0^{d_1} \int_0^{d_2} \frac{1}{z_{1}+z_{2}}e^{-(z_{1}+z_{2})}dz_{1}dz_{2}.\]
Therefore the null variance of \(T_{(n),c}\) is infinite.

Similarly to the discussion in the uncensored case, the score statistic \(T_{(n),c}\) is the sum of \(g_1(y_{11}, y_{12}), g_2(y_{21}, y_{22}), \ldots, g_n(y_{n1}, y_{n2})\). The dominant term of \(\text{Var}(T_{(n),c}; \omega)\), is \(\int_{|u|<\omega} u_i^2dF(u_i)\), where \(u_i = \frac{1}{s_i}\) with density function \(f(u_i) = \frac{1}{u_i^2}e^{-\frac{1}{u_i}}\). However, note that, here \(u_i \in (\frac{1}{d}, \infty)\), where \(d = d_1 + d_2 = \xi_1c_1^{\phi_1} + \xi_2c_2^{\phi_2}\). Hence, we have

\[\text{Var}(T_i; \omega) \sim \int_{|u|<\omega} u_i^2dF(u) = \int_{\frac{1}{d}}^{\infty} \frac{1}{s}e^{-s}ds = \log \omega + \log d + \sum_{k=1}^{\infty} \frac{(-1)^k}{kk!}[d^k - \frac{1}{\omega}].\]

Let the sequence \(C_n\) be defined by \(C_n = \sqrt{\frac{1}{2}n \log n}\) as before. Let \(\varepsilon > 0\), then

\[
\lim_{n \to \infty} \frac{n}{C_n^2} \text{Var}(T_i, \varepsilon C_n) = \lim_{n \to \infty} \frac{n[\log \varepsilon + \frac{1}{2} \log(\frac{1}{2}n \log n) + \log d + O(\log n)]}{\frac{1}{2}n \log n} = 1
\]

Applying Lemma 1, the central limit theorem for infinite variance, we obtain an asymptotic normal limit by giving \(T_{(n),c}\) the same scaling \(C_n\) as in Section 5.1. Hence,

\[
\frac{T_{(n),c}}{\sqrt{\frac{1}{2}n \log n}} \to N(0, 1), \quad \text{as } n \to \infty.
\]

### 5.2.2 Censored case with nuisance parameters

With the parameters \(\xi_j, \phi_j\) replaced by their null maximum likelihood estimators, we obtain the score statistic \(T_{(n),c}^{*}\) for the censored case with nuisance parameters. That is

\[
T_{(n),c}^{*} = \sum_{i=1}^{n} \left\{ I_i(1 + \log \hat{s}_i) - \frac{I_i(I_i - 1)}{2\hat{s}_i} - \hat{s}_i \log \hat{s}_i \right\}
\]

where \(\hat{s}_i = \xi_1y_{i1}^{\phi_1} + \xi_2y_{i2}^{\phi_2}\) and \(\hat{\xi}_j, \hat{\phi}_j\) are the maximum likelihood estimators of \(\xi_j, \phi_j\) under \(H_0\). It may be anticipated that, under the null hypothesis, \(T_{(n),c}^{*}\) has similar asymptotic properties to those of \(T_{(n),c}\). The properties of \(T_{(n),c}^{*}\) are studied by simulation in Section 5.3.
5.3 Simulation study

Simulation work was carried out on three aspects to investigate the behaviour of the score statistics with and without censoring. First, the rate of convergence to the normal limit of the score was examined. Secondly, we obtained critical values of the score tests by simulation, and investigated the influence on the critical values when there is censoring or there are nuisance parameters. The third part of the simulation work was to evaluate the power of the score test. Programs were written in Fortran77 and NAG (1995) library routines were called for generating random variables and performing maximum likelihood estimations, see Appendix A.

5.3.1 Evaluation of the rate of convergence

To investigate the rate of convergence of the normalised score test $T(n)/C_n$, we generated a set of random variables $s_i, (i = 1, \ldots, n)$ by using the NAG (1995) routine G05FFF, which have gamma distribution with scale parameter $1$ and shape parameter $2$. Assuming there are no nuisance parameters and no censoring, score statistic $T(n)$ was calculated from $s_1, \ldots, s_n$ through (5.2). Sample size $n$ was chosen varying from 100 to 100,000. For each sample size, the simulation was repeated 10,000 times. To estimate the stability of the simulated results, simulation on each sample size was done four times. The average quantiles of $\frac{T(n)}{C_n}$ at certain points ($P = 0.10, P = 0.05, P = 0.025, P = 0.01$) and corresponding standard deviations of the estimated quantiles are listed in Table 5.1, where $C_n = \sqrt{\frac{3}{2} n \log n}$. The bottom line of Table 5.1 lists the quantiles of the standard normal distribution for comparison. The results show that the rate of convergence to the normal is very slow in spite of the adjustment of $C_n$. A number of different adjusted $C_n$ values were tried to improve the convergence rate, but none of the attempts worked well. We have tried the Box-Cox transformation to normalise the behaviour of the $T(n)$, but that did not work well either. It therefore indicates that the normal approximation is very poor in finite samples.

Furthermore, to compare the performance of the convergence of the score statistics $T_{n}^*, T_{(n),c}$, and $T_{(n),c}$ with $T_{(n)}$, random variables were generated as follows. Assuming that the parameters of each Weibull distributed component are un-
Table 5.1: Normalised critical values of the score test $T_n/C_n$
(no censoring, marginal parameters known)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>$C_n$</th>
<th>Quantiles(s.d.)</th>
<th>0.10</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>15.17</td>
<td>-2.33(0.02)</td>
<td>-3.27(0.04)</td>
<td>-4.28(0.12)</td>
<td>-5.90(0.31)</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>39.42</td>
<td>-2.12(0.04)</td>
<td>-2.91(0.07)</td>
<td>-3.74(0.08)</td>
<td>-5.12(0.18)</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>58.77</td>
<td>-2.06(0.04)</td>
<td>-2.84(0.03)</td>
<td>-3.66(0.07)</td>
<td>-4.87(0.08)</td>
<td></td>
</tr>
<tr>
<td>5000</td>
<td>145.92</td>
<td>-1.95(0.04)</td>
<td>-2.67(0.07)</td>
<td>-3.42(0.04)</td>
<td>-4.46(0.05)</td>
<td></td>
</tr>
<tr>
<td>10000</td>
<td>214.60</td>
<td>-1.85(0.03)</td>
<td>-2.53(0.07)</td>
<td>-3.24(0.10)</td>
<td>-4.28(0.17)</td>
<td></td>
</tr>
<tr>
<td>50000</td>
<td>520.09</td>
<td>-1.77(0.03)</td>
<td>-2.42(0.03)</td>
<td>-3.04(0.04)</td>
<td>-4.04(0.17)</td>
<td></td>
</tr>
<tr>
<td>100000</td>
<td>758.71</td>
<td>-1.78(0.03)</td>
<td>-2.38(0.03)</td>
<td>-3.01(0.04)</td>
<td>-3.92(0.09)</td>
<td></td>
</tr>
<tr>
<td>$\infty$</td>
<td></td>
<td>-1.28</td>
<td>-1.64</td>
<td>-1.96</td>
<td>-2.33</td>
<td></td>
</tr>
</tbody>
</table>

known and that there is no censoring, we generated random variables $y_{i1}$ and $y_{i2}$, $i = 1, \ldots, n$, from two independent Weibull distributions with parameters $\xi_j = 1$ and $\phi_j = 1$ ($j = 1, 2$) by calling NAG (1995) subroutine G05DPF. Then, random variable $\hat{\phi}_i$ was obtained by $\hat{\phi}_i = \hat{\xi}_1 y_{i1}^{\hat{\phi}_1} + \hat{\xi}_2 y_{i2}^{\hat{\phi}_2}$, where $\hat{\xi}_j$, $\hat{\phi}_j$ are the maximum likelihood estimators of $\xi_j$, $\phi_j$ estimated from sample $(y_{i1}, y_{i2})$, $i = 1, \ldots, n$, under $H_0$. $T_n$ is the summation of a function of $\hat{\phi}_i$ as specified in (5.10).

Assume that $Y_1$ and $Y_2$ are censored at fixed times $c_1$ and $c_2$, respectively. For the case in which there is censoring, and the Weibull parameters are known, the population proportion of censored observations $p_c$ has the following relationship with the marginal cumulative hazards $H_1(c_1)$ and $H_2(c_2)$:

$$p_c = P(Y_1 > c_1) + P(Y_2 > c_2) - P(Y_1 > c_1, Y_2 > c_2)$$
$$= \exp(-\xi_1 c_1^{\phi_1}) + \exp(-\xi_2 c_2^{\phi_2}) - \exp(-\xi_1 c_1^{\phi_1} - \xi_2 c_2^{\phi_2})$$
$$= 1 - (1 - \exp(-d_1))(1 - \exp(-d_2)), \quad (5.13)$$

where $H_j(c_j) = d_j, j = 1, 2$. In addition, it is clear that $\exp(-H_j) = \exp(-d_j)$ ($j = 1, 2$) is the population proportion of censored observations for the component $j$.

Random variables $y_{i1}$ and $y_{i2}$ from two independent Weibull distribution were generated with parameters $\xi_j = 1$ and $\phi_j = 1$ ($j = 1, 2$). According to the above relationship, we chose fixed censoring time $c_1 = c_2 = c = 1.8$ for both components, so that, on average, the generated data has 30% censoring for one or both components of all the pairs. Then, all $y_{ij}$ ($i = 1, \ldots, n, j = 1, 2$) values
Table 5.2: Normalised critical values of the score test $T^*_n/C_n$
(no censoring, marginal parameters unknown)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>$C_n$</th>
<th>Quantiles(s.d.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td>100</td>
<td>15.17</td>
<td>-1.63(0.02)</td>
</tr>
<tr>
<td>500</td>
<td>39.42</td>
<td>-1.57(0.03)</td>
</tr>
<tr>
<td>1000</td>
<td>58.77</td>
<td>-1.52(0.02)</td>
</tr>
<tr>
<td>5000</td>
<td>145.92</td>
<td>-1.46(0.04)</td>
</tr>
<tr>
<td>10000</td>
<td>214.60</td>
<td>-1.30(0.03)</td>
</tr>
<tr>
<td>$\infty$</td>
<td></td>
<td>-1.28</td>
</tr>
</tbody>
</table>

greater than $c = 1.8$ were substituted by 1.8. Hence, the score statistics $T_{(n),c}$ were generated using (5.11). Similarly, for the case in which there is censoring and the Weibull parameters are unknown, the same procedure was used with the addition of a maximum likelihood step to obtain $T^*_{(n),c}$.

Normalised critical values of the score tests $T^*_n/C_n, T_{(n),c}/C_n, T^*_n/c/C_n$ are presented in Tables 5.2, 5.3 and 5.4, respectively. Note that, to compare the results with Table 5.1, the sample sizes are chosen from 100 to 10000. When sample size is more than 10000, it is very time consuming to perform the program in the case involving maximum likelihood estimation. The results show that the quantiles of $T^*_n$ and $T^*_{(n),c}$ are comparatively nearer zero than those of the $T_{(n)}$ and $T_{(n),c}$. The convergence to the normal for the case with nuisance parameters is slightly better than the case without nuisance parameters, no matter whether there is censoring or not. That is because the restriction has been imposed on the log-likelihood function when the nuisance parameters are estimated by maximising the log-likelihood function. This restriction, on the whole, makes the score statistic, which is from the first derivative of the log-likelihood function, have fewer extreme negative values for the case with nuisance parameters. On the other hand, Tables 5.1 - 5.4 also show that the quantiles of $T_{(n),c}$ and $T^*_{(n),c}$ at each level tend to be slightly nearer zero than those of $T_{(n)}$ and $T^*_{(n)}$, respectively. However, the rate of convergence for the score test with censored data, i.e. $T_{(n),c}$ and $T^*_{(n),c}$, is little different from that of the score test without censored data, i.e. $T_{(n)}$ and $T^*_{(n)}$.

The standard deviations of the estimated quantiles in the tables indicate reasonable precision apart from the ones on the extreme left tail (0.01 quantiles).
Table 5.3: Normalised critical values of the score test $T_{(n),c}/C_n$
(with 30% censoring, marginal parameters known)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>$C_n$</th>
<th>Quantiles(s.d.)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.10</td>
<td>0.05</td>
<td>0.025</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>15.17</td>
<td>-2.04(0.05)</td>
<td>-2.97(0.08)</td>
<td>-3.99(0.06)</td>
<td>-5.68(0.13)</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>39.42</td>
<td>-1.94(0.04)</td>
<td>-2.70(0.06)</td>
<td>-3.56(0.04)</td>
<td>-4.89(0.12)</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>58.77</td>
<td>-1.86(0.02)</td>
<td>-2.62(0.04)</td>
<td>-3.42(0.09)</td>
<td>-4.75(0.15)</td>
<td></td>
</tr>
<tr>
<td>5000</td>
<td>145.92</td>
<td>-1.79(0.03)</td>
<td>-2.44(0.05)</td>
<td>-3.18(0.07)</td>
<td>-4.27(0.18)</td>
<td></td>
</tr>
<tr>
<td>10000</td>
<td>214.60</td>
<td>-1.73(0.04)</td>
<td>-2.39(0.03)</td>
<td>-3.04(0.06)</td>
<td>-4.11(0.07)</td>
<td></td>
</tr>
<tr>
<td>$\infty$</td>
<td></td>
<td>-1.28</td>
<td>-1.64</td>
<td>-1.96</td>
<td>-2.33</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4: Normalised critical values of the score test $T^*_{(n),c}/C_n$
(with 30% censoring, marginal parameters unknown)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>$C_n$</th>
<th>Quantiles(s.d.)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.10</td>
<td>0.05</td>
<td>0.025</td>
<td>0.01</td>
</tr>
<tr>
<td>100</td>
<td>15.17</td>
<td>-1.62(0.04)</td>
<td>-2.34(0.05)</td>
<td>-3.15(0.11)</td>
<td>-4.67(0.11)</td>
</tr>
<tr>
<td>500</td>
<td>39.42</td>
<td>-1.55(0.01)</td>
<td>-2.22(0.04)</td>
<td>-2.99(0.05)</td>
<td>-4.34(0.20)</td>
</tr>
<tr>
<td>1000</td>
<td>58.77</td>
<td>-1.53(0.01)</td>
<td>-2.20(0.01)</td>
<td>-2.95(0.05)</td>
<td>-4.36(0.16)</td>
</tr>
<tr>
<td>5000</td>
<td>145.92</td>
<td>-1.48(0.02)</td>
<td>-2.08(0.01)</td>
<td>-2.73(0.02)</td>
<td>-3.95(0.14)</td>
</tr>
<tr>
<td>10000</td>
<td>214.60</td>
<td>-1.37(0.03)</td>
<td>-1.80(0.02)</td>
<td>-2.50(0.04)</td>
<td>-3.32(0.06)</td>
</tr>
<tr>
<td>$\infty$</td>
<td></td>
<td>-1.28</td>
<td>-1.64</td>
<td>-1.96</td>
<td>-2.33</td>
</tr>
</tbody>
</table>
Table 5.5: Estimated coefficients in the model $CT = c_0 + c_1 n + c_2 n^2$

<table>
<thead>
<tr>
<th>$P$</th>
<th>coefficients</th>
<th>$T(n)$</th>
<th>$T^*(n)$</th>
<th>$T(n),c$</th>
<th>$T^*(n),c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>$c_0$</td>
<td>-20.267</td>
<td>-15.153</td>
<td>-19.370</td>
<td>-8.604</td>
</tr>
<tr>
<td></td>
<td>$c_1$</td>
<td>-1.079</td>
<td>-0.874</td>
<td>-1.007</td>
<td>-0.936</td>
</tr>
<tr>
<td></td>
<td>$c_2$</td>
<td>$3.99 \times 10^{-3}$</td>
<td>$2.29 \times 10^{-3}$</td>
<td>$3.44 \times 10^{-3}$</td>
<td>$3.08 \times 10^{-3}$</td>
</tr>
<tr>
<td>0.025</td>
<td>$c_0$</td>
<td>-13.550</td>
<td>-8.323</td>
<td>-12.550</td>
<td>-7.318</td>
</tr>
<tr>
<td></td>
<td>$c_1$</td>
<td>-0.789</td>
<td>-0.659</td>
<td>-0.687</td>
<td>-0.590</td>
</tr>
<tr>
<td></td>
<td>$c_2$</td>
<td>$2.84 \times 10^{-3}$</td>
<td>$2.63 \times 10^{-3}$</td>
<td>$2.24 \times 10^{-3}$</td>
<td>$1.74 \times 10^{-3}$</td>
</tr>
<tr>
<td>0.05</td>
<td>$c_0$</td>
<td>-9.748</td>
<td>-7.255</td>
<td>-8.222</td>
<td>-5.823</td>
</tr>
<tr>
<td></td>
<td>$c_1$</td>
<td>-0.604</td>
<td>-0.394</td>
<td>-0.521</td>
<td>-0.429</td>
</tr>
<tr>
<td></td>
<td>$c_2$</td>
<td>$2.12 \times 10^{-3}$</td>
<td>$1.03 \times 10^{-3}$</td>
<td>$1.72 \times 10^{-3}$</td>
<td>$1.33 \times 10^{-3}$</td>
</tr>
<tr>
<td>0.10</td>
<td>$c_0$</td>
<td>-6.140</td>
<td>-4.762</td>
<td>-5.020</td>
<td>-4.247</td>
</tr>
<tr>
<td></td>
<td>$c_1$</td>
<td>-0.439</td>
<td>-0.280</td>
<td>-0.351</td>
<td>-0.292</td>
</tr>
<tr>
<td></td>
<td>$c_2$</td>
<td>$1.55 \times 10^{-3}$</td>
<td>$7.53 \times 10^{-4}$</td>
<td>$1.08 \times 10^{-3}$</td>
<td>$9.19 \times 10^{-4}$</td>
</tr>
</tbody>
</table>

5.3.2 Evaluation of critical values

Since the asymptotic null distribution of the score statistic is a poor approximation for realistic sample sizes, we therefore focus on estimating critical values of the score statistic for realistic sample sizes by simulation. The approach to generating the score statistics $T(n)$, $T^*(n)$, $T(n),c$ and $T^*(n),c$ is described in Section 5.3.1. For various sample sizes, we estimate the selected quantiles ($P = 0.01, P = 0.025, P = 0.05, P = 0.10$) from the null distribution of $T(n)$, $T^*(n)$, $T(n),c$ and $T^*(n),c$. We found that, approximately, the estimated critical values at each significance level increase quadratically with sample size up to $n = 100$. Figure 5.1 shows that, at least for $n \leq 100$, quadratic regression models can be used to fit the estimated critical values of $T(n)$. Similar results hold for $T^*(n)$, $T(n),c$ and $T^*(n),c$. We list the estimated coefficients in the regression model $CT = c_0 + c_1 n + c_2 n^2$ corresponding to each score test $T(n)$, $T^*(n)$, $T(n),c$ and $T^*(n),c$, and each significance level, where $C_T$ is the critical value, $n$ is the sample size, and $T(n),c$ and $T^*(n),c$ refer to the case in which data has about 30% censoring, see Table 5.5. Note that the critical values showed in Figure 5.1 and Table 5.5 both refer to the un-normalised score, which results in different values from those in Tables 5.1 - Table 5.4.

Simulation was also used to investigate the variation of the statistics $T(n),c$ and $T^*(n),c$ in terms of the change of the censoring point. For the case in which there are no nuisance parameters, we generated $T(n),c$ with Weibull parameter $\xi_1 = \xi_2 =$
Figure 5.1: Estimated critical values of un-normalised score statistic $T_{(n)}$ for various sample sizes with superimposed quadratic fit

$\phi_1 = \phi_2 = 1$ and censoring point $c$ varying from 0.2 to 6 for sample sizes 100, 500, 1000 and 5000. Critical values of the normalised $T_{(n),c}$ at significant level $P = 0.05$ for various censoring points are plotted in Figure 5.2, which indicates that the critical value decreases as the censoring point increases. Since an increase of the censoring point tends to increase of the percentage of uncensored observations, the critical value is decreasing with respect to the decrease of the degree of censoring.

The pattern of the change according to the censoring point is found by fitting nonlinear models. We found that, for different sample sizes, the relationship between the critical value and the censoring point may be approximated by the form $C_T = a_0 + \frac{a_1}{c-a_2}$, where $C_T$ is the critical value, $c$ is the censoring point. Estimated coefficients $a_0, a_1$ and $a_2$ are listed in Table 5.6. The fitted curves, see Figure 5.2, each approach an asymptote as the censoring point increases, which is the limiting case of no censoring. Since 99.5% of observations are uncensored when $c = 6$, the critical value near this point is very close to that obtained for $T_{(n)}$ at the corresponding sample size.
Figure 5.2: The variation of critical values of the normalised version of $T_{n,c}$ in terms of censoring points

Similarly, for the case in which there are nuisance parameters, the simulated results show that the decrease of the critical values in terms of the increase of the censoring point follows a similar pattern as in the parameters known case. However, the decrease of the critical values is much slower than that in the parameters known case. Simulation in this case was carried out for sample sizes 100, 500 and 1000 only. The corresponding plot and fitted lines are shown in Figure 5.3. The estimated coefficients fitting the non-linear model $CT = a_0 + \frac{a_1}{c-a_2}$ are listed in Table 5.6.

The simulation for the parameters unknown case was carried out assuming $\xi_j = \phi_j = 1$. In practice, we usually only know the fixed censoring time but have no knowledge of $\xi_j$, $\phi_j$. Therefore, Figure 5.3 is only applicable when we assume the two marginals are the same, i.e. $\xi_1 = \xi_2 = \xi$, $\phi_1 = \phi_2 = \phi$. For example, assume $Y_1$ and $Y_2$ are censored at fixed time $c$, the censoring point $d$ corresponding to the Figure 5.3 might be estimated from $d = \hat{\xi} c \hat{\phi}$, where $\hat{\xi}$ and $\hat{\phi}$ are the maximum likelihood estimates of Weibull parameters $\xi$ and $\phi$. Alternatively, $\exp(-d_j)$ may be estimated by the mean of the sample proportion of censored values in each
Figure 5.3: The variation of critical values of the normalised version of $T_{(n),c}^*$ in terms of censoring points.

Table 5.6: Estimated coefficients in the model $C_T = a_0 + \frac{a_1}{c-a_2}$

<table>
<thead>
<tr>
<th>Score Statistic</th>
<th>Sample Size</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{(n),c}$</td>
<td>100</td>
<td>$a_0$</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-3.78</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>-3.00</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>-2.90</td>
</tr>
<tr>
<td>$T_{(n),c}^*$</td>
<td>100</td>
<td>-2.39</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-2.25</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>-2.09</td>
</tr>
</tbody>
</table>
From Figure 5.2 and Figure 5.3, we can see that, when data has less than 50% censoring (c > 1.2), each fitted curve declines more slowly as the censoring point increases, especially for the parameters unknown case. This subtle variation is relatively minor in practical terms. We suggest using the critical values from $T(n)$ for use in the censored case instead of working out the critical values for all the different censoring cases, if the percentage of censoring is low enough (less than 50%). However, we will lose some power by doing this, since such tests will necessarily be slightly conservative.

5.3.3 Evaluation of power

Power for detecting heterogeneity based on the PSW model

We performed simulations to estimate the power of the score statistics $T_n$, $T_{(n)}$, $T_{(n),c}$ and $T_{(n),c}^*$ for various sample sizes. We generated random variables $x_{i1}$ and
$x_{i2}$ from two independent Weibull distributions with parameters $\xi_j = 1$ and $\phi_j = 1$ ($j = 1, 2$), and random variable $w_i$ from a positive stable distribution with characteristic exponent $\nu$, $i = 1, \ldots, n$. The simulation method of generating stable random variables is based on the algorithm of Chambers and Stuck (1976). Then, the random variable $y_{i1}$ and $y_{i2}$ based on the PSW were obtained from $y_{ij} = x_{ij}/w_i$, $(i = 1, \ldots, n, j = 1, 2)$. In the simulations, $\nu$ was chosen to vary from 0.5 to 0.99 as in Tables 5.7 and 5.8. Each simulated case for each sample size was repeated 10,000 times. The estimated power of the score statistics $T(n)$, $T^*_n$, $T_{(n),c}$ and $T^*_{(n),c}$ for various sample sizes are listed in Tables 5.7 and 5.8 for comparison. For sample sizes $n \geq 100$, the critical values at significance level $P = 0.05$ for the test statistics $T(n)$, $T^*_n$, $T_{(n),c}$ and $T^*_{(n),c}$ are taken from Tables 5.1 - 5.4, whereas the corresponding critical values are obtained from Table 5.5 when sample sizes $n \leq 100$. The results in Tables 5.7 and 5.8 show that even for small sample sizes ($n \geq 10$), the power of the score test for detecting the heterogeneity in the PSW model is reasonable. The power of the score test without censoring is slightly higher than that with censoring, with and without nuisance parameters. On the other hand, the power of the score test without nuisance parameters is higher than that with nuisance parameters, both with and without censoring. This is caused by lack of marginal parameter information. Figure 5.4 shows the estimated power curve of the score statistic $T(n)$ for $P = 0.05$ for various values of $\nu$ and various sample sizes.

**Power for detecting heterogeneity based on the GW model**

We also investigated the power of the score tests in the situation that a sample of observations is thought to be from a PSW model, but it is in fact from a GW model. This situation might occur in practice. Simulation was carried out to examine the power of our score tests to detect heterogeneity in the GW model.

Random variables $(x_{i1}, x_{i2})$ were generated from two independent Weibull distributions with parameters $\xi_j = 1$ and $\phi_j = 1$, ($j = 1, 2$). Without loss of generality, random variables $w_i$ were generated from a gamma distribution with mean one and shape parameter $\nu$. Then, bivariate random variables $(y_{i1}, y_{i2})$ from
Table 5.7: Estimated power of score tests $T_{(n)}$ and $T^*_n$ for different sample sizes and $\nu = 0.05$ (with no censoring)

<table>
<thead>
<tr>
<th>sample size</th>
<th>$\nu$</th>
<th>$T_{(n)}$ power(%)</th>
<th>$T^*_n$ power(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.50</td>
<td>98.39</td>
<td>63.10</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>91.90</td>
<td>44.47</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>77.64</td>
<td>33.71</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>50.97</td>
<td>21.80</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>21.95</td>
<td>11.15</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>11.60</td>
<td>5.80</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>4.30</td>
<td>4.69</td>
</tr>
<tr>
<td>20</td>
<td>0.50</td>
<td>99.98</td>
<td>91.88</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>92.29</td>
<td>80.75</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>93.91</td>
<td>58.77</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>71.73</td>
<td>39.94</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>31.72</td>
<td>19.64</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>14.57</td>
<td>12.45</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>7.50</td>
<td>6.81</td>
</tr>
<tr>
<td>30</td>
<td>0.50</td>
<td>100.00</td>
<td>98.80</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>99.97</td>
<td>90.19</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>98.55</td>
<td>75.39</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>85.06</td>
<td>48.93</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>42.81</td>
<td>24.31</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>19.29</td>
<td>13.03</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>6.74</td>
<td>7.30</td>
</tr>
<tr>
<td>40</td>
<td>0.50</td>
<td>100.00</td>
<td>99.81</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>99.98</td>
<td>96.93</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>99.60</td>
<td>85.78</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>92.47</td>
<td>60.98</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>51.72</td>
<td>29.94</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>22.42</td>
<td>16.44</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>6.77</td>
<td>7.42</td>
</tr>
<tr>
<td>50</td>
<td>0.50</td>
<td>100.00</td>
<td>99.97</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>100.00</td>
<td>99.02</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>99.93</td>
<td>91.52</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>96.50</td>
<td>67.29</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>64.64</td>
<td>33.57</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>32.34</td>
<td>17.29</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>10.71</td>
<td>7.38</td>
</tr>
<tr>
<td>1000</td>
<td>0.50</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>100.00</td>
<td>99.89</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>97.16</td>
<td>85.28</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>17.79</td>
<td>20.02</td>
</tr>
</tbody>
</table>
Table 5.8: Estimated power of score tests $T_{(n),c}$ and $T^*_{(n),c}$ for different sample sizes and $P = 0.05$ (with 30% censoring)

<table>
<thead>
<tr>
<th>sample size</th>
<th>$\nu$</th>
<th>power(%) $T_{(n),c}$</th>
<th>power(%) $T^*_{(n),c}$</th>
<th>sample size</th>
<th>$\nu$</th>
<th>power(%) $T_{(n),c}$</th>
<th>power(%) $T^*_{(n),c}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.50</td>
<td>90.07  54.01</td>
<td></td>
<td>100</td>
<td>0.50</td>
<td>100.00  100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>78.16  34.56</td>
<td></td>
<td></td>
<td>0.60</td>
<td>100.00  97.89.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>60.59  27.78</td>
<td></td>
<td></td>
<td>0.70</td>
<td>99.99  90.96</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>39.82  19.44</td>
<td></td>
<td></td>
<td>0.80</td>
<td>97.43  80.68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>20.16  9.76</td>
<td></td>
<td></td>
<td>0.90</td>
<td>66.55  50.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>12.49  5.30</td>
<td></td>
<td></td>
<td>0.95</td>
<td>32.30  21.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>6.48   4.32</td>
<td></td>
<td></td>
<td>0.99</td>
<td>8.85   7.16</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.50</td>
<td>98.97  86.70</td>
<td></td>
<td>200</td>
<td>0.50</td>
<td>100.00  100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>93.98  78.98</td>
<td></td>
<td></td>
<td>0.60</td>
<td>100.00  100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>80.69  55.80</td>
<td></td>
<td></td>
<td>0.70</td>
<td>100.00  93.41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>57.01  37.62</td>
<td></td>
<td></td>
<td>0.80</td>
<td>99.92  89.76</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>27.41  15.65</td>
<td></td>
<td></td>
<td>0.90</td>
<td>87.27  57.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>14.17  11.11</td>
<td></td>
<td></td>
<td>0.95</td>
<td>48.65  29.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>6.27   6.46</td>
<td></td>
<td></td>
<td>0.99</td>
<td>10.96  7.96</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.50</td>
<td>99.88  90.96</td>
<td></td>
<td>300</td>
<td>0.50</td>
<td>100.00  100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>98.34  83.57</td>
<td></td>
<td></td>
<td>0.60</td>
<td>100.00  100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>91.19  65.92</td>
<td></td>
<td></td>
<td>0.70</td>
<td>100.00  100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>67.99  38.18</td>
<td></td>
<td></td>
<td>0.80</td>
<td>100.00  90.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>33.71  21.17</td>
<td></td>
<td></td>
<td>0.90</td>
<td>95.96  63.55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>17.20  12.28</td>
<td></td>
<td></td>
<td>0.95</td>
<td>61.59  35.78</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>6.61   6.90</td>
<td></td>
<td></td>
<td>0.99</td>
<td>13.74  8.96</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.50</td>
<td>100.00 98.40</td>
<td></td>
<td>400</td>
<td>0.50</td>
<td>100.00  100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>99.78  90.95</td>
<td></td>
<td></td>
<td>0.60</td>
<td>100.00  100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>96.27  80.48</td>
<td></td>
<td></td>
<td>0.70</td>
<td>100.00  100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>79.90  55.16</td>
<td></td>
<td></td>
<td>0.80</td>
<td>100.00  97.90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>42.47  27.73</td>
<td></td>
<td></td>
<td>0.90</td>
<td>98.41  85.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>21.28  14.36</td>
<td></td>
<td></td>
<td>0.95</td>
<td>70.02  40.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>8.38   6.74</td>
<td></td>
<td></td>
<td>0.99</td>
<td>14.30  11.15</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.50</td>
<td>100.00 99.98</td>
<td></td>
<td>500</td>
<td>0.50</td>
<td>100.00  100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>99.92  95.68</td>
<td></td>
<td></td>
<td>0.60</td>
<td>100.00  100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>98.36  89.28</td>
<td></td>
<td></td>
<td>0.70</td>
<td>100.00  100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>84.86  50.83</td>
<td></td>
<td></td>
<td>0.80</td>
<td>100.00  100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>46.34  31.23</td>
<td></td>
<td></td>
<td>0.90</td>
<td>99.60  91.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>23.32  15.42</td>
<td></td>
<td></td>
<td>0.95</td>
<td>76.13  58.59</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>8.17   7.03</td>
<td></td>
<td></td>
<td>0.99</td>
<td>15.19  12.62</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>0.50</td>
<td>100.00 100.00</td>
<td></td>
<td></td>
<td>0.50</td>
<td>100.00  100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>100.00 100.00</td>
<td></td>
<td></td>
<td>0.60</td>
<td>100.00  100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>100.00 100.00</td>
<td></td>
<td></td>
<td>0.70</td>
<td>100.00  100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>100.00 100.00</td>
<td></td>
<td></td>
<td>0.80</td>
<td>100.00  100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>100.00 95.00</td>
<td></td>
<td></td>
<td>0.90</td>
<td>100.00  95.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>94.45  79.53</td>
<td></td>
<td></td>
<td>0.95</td>
<td>94.45  79.53</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>22.08  18.46</td>
<td></td>
<td></td>
<td>0.99</td>
<td>22.08  18.46</td>
<td></td>
</tr>
</tbody>
</table>
a GW model were obtained from $y_{ij} = x_{ij}/w_i$, $i = 1, \ldots, n$, $j = 1, 2$. Table 5.9 shows the estimated powers of the score tests $T(n)$ and $T^*(n)$ for sample sizes $n = 20, 50, 100, 500$, each based on 10,000 replications. Parameter $\nu$ was chosen as 1.0, 2.0, 4.0, 8.0, 16.0, 50.0.

The results indicate that the score tests still have good powers even if the frailty arises from the GW distribution. The score test are relatively insensitive to model misspecification.

Kimber (1998) proposed a test based on a decomposition of a score test (Crowder & Kimber, 1997) into marginal and associational terms. The associational term, which is actually a measure of covariance between components on the cumulative hazard scale, is designed as a test statistic to detect the presence of heterogeneity. In the case of bivariate Weibull based random effects models, this test statistic is

$$T_{(1,2)} = n^{-1} \sum_{i=1}^{n} (\xi_1 y_{i1}^{\phi_1} - 1)(\xi_2 y_{i2}^{\phi_2} - 1),$$

where parameters $\xi_j$ and $\phi_j$ ($j = 1, 2$) are known. Under the null hypothesis, $\sqrt{n}T_{(1,2)}$ asymptotically has a standard normal distribution. The corresponding test statistic $T^*_{(1,2)}$ in the case with unknown Weibull parameters is defined as

$$T^*_{(1,2)} = n^{-1} \sum_{i=1}^{n} (\hat{\xi}_1 y_{i1}^{\hat{\phi}_1} - 1)(\hat{\xi}_2 y_{i2}^{\hat{\phi}_2} - 1),$$

where $\hat{\xi}_j$ and $\hat{\phi}_j$ ($j = 1, 2$) are the null maximum likelihood estimates of the Weibull parameters. The asymptotic null distribution of $\sqrt{n}T^*_{(1,2)}$ remains the same as that of $\sqrt{n}T_{(1,2)}$, making this test easy to use.

Kimber (1998) also investigated the powers of his proposed test by a simulation experiment. In the case with no censoring, the estimated powers of Kimber’s test $T^*_{(1,2)}$ at $P = 0.05$ for the bivariate PSW model with sample size $n = 50$ and selected values of $\nu = 0.5, 0.6, 0.7, 0.8, 0.9$ are 99%, 91%, 69%, 41%, and 18% respectively. In contrast, the estimated powers of our score test $T^*_n(n)$ for the same sample size and corresponding $\nu$ values from Table 5.7 are 99.97%, 99.02%, 91.52%, 67.29% and 33.57%, which indicate that $T^*_n(n)$ has higher power than that of Kimber’s test. Conversely, simulation results also show that our score test has less power than
Table 5.9: Estimated power of score tests $T(n)$, $T^*_n$ and $T^*_{(1,2)}$ when data are from a GW model with gamma shape parameter $\nu$ and $P = 0.05$

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$\nu$</th>
<th>$T(n)$</th>
<th>$T^*_n$</th>
<th>$T^*_{(1,2)}$</th>
<th>Sample size</th>
<th>$\nu$</th>
<th>$T(n)$</th>
<th>$T^*_n$</th>
<th>$T^*_{(1,2)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>1.0</td>
<td>99.32</td>
<td>35.71</td>
<td>79</td>
<td>100</td>
<td>1.0</td>
<td>100.00</td>
<td>98.52</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>87.93</td>
<td>15.64</td>
<td>51</td>
<td></td>
<td>2.0</td>
<td>100.00</td>
<td>63.21</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>4.0</td>
<td>53.70</td>
<td>9.36</td>
<td>27</td>
<td></td>
<td>4.0</td>
<td>96.30</td>
<td>20.05</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>8.0</td>
<td>22.46</td>
<td>8.03</td>
<td>14</td>
<td></td>
<td>8.0</td>
<td>57.83</td>
<td>8.08</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>16.0</td>
<td>11.23</td>
<td>5.94</td>
<td>9</td>
<td></td>
<td>16.0</td>
<td>21.55</td>
<td>7.32</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>50.0</td>
<td>5.92</td>
<td>4.91</td>
<td>-</td>
<td></td>
<td>50.0</td>
<td>7.70</td>
<td>5.94</td>
<td>-</td>
</tr>
<tr>
<td>50</td>
<td>1.0</td>
<td>100.00</td>
<td>82.40</td>
<td>98</td>
<td>500</td>
<td>1.0</td>
<td>100.00</td>
<td>100.00</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>98.98</td>
<td>28.98</td>
<td>79</td>
<td></td>
<td>2.0</td>
<td>100.00</td>
<td>99.90</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>4.0</td>
<td>76.89</td>
<td>10.10</td>
<td>45</td>
<td></td>
<td>4.0</td>
<td>100.00</td>
<td>71.05</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>8.0</td>
<td>31.51</td>
<td>5.22</td>
<td>23</td>
<td></td>
<td>8.0</td>
<td>98.42</td>
<td>20.98</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>16.0</td>
<td>11.81</td>
<td>4.25</td>
<td>12</td>
<td></td>
<td>16.0</td>
<td>56.97</td>
<td>9.81</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>50.0</td>
<td>5.67</td>
<td>3.51</td>
<td>-</td>
<td></td>
<td>50.0</td>
<td>13.25</td>
<td>6.03</td>
<td>-</td>
</tr>
</tbody>
</table>

that of Kimber's test in the case of gamma frailty. Table 5.9 lists the estimated powers of Kimber's test $T^*_{(1,2)}$ at $P = 0.05$ for selected sample sizes and $\nu$ values, which is for comparison with our score test $T^*_n$.

5.4 Extension of the score test

5.4.1 Multivariate case

Trivariate case

We consider the extension of the score test to the trivariate case, $p = 3$. In this case, the joint survivor function for the PSW model is

$$S(y_1, y_2, y_3) = \exp(-s^\nu),$$

where $s = \sum_{j=1}^{3} \xi_j \phi_j^{y_j}$. The null and alternative hypotheses for the score test are the same as for the bivariate case, that is $H_0 : \nu = 1$, and $H_1 : 0 < \nu < 1$.

For the case in which there is no censoring and parameters $\xi_j$ and $\phi_j$ ($j = 1, 2, 3$)
are known, the log-likelihood of a sample $y_{ij}$, $i = 1, \ldots, n$, $j = 1, 2, 3$, is

$$l_n(\nu) = -\sum_{i=1}^{n} s_i^\nu + \sum_{i=1}^{n} \sum_{j=1}^{3} \log(\xi_j \phi_j y_{ij}^{\phi_j-1})$$

$$+ \sum_{i=1}^{n} \log[\nu^2 s_i^{3\nu-3} - (\nu - 1)(\nu - 2)s_i^{\nu-3} - 3\nu(\nu - 1)s_i^{2\nu-3}] + n \log \nu,$$

where $s_i = \sum_{j=1}^{3} \xi_j y_{ij}^{\phi_j}$. Hence the corresponding score statistic $T_{(n)}$ is obtained by evaluating the first derivative of the log-likelihood at $\nu = 1$, giving

$$T_{(n)} = \sum_{i=1}^{n} \left(3 - s_i \log s_i + 3 \log s_i - \frac{3}{s_i} - \frac{1}{s_i^2}\right). \quad (5.14)$$

The score statistic $T_{(n)}$ depends on the observations only through the variables $s_i$. Under the null hypothesis $H_0 : \nu = 1$, the $s_i$ are independent Gamma variables with shape parameter 3 and scale parameter 1. The density function of $s_i$ is

$$f_0(s_i) = \frac{s_i^{2e^{-s_i}}}{\Gamma(3)}.$$

Let $T_i = 3 - s_i \log s_i + 3 \log s_i - \frac{3}{s_i} - \frac{1}{s_i^2}$. The expected value of $T_i$ is

$$E(T_i) = 3 - E(s_i \log s_i) + 3E(\log s_i) - E\left(\frac{3}{s_i}\right) - E\left(\frac{1}{s_i^2}\right)$$

$$= 3 - 3\psi(4) - 3 \times \frac{1}{2} - \frac{1}{2} + 3\psi(3)$$

$$= 0.$$

Therefore the expected value of the score statistic $T_{(n)}$ is zero.

The variance of $T_i$ is

$$Var(T_i) = E(T_i^2) = E(3 - s_i \log s_i + 3 \log s_i - \frac{3}{s_i} - \frac{1}{s_i^2})^2$$

$$= E(9 + s_i^2 \log^2 s_i + \frac{3}{s_i^2} + \frac{1}{4} + 9 \log^2 s_i - 6s_i \log s_i - \frac{18}{s_i}$$

$$+ 24 \log s_i - 16 \frac{\log s_i}{s_i} - 6s_i \log^2 s_i + \frac{6}{s_i^3} - 6 \frac{\log s_i}{s_i^2}).$$

However, in the above expression, $E(\frac{1}{s_i^2}) = \infty$ so that $Var(T_i) = \infty$. Therefore the variance of $T_{(n)}$ is infinite. Analogous to the bivariate case, we apply Lemma 1 (central limit theorem for infinite variance), and obtain the null asymptotic distribution of $T_{(n)}$ as follows,

$$\frac{T_{(n)}}{\sqrt{\frac{3}{2}n \log n}} \rightarrow N(0, 1), \text{ as } n \rightarrow \infty.$$
Table 5.10: Normalised critical values of the score test $T(n)/C_n$
(trivariate case, marginal parameters are known)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>$C_n$</th>
<th>Quantiles(s.d.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td>100</td>
<td>26.28</td>
<td>-1.79</td>
</tr>
<tr>
<td>500</td>
<td>68.27</td>
<td>-1.85</td>
</tr>
<tr>
<td>1000</td>
<td>101.79</td>
<td>-1.85</td>
</tr>
<tr>
<td>5000</td>
<td>252.70</td>
<td>-1.90</td>
</tr>
<tr>
<td>10000</td>
<td>371.69</td>
<td>-2.07</td>
</tr>
<tr>
<td>$\infty$</td>
<td>-1.28</td>
<td>-1.64</td>
</tr>
</tbody>
</table>

A simulation was performed to investigate the rate of convergence of the normalised score test $T(n)/C_n$ in the trivariate case, where $T(n)$ is defined in (5.14) and $C_n = \sqrt{\frac{3}{2}n \log n}$. Sample size $n$ was chosen from 100 to 10,000. The simulation for each sample size was based on 10,000 replications. Random variables $y_{ij}$, $(i = 1, \ldots, n, j = 1, 3)$ were generated from independent Weibull distributions with parameters $\xi_j = 1$ and $\phi_j = 1$ by calling NAG (1995) subroutine G05DPF. The rate of convergence of normalised score test $T(n)/C_n$ is shown in Table 5.10, which indicates that convergence is very slow.

Similarly, the corresponding score statistic in the case with censoring may be obtained:

$$T_{(n),c} = \sum_{i=1}^{n} \left\{ I_i (1 + \log s_i) - s_i \log s_i - \frac{I_i (I_i - 1)}{2s} - \frac{I_i (I_i - 1)(I_i - 2)}{6s_i^2} \right\},$$  \hspace{1cm} (5.15)

where $I_i$ is a variable indicating the number of uncensored components in the tri-variate observation with index $i$.

In the case in which the parameters $\xi_j$ and $\phi_j$ ($j = 1, 2, 3$) are unknown, corresponding statistics can be obtained by replacing parameters $\xi_j$ and $\phi_j$ ($j = 1, 2, 3$) with their maximum likelihood estimators under $H_0$ in equations (5.14) and (5.15), respectively.
The behaviour of these score statistics may be investigated by simulation work. Corresponding critical values of the score test in different situations can be obtained by simulation.

The $p$-variate case

The score statistic for the general $p$ variate case may be derived. For the case in which there is no censoring,

$$
T_{(n)} = \sum_{i=1}^{n} \left\{ p(1 + \log s_i) - s_i \log s_i - \frac{1}{s_i^{k+1}} \sum_{k=0}^{p-2} k! C_p^{k+2} \frac{1}{s_i^{k+1}} \right\}.
$$

For the case in which there is censoring,

$$
T_{(n),c} = \begin{cases} 
\sum_{i=1}^{n} \left\{ I_i(1 + \log s_i) - s_i \log s_i - \frac{1}{s_i^{k+1}} \sum_{k=0}^{I_i-2} k! C_{I_i}^{k+2} \frac{1}{s_i^{k+1}} \right\} & \text{as } I_i \geq 2 \\
\sum_{i=1}^{n} \left\{ I_i(1 + \log s_i) - s_i \log s_i \right\} & \text{as } I_i = 0, 1
\end{cases}
$$

where $I_i$ is a variable indicating the number of uncensored components for the $p$-variate observation with index $i$.

In general, we have, under $H_0$, $E(T_{(n)}) = 0$ and $Var(T_{(n)}) = \infty$ for the $p$-variate case because of the $s_i^{-(p-1)}$ term. The behaviour of the $p$-variate score statistic may be explored by simulation.

5.4.2 Case with covariates

In many applications of survival analysis, lifetimes depend on explanatory variables or covariates. For example, in a reliability context, failure times of components in a system depend on the load applied to the components; in a medical context, the survival times for patients depend on the treatment received; the recurrence time of a disease depends on the dose of the medicine given. The score test may be extended to deal with data involving covariates. Suppose that we have a bivariate lifetime sample $(y_{11}, y_{12}), \ldots, (y_{n1}, y_{n2})$. Let the vector $x_{ij}$ denote the covariates associated with the $i$th observation. In the case without nuisance parameters, the
score statistic $T_{(n)}$ involving covariate information is defined as

$$T_{(n)} = \sum_{i=1}^{n} (2 - s_i \log s_i + 2 \log s_i - \frac{1}{s_i}),$$

where $s_i = \xi_{ij} y_{ij1}^{\phi_1} + \xi_{ij2} y_{ij2}^{\phi_2}$, $\log \xi_{ij} = x_{ij}^T \beta_j$, and $\beta_j = (\beta_{1j}, \ldots, \beta_{qj})$.

It is apparent that score statistic $T_{(n)}$ still depends on the observations only through the variable $s_i$. Although the parameters $\xi_{ij}$ vary over $i$ and $j$ because of the covariate information, the distribution of $s_i$ remains the same, which is a gamma with shape parameter 2 and scale parameter 1. Therefore, the score statistic $T_{(n)}$ in the case with covariates has the same properties as that in the without covariates case.

In the case in which there are nuisance parameters, the corresponding score statistic $T^*_{(n)}$ is

$$T^*_{(n)} = \sum_{i=1}^{n} (2 - \hat{s}_i \log \hat{s}_i + 2 \log \hat{s}_i - \frac{1}{\hat{s}_i}),$$

where $\hat{s}_i = \exp(x_{ij1}^T \hat{\beta}_1) y_{ij1}^{\hat{\phi}_1} + \exp(x_{ij2}^T \hat{\beta}_2) y_{ij2}^{\hat{\phi}_2}$. $\hat{\beta}_j$ and $\hat{\phi}_j$ are the maximum likelihood estimates of $\beta_j$ and $\phi_j$, respectively, under the null model.

The behaviour of the score statistic $T^*_{(n)}$ in the case with covariates was examined by a simulation study. Samples of pairs from independent Weibull distributions were generated with shape parameter $\phi_1 = \phi_2 = 1$ and scale parameter $\exp(\beta_{0j} + \beta_{1j} x_i)$. The coefficients $\beta_j$ were set as $\beta_1 = (1, 1)$, $\beta_2 = (1, -1)$. The $x_i$ were time constant dummy variables with values 1 for 50% of bivariate observations and values −1 for the remaining 50%. Sample size $n$ was chosen from 100 to 10,000. For each sample size, simulations were repeated 1000 times.

Table 5.12 shows the rate of convergence of normalised score test $T^*_{(n)}/C_n$ in the case with covariates. The results indicate that the rate of convergence to the normal is similar to that in the without covariates case. Critical values for sample size $n \leq 100$ obtained from simulations are also similar to those in the without covariates case. Figure 5.5 shows that the quadratic regression models can also be used to fit the estimated critical values of $T^*_{(n)}$ in the case with covariates. The
Figure 5.5: Estimated critical values of un-normalised score statistic $T_n^*$ for various sample sizes with superimposed quadratic fit in the case with covariate estimated coefficients of the regression model $C_T = c_0 + c_1 n + c_2 n^2$ are listed in Table 5.11 for comparison with the estimated coefficients in Table 5.5.

Therefore, we tentatively suggest that we may apply the critical values obtained from the without covariates case to the score test in the covariates case.

5.5 Examples

5.5.1 Infant nutrition data

The infant nutrition data is fitted by a bivariate PSW model as shown in the examples in Chapter 4. To detect whether two independent Weibull models are adequate to fit the data, we performed the score test with the null hypothesis of independent Weibulls against the alternative hypothesis of the PSW model. The
Table 5.11: Estimated coefficients in the model $C_T = c_0 + c_1 n + c_2 n^2$

<table>
<thead>
<tr>
<th>$P$</th>
<th>$c_0$</th>
<th>$c_1$</th>
<th>$c_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>-17.771</td>
<td>-0.8426</td>
<td>2.32x10^{-3}</td>
</tr>
<tr>
<td>0.025</td>
<td>-12.536</td>
<td>-0.5134</td>
<td>1.26x10^{-3}</td>
</tr>
<tr>
<td>0.05</td>
<td>-8.168</td>
<td>-0.4145</td>
<td>1.32x10^{-3}</td>
</tr>
<tr>
<td>0.10</td>
<td>-5.995</td>
<td>-0.2615</td>
<td>7.20x10^{-3}</td>
</tr>
</tbody>
</table>

Table 5.12: Normalised critical values of the score test $T^*_{(n)}/C_n$
(with a covariate, parameters unknown)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>$C_n$</th>
<th>Quantiles(s.d.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td>100</td>
<td>15.17</td>
<td>-1.60(0.02)</td>
</tr>
<tr>
<td>500</td>
<td>39.42</td>
<td>-1.54(0.04)</td>
</tr>
<tr>
<td>1000</td>
<td>58.77</td>
<td>-1.51(0.03)</td>
</tr>
<tr>
<td>5000</td>
<td>145.92</td>
<td>-1.44(0.02)</td>
</tr>
<tr>
<td>10000</td>
<td>214.60</td>
<td>-1.32(0.04)</td>
</tr>
<tr>
<td>$\infty$</td>
<td></td>
<td>-1.28</td>
</tr>
</tbody>
</table>
Weibull maximum likelihood estimates for the age of introducing fish and for the age of introducing egg are given in Chapter 1. Therefore, the observed value of 

$$T^*_n = \sum_{i=1}^{n} \left( -\hat{s}_i \log \hat{s}_i + 2 \log \hat{s}_i - \frac{1}{\hat{s}_i} + 2 \right) = -31.92,$$

where $n = 55$, $\hat{s}_i = \hat{\xi}_{11}y_{i1} + \hat{\xi}_{21}y_{i2}$. Using the estimated coefficients listed in Table 5.5, we see that the critical value of test statistic $T^*_n$ for a significance level of $P = 0.05$ and sample size $n = 55$ is $C_T = -7.255 - 0.394 \times 55 + 0.00103 \times 55^2 = -25.81$. $T^*_n$ is significant at the 0.05 level. Therefore, we conclude that there is evidence against the null hypothesis on the basis of the score test.

### 5.5.2 Repeated endurance exercise tests data

Consider the exercise times 1 hour and 3 hours after the drug treatment in the repeated endurance exercise tests data set in Table 1.2. In order to detect any heterogeneity in the pairs of observations, we performed score test. According to the maximum likelihood estimates given in the corresponding example of Chapter 1, the estimated score statistic without considering the covariate, dose, is $T^*_{(n),c} = -28.82$, where $T^*_{(n),c}$ was obtained from (5.12), the censored case. Since there are seven cases with censored values in the data, which gives approximately a 30% observed censoring proportion, column 4 in Table 5.5 was used to calculate the critical value of the score test in the censoring case, at $P = 0.05$. It is $C_T = -5.823 - 0.429 \times 21 + 0.00133 \times 21^2 = -14.25$. Similarly, the critical value of $T^*_{(n),c}$ at $P = 0.01$ may be obtained from $C_T = -8.604 - 0.936 \times 21 + 0.00308 \times 21^2 = -26.90$. Therefore, the score test gives very strong evidence against the null hypothesis.

In this example, it might be recognised that the exercise times are associated with the dose provided. Those patients who showed more severe initial incapacitation were given higher dose of drug. Hence, there is frailty after allowing for the covariate, dose. Taking the covariate, dose, into account, we obtained the maximum likelihood estimates $\hat{\phi}_1 = 1.61$, $\hat{\beta}_{10} = -10.14$ and $\hat{\beta}_{11} = -0.12$, for the exercise time 1 hour after treatment; $\hat{\phi}_2 = 1.98$, $\hat{\beta}_{20} = -11.52$ and $\hat{\beta}_{21} = -1.22$, for the exercise time 3 hours after treatment, where $\log \xi_{ij} = \beta_{j0} + \beta_{j1} \text{dose}(i), i = 1, \ldots, 21, j = 1, 2$. 119
The corresponding score statistic is estimated, which is $T_{(n),c}^* = -28.93$. This estimated score is similar to that obtained without considering the covariate. Hence the significant result found earlier cannot be explained by the covariate effect.

### 5.5.3 Fibre failure strength data

The null maximum likelihood estimates of the underlying Weibull parameters for each variable are given as an example in Chapter 1. Since there are missing values in the data, we only consider the pairs with both components observed. On the basis of the maximum likelihood estimates, the observed score statistics are $T_{n}^* = -41.36$, $(n = 42)$ for the pair $(Y_1, Y_2)$, $T_{n}^* = -42.09$ for the pair $(Y_1, Y_3)$, $(n = 39)$, $T_{n}^* = -27.78$ for the pair $(Y_1, Y_4)$, $(n = 42)$, $T_{n}^* = -39.95$, $(n = 39)$ for the pair $(Y_2, Y_3)$, $T_{n}^* = -20.10$, $(n = 42)$ for the pair $(Y_2, Y_4)$, $T_{n}^* = -24.94$, $(n = 39)$ for the pair $(Y_3, Y_4)$. The critical value of test statistic $T_{n}^*$ for a significance level of $P = 0.05$ are $C_T = -7.255 - 0.394 \times 39 + 0.00103 \times 39^2 = -21.05$, and $C_T = -7.255 - 0.394 \times 42 + 0.00103 \times 42^2 = -21.99$, for sample size $n = 39, n = 42$ respectively. Hence, there exists significant heterogeneity in all pairs of variables except for the pair $(Y_2, Y_4)$.

We also tried calculating the score value for the four-variate PSW model. On the basis on the maximum likelihood estimates and the 39 non-missing observations, the observed score statistic $T_{n}^* = -375.93$. The corresponding critical value at $P = 0.05$ obtained by simulation is $C_T = -253.72$. This result suggests that there are frailty effects in this data set.
Chapter 6

Likelihood ratio test for heterogeneity based on the PSW model

The likelihood ratio (LR) test is considered in this chapter as an alternative approach to the score test for detecting the heterogeneity in the PSW model. First we discuss the maximum likelihood (ML) estimator of \( \nu \) under the null hypothesis \( H_0 : \nu = 1 \), where a boundary problem occurs. The LR test and its asymptotic distribution are explored thereafter. A simulation study is used to examine the rate of convergence of the ML estimator of \( \nu \) and the LR statistic to their asymptotic distributions, and to compare the power of the LR test with that of the score test discussed in Chapter 5. Examples are given to illustrate the method and for comparison with the results from the score test.

We shall only consider the bivariate case, \( p = 2 \), in this chapter in detail. That is the model in the form

\[
S(y_1, y_2) = \exp(-s^\nu),
\]

where \( s = \xi_1 y_1^{\phi_1} + \xi_2 y_2^{\phi_2} \).
6.1 ML estimation of $\nu$ when $H_0 : \nu = 1$ is true

When the parameter $\nu$ has true value within the parameter space $0 < \nu < 1$, the ML estimator, $\hat{\nu}$, satisfies standard regularity conditions, see Cox & Hinkley (1982, Section 9.2). That is to say, $\hat{\nu}$ is consistent, asymptotically normal and asymptotically efficient. We are now interested in the special case when the true value is $\nu = 1$, which is on the boundary of the parameter space. Such situations are studied by Moran (1971). However, some of the regularity conditions that he considers fail in our case because, as stated in Chapter 4, the expected information at $H_0 : \nu = 1$ is infinite. A similar situation is identified and investigated by Tawn (1988) in the context of bivariate extremes, where the problem of infinite expected information also arises.

Suppose that $l_n(\nu)$ is the log-likelihood function for a bivariate sample of size $n$ without censoring, which is given by

$$ l_n(\nu) = -\sum_{i=1}^{n} s_i^\nu + \sum_{i=1}^{n} \sum_{j=1}^{2} \log(\xi_j \phi_j y_{ij}^{\phi_j - 1}) + \sum_{i=1}^{n} \log[\nu^2 s_i^{2\nu - 2} - \nu (\nu - 1) s_i^{\nu - 2}], \quad (6.1) $$

where $s_i = \xi_1 y_{i1}^{\phi_1} + \xi_2 y_{i2}^{\phi_2}$. Let $\hat{\nu}_0$ denote the ML estimate of $\nu$ when $H_0$ is true and when $\xi_j$ and $\phi_j$ ($j = 1, 2$) are known. Some of the properties which are relevant to $\hat{\nu}_0$ are presented in the following.

**Property 1.** The score statistic derived in Chapter 5 is $T_{(n)} = \sum_{i=1}^{n} T_i$, where $T_i$ is the first derivative of the log-likelihood contribution from the observation $i$ evaluated at $H_0$. For $\forall \nu \in (0, 1]$, we have, as $n \to \infty$, $\frac{\partial l_n(\nu)}{\partial \nu} \geq 0$, if $T_{(n)} \geq 0$.

**Proof:** For $\nu \in (0, 1]$, we first expand the log-likelihood $l_n(\nu)$ in a Taylor expansion at $\nu = 1$:

$$ l_n(\nu) = l_n(1) + (\nu - 1) \frac{\partial l_n(1)}{\partial \nu} + \frac{1}{2} (\nu - 1)^2 \frac{\partial^2 l_n(\nu^*)}{\partial \nu^2}, \quad (6.2) $$

where $\nu^* \in (\nu, 1)$. Note that $\frac{\partial l_n(1)}{\partial \nu} = T_{(n)}$. Thus, the first derivative of the equation (6.2) is

$$ \frac{\partial l_n(\nu)}{\partial \nu} = T_{(n)} + (\nu - 1) \frac{\partial^2 l_n(\nu^*)}{\partial \nu^2}. \quad (6.3) $$
For \( n \to \infty \), there is \( E(\frac{\partial^2 l_n}{\partial \nu^2}) = -E(\frac{\partial l_n}{\partial \nu})^2 \), so that
\[
\frac{\partial^2 l_n(\nu^*)}{\partial \nu^2} \sim -\left(\frac{\partial l_n(\nu^*)}{\partial \nu}\right)^2 < 0.
\]
Therefore, when \( T(n) \geq 0 \), we have
\[
\frac{\partial l_n(\nu)}{\partial \nu} \geq 0, \quad \text{for } \forall \nu \in (0, 1].
\]

**Property 2.** For the ML estimator \( \hat{\nu}_0 \), as \( n \to \infty \), we have
\[
P\{\hat{\nu}_0 = 1\} \to \frac{1}{2} \quad \text{and} \quad P\{0 < \hat{\nu}_0 < 1\} \to \frac{1}{2}.
\]
That is to say, for large samples, we expect to find about 50% of ML estimators are at the boundary \( \nu = 1 \), while another 50% of ML estimators are within the parameter space \( 0 < \nu < 1 \).

**Proof:** Following the results in Chapter 5, the first derivative of the log-likelihood under the null hypothesis is
\[
T(n) = \frac{\partial l_n(1)}{\partial \nu} = \sum_{i=1}^{n} \left(2 - s_i \log s_i + 2 \log s_i - \frac{1}{s_i}\right).
\]
Furthermore, when \( \nu = 1 \), we have proved that, as \( n \to \infty \), \( \frac{T(n)}{C_n} \) has asymptotically a standard normal distribution, where \( C_n = \sqrt{\frac{1}{2}n \log n} \). That is
\[
\frac{T(n)}{C_n} \to Z, \quad \text{as} \quad n \to \infty,
\]
where \( Z \) is a standard normal random variable. Therefore, we have that, when \( \nu = 1 \),
\[
Pr\{T(n) \geq 0\} = Pr\left\{\frac{T(n)}{C_n} \geq 0\right\} \to Pr\{Z \geq 0\} = \frac{1}{2}.
\]
In addition to Property 1, we have
\[
Pr\left\{\frac{\partial l_n(\nu)}{\partial \nu} \geq 0\right\} = \frac{1}{2}, \quad \text{for } \forall \nu \in (0, 1].
\]
The fact \( \frac{\partial l_n(\nu)}{\partial \nu} \geq 0 \) indicates that the log-likelihood function \( l_n(\nu) \) is increasing at \( \nu \) for all \( \nu \in (0, 1] \), which means that the ML estimators \( \hat{\nu}_0 \) has to be at the boundary \( \nu = 1 \). Hence, we conclude that, as \( n \to \infty \),
\[
P\{\hat{\nu}_0 = 1\} \to \frac{1}{2} \quad \text{and} \quad P\{0 < \hat{\nu}_0 < 1\} \to \frac{1}{2}.
\]
Property 3. Suppose that $H_0$ is true. Let $T_{(n)}$ and $C_n$ be as above. Then, as $n \to \infty$,

$$
\frac{\sum_{i=1}^{n} T_i^2}{C_n^2} \to 1
$$

(6.4)

in distribution.

Proof: According to the discussion in Chapter 5, we may conclude that, under $H_0$, the leading term of $\sum_{i=1}^{n} T_i^2$ is $\sum_{i=1}^{n} \frac{1}{s_i^2}$, that is

$$
\sum_{i=1}^{n} T_i^2 = \sum_{i=1}^{n} \frac{1}{s_i^2} + o(\sum_{i=1}^{n} \frac{1}{s_i^2}).
$$

Let $u_i = \frac{1}{s_i}$. Hence the density function of $u_i$ is $f(u_i) = \frac{1}{u_i} e^{-\frac{1}{u_i}}$. Therefore, we have for $w \to \infty$,

$$
E\left(\frac{1}{s_i^2}\right) = E(u_i^2) \sim \int_0^w \frac{1}{u_i} e^{-\frac{1}{u_i}} du_i = \int_{\frac{1}{w}}^{\infty} \frac{1}{s_i} e^{-s_i} ds_i
$$

$$
= \log w - \gamma - \sum_{k=1}^{\infty} \frac{(-1)^k}{kk!} \left(\frac{1}{\omega}\right)^k,
$$

see Abramowitz et al (1972). Furthermore,

$$
\sum_{i=1}^{n} \frac{1}{s_i^2} \sim n\left\{\log w - \gamma - \sum_{k=1}^{\infty} \frac{(-1)^k}{kk!} \left(\frac{1}{\omega}\right)^k\right\}, \quad \text{as } n \to \infty.
$$

Let $\omega = C_n$. Then, we have that

$$
\frac{\sum_{i=1}^{n} T_i^2}{C_n^2} \sim \frac{\sum_{i=1}^{n} \frac{1}{s_i^2} + o(\sum_{i=1}^{n} \frac{1}{s_i^2})}{C_n^2} \sim n\left\{\frac{1}{2} \log\left(\frac{1}{2} n \log n\right) - \gamma\right\} = \frac{1}{2} n \log n
$$

as $n \to \infty$.

Property 4: For the ML estimator $\hat{\nu}_0$, as $n \to \infty$, we have

$$
(1 - \hat{\nu}_0) C_n \to Z I(Z > 0) \quad \text{for } 0 < \hat{\nu}_0 < 1,
$$

(6.5)

where $C_n = \sqrt{\frac{1}{2} n \log n}$, $Z$ is a standard normal random variable, and $I$ is an indicator function. This property is analogous to the results of Tawn (1988), who studied estimation of bivariate extreme value distributions. In his work, the parameter space is $\nu \geq 1$ so that the boundary is also at $\nu = 1$.
**Proof:** On the basis of equation (6.3), the ML estimator $\hat{v}_0$, as $0 < \hat{v}_0 < 1$, should satisfy

$$0 = T(n) + (\hat{v}_0 - 1)\frac{\partial^2 l_n(\nu^*)}{\partial \nu^2},$$

where $\nu^* \in (\hat{v}_0, 1)$. Hence, we have, for $0 < \hat{v}_0 < 1$,

$$(1 - \hat{v}_0)C_n = \frac{T(n)C_n}{\partial^2 l_n(\nu^*)} = -\frac{T(n)}{\partial^2 l_n(\nu^*)}.$$

Note that, here $T(n) < 0$, since $0 < \hat{v}_0 < 1$. It follows that

$$\frac{T(n)}{C_n} \to -IZI(Z > 0), \quad \text{as} \quad n \to \infty,$$

where $Z$ is a random variable from a standard normal distribution, $I$ is an indicator function. On the other hand, as $n \to \infty$, $\hat{v}_0 \to 1$ so that $\nu^* \to 1$. Thus,

$$\frac{\partial^2 l_n(\nu^*)}{\partial \nu^2} = \sum_{i=1}^{n} \frac{\partial^2 l_i(\nu^*)}{\partial \nu^2} \sim -\sum_{i=1}^{n} \left(\frac{\partial l_i(\nu^*)}{\partial \nu}\right)^2 \sim -\sum_{i=1}^{n} T_i^2.$$

Combining the result that $\sum_{i=1}^{n} \frac{T_i^2}{C_n^2} \to 1$, as $n \to \infty$, we have the following asymptotic results:

$$(1 - \hat{v}_0)C_n \to ZI(Z > 0) \quad \text{for} \quad 0 < \hat{v}_0 < 1.$$

Therefore, combining the above results, we have that, under $H_0$, the asymptotic behaviour of the ML estimator $\hat{v}_0$ under $H_0$ is

$$(1 - \hat{v}_0)\sqrt{\frac{1}{2} n \log n} \to W, \quad \text{as} \quad n \to \infty,$$

where convergence is in distribution and $W$ is a non-negative random variable. Moreover, $W$ has the distribution function

$$P(W \leq w) = \begin{cases} \Phi(w) & \text{if} \quad w \geq 0 \\ 0 & \text{if} \quad w < 0 \end{cases}$$

where $\Phi(\cdot)$ is the standard normal distribution function.
6.2 The likelihood ratio test and its asymptotic properties

Suppose that we observe a bivariate sample \((y_{11}, y_{12}), \ldots, (y_{n1}, y_{n2})\). The likelihood ratio (LR) test for the null hypothesis that the data are from independent Weibull distributions, i.e. \(H_0 : \nu = 1\), against the alternative hypothesis that they are from the PSW model, i.e. \(H_1 : 0 < \nu < 1\), is derived as follows. At present, we assume that \(\xi_j, \phi_j \ (j = 1, 2)\) are known. The corresponding LR statistic is

\[
\Lambda(n) = 2\{l_n(\hat{\nu}) - l_n(1)\}, \quad (6.9)
\]

where \(l_n(.)\) is the log-likelihood function as shown in equation (6.1), and \(\hat{\nu}\) is the unrestricted maximum likelihood estimate of \(\nu\) when \(\xi_j, \phi_j \ (j = 1, 2)\) are known. \(l_n(\hat{\nu})\) is the unrestricted maximised value of \(l_n\), while \(l_n(1)\) is \(l_n\) evaluated at \(\nu = 1\). Since the null hypothesis for the LR test based on the PSW model is on the boundary of the parameter space, we expect a nonstandard asymptotic null distribution for the LR statistic.

The asymptotic properties for the LR statistic \(\Lambda(n)\) under \(H_0 : \nu = 1\) are explored in the following. Let \(\hat{\nu}_0\) be the maximum likelihood estimate of \(\nu\) under the null hypothesis, and suppose that \(0 < \hat{\nu}_0 < 1\). By expanding the log-likelihood function \(l_n(\hat{\nu})\) at \(\nu = 1\), we have

\[
2\{l_n(\hat{\nu}_0) - l_n(1)\} = 2(\hat{\nu}_0 - 1) \frac{\partial l_n(1)}{\partial \nu} + (\hat{\nu}_0 - 1)^2 \frac{\partial^2 l_n(\nu^*)}{\partial \nu^2},
\]

where \(\nu^* \in (\hat{\nu}_0, 1)\). Following equation (6.3) and the fact that \(\frac{\partial l_n(\hat{\nu}_0)}{\partial \nu} = 0\), we have

\[
\frac{\partial l_n(1)}{\partial \nu} = - (\hat{\nu}_0 - 1) \frac{\partial^2 l_n(\nu^*)}{\partial \nu^2},
\]

so that the LR statistic under \(H_0\) may be expressed as

\[
\Lambda(n) = -2(\hat{\nu}_0 - 1)^2 \frac{\partial^2 l_n(\nu^*)}{\partial \nu^2} + (\hat{\nu}_0 - 1)^2 \frac{\partial^2 l_n(\nu^*)}{\partial \nu^2} = - (\hat{\nu}_0 - 1)^2 \frac{\partial^2 l_n(\nu^*)}{\partial \nu^2}.
\]
Therefore, using the results in (6.4), (6.5) and (6.6), the asymptotic distribution of $A(n)$ for $0 < \hat{\nu}_0 < 1$ is obtained, which is, as $n \to \infty$,

$$A(n) = -(\hat{\nu}_0 - 1)^2 \frac{\partial^2 l_n(\nu^*)}{\partial \nu^2} \sum_{i=1}^{n} \frac{T_i^2}{C_n^2} \to Z^2,$$

where $Z$ is a standard normal random variable.

In terms of the results that, for large samples, 50 percent of ML estimates $\hat{\nu}_0$ under $H_0$ are at the boundary, we conclude that the LR statistic $A(n)$ has null asymptotic distribution such that

$$A(n) = -2\{l_n(\hat{\nu}_0) - l_n(1)\} \to W^2, \quad \text{as} \quad n \to \infty,$$

where $W$ is as defined in (6.8). That is to say, if $A(n)$ asymptotically has a $\chi^2_{(1)}$ distribution with probability one half and is zero with probability one half when $H_0$ is true. This result agrees with the general results obtained by Self & Liang (1987). Their paper summarises all the earlier work on the boundary problem for ML estimation and generalised LR tests, and provides a uniform framework for the large sample distribution of the ML estimator and LR statistic. The situation we encounter here is a special case in which only one parameter is on the boundary. A similar conclusion was drawn by Tawn (1988) for a related problem.

So far, we have ignored the fact that the parameters $\xi_j, \phi_j$ ($j = 1, 2$) are usually unknown in practice. The corresponding LR statistic $A^*_n$ for the case with nuisance parameters and its asymptotic properties can be obtained as follows. Let $\theta$ denote the nuisance parameters, i.e. $\theta = (\xi_1, \phi_1, \xi_2, \phi_2)$, the LR statistic $A^*_n$ is

$$A^*_n = 2\{l_n(\hat{\nu}, \hat{\theta}) - l_n(1, \hat{\theta})\}, \quad (6.10)$$

where $l_n(.)$ is given in equation (6.1); $\hat{\nu}$ and $\hat{\theta}$ are the unrestricted ML estimators of $\nu$ and $\theta$ respectively; $\hat{\theta}$ is the ML estimator of $\theta$ under $H_0$.

Since the ML estimators of the Weibull parameters $\xi_j, \phi_j$ ($j = 1, 2$), are regular, it is not difficult to obtain the asymptotic distribution of $A^*_n$ using similar methods.
to those applied to the case with no nuisance parameters. Hence, as for $A_{(n)}$, under $H_0$, the LR statistic $A^*_n$ is also asymptotically a $50:50$ mixture of a $\chi^2_{(1)}$ distribution and a probability mass at zero.

Now we consider the case in which observations may be right censored. Assume that $Y_1$ and $Y_2$ are censored at fixed times $c_1$ and $c_2$ respectively. For a bivariate sample with $n$ observations, $(y_{11}, y_{12}), \ldots, (y_{n1}, y_{n2})$, there are four possibilities for each pair of observations, which are

- $R_1$: $y_1$ is observed, $y_2$ is observed;
- $R_2$: $y_1$ is observed, $y_2$ is censored;
- $R_3$: $y_1$ is censored, $y_2$ is observed;
- $R_4$: $y_1$ is censored, $y_2$ is censored.

The likelihood function is

$$L_n = \prod_{i \in R_1} f(y_{i1}, y_{i2}) \frac{-\partial S(y_{i1}, c_2)}{\partial y_{i1}} \prod_{i \in R_3} \frac{-\partial S(c_1, y_{i2})}{\partial y_{i2}} \prod_{i \in R_4} S(c_1, c_2),$$

where $f(y_1, y_2)$ and $S(y_1, y_2)$ are the density and survivor function for the PSW model, respectively. Thus the corresponding log-likelihood function can be expressed as

$$l_{n,c}(\nu) = -\sum_{i=1}^{n} s_i^\nu + \sum_{i=1}^{n} \sum_{j=1}^{2} \delta_{ij} \log(\xi_j \phi_j y_{ij}^{\phi_j-1}) - \sum_{i=1}^{n} (\delta_{i1} + \delta_{i2}) \log s_i$$

$$+ \sum_{i=1}^{n} \delta_{i1} \delta_{i2} \log(n s_i^\nu - \nu + 1) + \sum_{i=1}^{n} (\delta_{i1} + \delta_{i2} - \delta_{i1} \delta_{i2}) (\nu \log s_i + \log \nu),$$

where $\delta_{ij}$ is an indicator variable defined as $\delta_{ij} = 1$ if $y_{ij}$ is observed; $\delta_{ij} = 0$ if $y_{ij}$ is censored.

Based on the log-likelihood function $l_{n,c}(\nu)$, the corresponding LR statistic for the case with censoring may be obtained. When the parameters $\xi_j, \phi_j$ are known, the LR statistics for the censored case is

$$A_{(n),c} = 2\{l_{n,c}(\hat{\nu}) - l_{n,c}(1)\},$$

where $l_{n,c}(\cdot)$ is shown in equation (6.11), and $\hat{\nu}$ is the unrestricted maximum likelihood estimate of $\nu$ when $\xi_j, \phi_j$ ($j = 1, 2$) are known. When the parameters $\xi_j$,
\( \phi_j \) are unknown, the LR statistic for the censored case is

\[
\Lambda_{(n),c}^* = 2\{l_{n,c}(\hat{\nu}, \hat{\theta}) - l_{n,c}(1, \hat{\theta})\},
\]

(6.13)

where \( l_{n,c}(\cdot) \) is given in equation (6.11), and \( \hat{\nu} \) and \( \hat{\theta} \) are the unrestricted ML estimates of \( \nu \) and \( \theta \), respectively; \( \hat{\theta} \) is the ML estimator of \( \theta \) under \( H_0 \).

6.3 Simulation study

The simulation work here covers three aspects. First, in order to see how applicable the asymptotic results are for the LR statistics, the rate of convergence to the mixture distribution is examined. Secondly, for the purpose of application, critical values for the LR tests in small sample sizes are estimated. Finally, the powers of the LR tests in different situations are evaluated, and compared with those of the corresponding score tests.

6.3.1 Evaluation of the rate of convergence

To examine how applicable the null asymptotic results are, we performed a series of simulations to evaluate the rate of convergence of the LR statistics. Sample size \( n \) was chosen to vary from 100 to 10,000. For each sample size, the simulation was repeated 10,000 times for the case in which there are no nuisance parameters. Since the simulation in the case in which there are nuisance parameters is much more time-consuming, we reduced the number of simulations to 1,000 in this case. To estimate the stability of the simulated results, the simulation was replicated four times for each sample size and standard deviations of estimated quantiles were evaluated. To verify that the LR statistic has a mixture distribution with 50% values at zero under \( H_0 \), we counted the percentage of maximum likelihood estimates of \( \nu \) which are at the boundary \( \nu = 1 \). Without loss of generality, samples were generated from two independent Weibull distributions with parameters \( \xi_j = \phi_j = 1 \ (j = 1, 2) \).
For the case in which there are no nuisance parameters and no censoring, the unrestricted ML estimate of $\nu$ and its corresponding maximised log-likelihood $l_n(\hat{\nu})$ were calculated and the log-likelihood of $l_n(1)$ was evaluated. Therefore, the LR statistic $\Lambda_{(n)}$ can be obtained from equation (6.9). Conditionally on $\Lambda_{(n)} > 0$, we focused on the non-zero part of the LR statistic. The mean upper tail quantile of $\Lambda_{(n)}$ (excluding zero values) and the standard deviations of the estimated quantiles (in brackets) are listed in Table 6.1, where quantiles were selected at $P = 0.99$, $P = 0.975$, $P = 0.95$, $P = 0.90$. The bottom line of Table 6.1 lists the corresponding quantiles of the $\chi^2(1)$ distribution with one degree of freedom. The second column lists the mean percentages of ML estimates which are at the boundary $\hat{\nu} = 1$. The results show that the percentage approaches 50% as the sample size increases, which agrees with the properties discussed in section 6.2. Table 6.1 also shows that the convergence to the chi-squared distribution is reasonable. The rate of convergence to the chi-squared for the LR statistic $\Lambda_{(n)}$ is much quicker than the rate of convergence to the normal for the score statistic $T(n)$.

Furthermore, to compare the behaviour of the convergence of the LR statistics $\Lambda_{(n)}^*$, $\Lambda_{(n),c}$, and $\Lambda_{(n),c}^*$, more simulation work was done. In the case in which there is no censoring and the marginal parameters are unknown, the unrestricted ML estimates of $\nu$, $\xi_j$, $\phi_j$ ($j = 1, 2$) and the corresponding maximised log-likelihood $l_n$ were obtained, while the ML estimates of $\xi_j$, $\phi_j$ ($j = 1, 2$) under the null model and their corresponding log-likelihood were estimated. Then, the LR statistic $\Lambda_{(n)}^*$ can be obtained from (6.10). Similarly, by fixing the censoring time $c_1 = c_2 = c = 1.8$ for both components, the same procedures described above were applied to obtain the LR statistics $\Lambda_{(n),c}$ and $\Lambda_{(n),c}^*$, where the generated data has, on average, 30% censoring for at least one component of all the pairs.

Conditional on $\Lambda_{(n)}^* > 0$, $\Lambda_{(n),c} > 0$ and $\Lambda_{(n),c}^* > 0$, the quantiles of the LR statistics $\Lambda_{(n)}^*$, $\Lambda_{(n),c}$ and $\Lambda_{(n),c}^*$ for various sample sizes are presented in Tables 6.2 - 6.4, respectively. The convergence to the chi-squared distribution for the $\Lambda_{(n)}^*$ is also good. The standard deviations of the estimated percentiles listed in the brackets are higher in this case because of the reduced number of replications. The rate of
Table 6.1: Estimated critical values of the likelihood ratio test statistic $\Lambda_{(n)}$
(no censoring, marginal parameters known)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Percent($%$) ($\Lambda_{(n)} = 0$)</th>
<th>Quantiles(s.d.) conditional on $\Lambda_{(n)} &gt; 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.90 (0.09)  0.95 (0.12)  0.975 (0.03)  0.99 (0.25)</td>
</tr>
<tr>
<td>100</td>
<td>55.18</td>
<td>2.69 (0.09)  3.75 (0.12)  4.90 (0.03)  6.43 (0.25)</td>
</tr>
<tr>
<td>500</td>
<td>53.86</td>
<td>2.65 (0.07)  3.74 (0.08)  4.81 (0.13)  6.49 (0.22)</td>
</tr>
<tr>
<td>1000</td>
<td>53.68</td>
<td>2.68 (0.01)  3.78 (0.09)  4.97 (0.11)  6.74 (0.15)</td>
</tr>
<tr>
<td>5000</td>
<td>53.09</td>
<td>2.71 (0.04)  3.81 (0.07)  4.94 (0.11)  6.59 (0.24)</td>
</tr>
<tr>
<td>10000</td>
<td>52.85</td>
<td>2.66 (0.06)  3.80 (0.11)  5.09 (0.20)  6.61 (0.19)</td>
</tr>
<tr>
<td>$\infty$</td>
<td></td>
<td>2.71  3.84  5.02  6.64</td>
</tr>
</tbody>
</table>

convergence of $\Lambda_{(n),c}$ and $\Lambda^*_{(n),c}$ in the case with 30% censoring is slower than in the no censoring case, but still acceptable. For the same reason as stated above, the standard deviations of the estimated percentiles for $\Lambda^*_{(n),c}$ are greater than those for the no nuisance parameters case.

6.3.2 Estimation of critical values

Obviously, the asymptotic approximation may be poor for smaller sample sizes that occur in practice. We therefore estimate the critical values of the LR test for smaller sample sizes ($n \leq 100$) by simulation and examine how applicable the asymptotic results are for smaller samples. The simulation was based on 1,000 replications for each case.

Note that the critical values obtained here are unconditional. In other words, the simulated LR values when $\hat{v} = 1$ are also included in calculating the observed quantiles.

The simulation study shows that, in the case in which there are no nuisance parameters, there is an approximately linear relationship between the critical values and the sample sizes at each significance level ($P = 0.01$, $P = 0.025$, $P = 0.05$ and
Table 6.2: Estimated critical values of the likelihood ratio test statistic \( \Lambda_n^* \) (no censoring, marginal parameters unknown)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Percent(%) ((\Lambda_n = 0))</th>
<th>Quantiles(s.d.) conditional on ( \Lambda_n &gt; 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.90</td>
<td>0.95</td>
</tr>
<tr>
<td>100</td>
<td>55.8</td>
<td>2.75(0.36)</td>
</tr>
<tr>
<td>500</td>
<td>56.3</td>
<td>2.65(0.20)</td>
</tr>
<tr>
<td>1000</td>
<td>55.1</td>
<td>2.93(0.14)</td>
</tr>
<tr>
<td>5000</td>
<td>53.9</td>
<td>2.77(0.25)</td>
</tr>
<tr>
<td>10000</td>
<td>53.4</td>
<td>2.80(0.07)</td>
</tr>
<tr>
<td>( \infty )</td>
<td></td>
<td>2.71</td>
</tr>
</tbody>
</table>

Table 6.3: Estimated critical values of the likelihood ratio test statistic \( \Lambda_{n,c} \) (with 30% censoring, marginal parameters known)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Percent(%) ((\Lambda_n = 0))</th>
<th>Quantiles(s.d.) conditional on ( \Lambda_n &gt; 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.90</td>
<td>0.95</td>
</tr>
<tr>
<td>100</td>
<td>56.72</td>
<td>2.64(0.12)</td>
</tr>
<tr>
<td>500</td>
<td>55.08</td>
<td>2.71(0.06)</td>
</tr>
<tr>
<td>1000</td>
<td>54.17</td>
<td>2.68(0.03)</td>
</tr>
<tr>
<td>5000</td>
<td>53.63</td>
<td>2.73(0.07)</td>
</tr>
<tr>
<td>10000</td>
<td>53.62</td>
<td>2.62(0.08)</td>
</tr>
<tr>
<td>( \infty )</td>
<td></td>
<td>2.71</td>
</tr>
</tbody>
</table>

Table 6.4: Estimated critical values of likelihood ratio test statistic \( \Lambda_{n,c}^* \) (with 30% censoring, marginal parameters unknown)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Percent(%) ((\Lambda_n = 0))</th>
<th>Quantiles(s.d.) conditional on ( \Lambda_n &gt; 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.90</td>
<td>0.95</td>
</tr>
<tr>
<td>100</td>
<td>55.4</td>
<td>2.95(0.19)</td>
</tr>
<tr>
<td>500</td>
<td>56.1</td>
<td>2.85(0.14)</td>
</tr>
<tr>
<td>1000</td>
<td>55.6</td>
<td>2.66(0.08)</td>
</tr>
<tr>
<td>5000</td>
<td>56.0</td>
<td>2.76(0.10)</td>
</tr>
<tr>
<td>10000</td>
<td>55.4</td>
<td>2.50(0.21)</td>
</tr>
<tr>
<td>( \infty )</td>
<td></td>
<td>2.71</td>
</tr>
</tbody>
</table>

132
Table 6.5: Estimated coefficients in the model $C_A = c_0 + c_1 n + c_2 n^2$

<table>
<thead>
<tr>
<th>P</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$c_0$</td>
</tr>
<tr>
<td>0.01</td>
<td>6.4118</td>
</tr>
<tr>
<td>0.025</td>
<td>4.7782</td>
</tr>
<tr>
<td>0.05</td>
<td>3.5066</td>
</tr>
<tr>
<td>0.10</td>
<td>2.1165</td>
</tr>
</tbody>
</table>

$P = 0.10$). The fitted lines and the simulated data are plotted in Figure 6.1. The plots indicate that the fitted lines at each significance level are almost horizontal for sample sizes from 10 to 100. Therefore we suggest using the mean values obtained from simulation as critical values of the LR test $\Lambda(n)$. That is, when sample size $n$ satisfies $10 \leq n \leq 100$, the critical values at $P = 0.01$, $P = 0.025$, $P = 0.05$ and $P = 0.10$ are approximately 4.95, 3.49, 2.48 and 1.36, respectively. Alternatively, since these values are close to the corresponding $\chi^2$ values, $\chi^2$ values may be used as critical values even in samples with smaller size, i.e. $10 \leq n \leq 100$. The $\chi^2$ values at $P = 0.01$, $P = 0.025$, $P = 0.05$ and $P = 0.10$ are 5.41, 3.84, 2.71, and 1.64 respectively. Using the $\chi^2$ values as critical values will give a conservative test.

For the case in which there are nuisance parameters, we find that, approximately, the variations of the critical values within the range of $n = 10$ to $n = 100$ are quadratic in $n$, as shown in Figure 6.2. The estimated coefficients of the regression model $C_A = c_0 + c_1 n + c_2 n^2$ are listed in Table 6.5, where $C_A$ is the estimated critical value at the corresponding significance level $P$, $n$ is the sample size. It can be seen from Figure 6.2 that the fitted curves are very flat when the sample size is between 50 and 100. The means of the simulated values for $50 \leq n \leq 100$ at $P = 0.01$, $P = 0.025$, $P = 0.05$ and $P = 0.10$ are 5.44, 3.87, 2.64, 1.56, respectively. These values are very close to the corresponding $\chi^2$ values. Hence, we suggest that, as an alternative, $\chi^2$ values may be used as critical values for samples with size $50 \leq n \leq 100$ in the case in which there are nuisance parameters.
Figure 6.1: Estimated critical values of the LR test $\Lambda_{(n)}$ for various samples

Figure 6.2: Estimated critical values of the LR test $\Lambda_{(n)}^{*}$ for various samples
6.3.3 Estimation of power functions

Power for detecting heterogeneity based on the PSW model

A simulation experiment was run to estimate the power of the LR tests for various sample sizes. Random variables from the PSW distributions were generated in the same way as described in Chapter 5, Section 5.3.3. Simulation was based on 10,000 replications. The parameter $\nu$ was chosen to vary from 0.5 to 0.99. The estimated power of the LR statistics $\Lambda(n), \Lambda^*(n), \Lambda(n)_c$ and $\Lambda^*(n)_c$ for various sample sizes are listed in Table 6.6 and 6.7, where the critical value at significance level $P = 0.05$ is 2.71, i.e. corresponding $\chi^2$ value, for larger sample sizes ($n > 100$). The critical values for smaller sample sizes ($n \leq 50$) are from estimated results in Table 6.5. The results indicate that, for sample size $n > 100$, the estimated power of the LR tests for detecting the heterogeneity in the PSW model is very good in the four different cases. For smaller sample sizes, the estimated powers of the LR tests are lower, but still reasonable. The estimated powers of $\Lambda(n)$ and $\Lambda(n)_c$ are higher than those of $\Lambda^*(n)$ and $\Lambda^*(n)_c$. On the other hand, the estimated power is not reduced much in the case with 30% censoring.

Power for detecting heterogeneity based on the GW model

We also investigated the power of the LR tests in the case in which the model is misspecified, say, when a sample of observations are actually from a GW distribution instead of a PSW distribution.

The generation of bivariate random variables is analogous to that performed in the same situation for the score tests, see Section 5.3.3. The gamma parameter $\nu$ was chosen as $1.0, 2.0, 4.0, 8.0, 16.0, 50.0$. Simulation was based on 10,000 replications for each case. Table 6.8 shows the estimated power of the LR tests $\Lambda(n)$ and $\Lambda^*(n)$ for sample size $n = 20, 50, 100, 500$.

The results indicate that the LR tests still have good power even if the hetero-
Table 6.6: Estimated power of likelihood ratio tests $\Lambda_{(n)}$ and $\Lambda_{*(n)}$ for different sample sizes and $P = 0.05$
(with no censoring)

<table>
<thead>
<tr>
<th>sample size</th>
<th>$\nu$</th>
<th>$\Lambda_{(n)}$ power(%)</th>
<th>$\Lambda_{*(n)}$ power(%)</th>
<th>sample size</th>
<th>$\nu$</th>
<th>$\Lambda_{(n)}$ power(%)</th>
<th>$\Lambda_{*(n)}$ power(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.50</td>
<td>98.66</td>
<td>87.22</td>
<td>100</td>
<td>0.50</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>92.57</td>
<td>69.69</td>
<td></td>
<td>0.60</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>78.20</td>
<td>50.98</td>
<td></td>
<td>0.70</td>
<td>100.00</td>
<td>99.80</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>50.37</td>
<td>36.75</td>
<td></td>
<td>0.80</td>
<td>99.90</td>
<td>95.89</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>20.48</td>
<td>22.27</td>
<td></td>
<td>0.90</td>
<td>81.53</td>
<td>65.24</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>11.67</td>
<td>18.49</td>
<td></td>
<td>0.95</td>
<td>35.56</td>
<td>40.01</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>5.40</td>
<td>6.23</td>
<td></td>
<td>0.99</td>
<td>7.28</td>
<td>6.65</td>
</tr>
<tr>
<td>20</td>
<td>0.50</td>
<td>100.00</td>
<td>98.60</td>
<td>200</td>
<td>0.50</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>99.80</td>
<td>88.08</td>
<td></td>
<td>0.60</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>94.54</td>
<td>71.56</td>
<td></td>
<td>0.70</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>66.42</td>
<td>47.14</td>
<td></td>
<td>0.80</td>
<td>100.00</td>
<td>99.90</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>32.57</td>
<td>26.49</td>
<td></td>
<td>0.90</td>
<td>97.25</td>
<td>89.40</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>15.41</td>
<td>12.84</td>
<td></td>
<td>0.95</td>
<td>60.96</td>
<td>57.26</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>6.56</td>
<td>5.10</td>
<td></td>
<td>0.99</td>
<td>11.94</td>
<td>6.98</td>
</tr>
<tr>
<td>30</td>
<td>0.50</td>
<td>100.00</td>
<td>99.85</td>
<td>300</td>
<td>0.50</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>100.00</td>
<td>96.96</td>
<td></td>
<td>0.60</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>98.80</td>
<td>85.80</td>
<td></td>
<td>0.70</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>83.63</td>
<td>59.76</td>
<td></td>
<td>0.80</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>41.05</td>
<td>33.50</td>
<td></td>
<td>0.90</td>
<td>99.29</td>
<td>97.20</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>18.63</td>
<td>17.02</td>
<td></td>
<td>0.95</td>
<td>73.47</td>
<td>69.56</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>7.04</td>
<td>6.80</td>
<td></td>
<td>0.99</td>
<td>13.46</td>
<td>7.49</td>
</tr>
<tr>
<td>40</td>
<td>0.50</td>
<td>100.00</td>
<td>100.00</td>
<td>400</td>
<td>0.50</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>100.00</td>
<td>98.81</td>
<td></td>
<td>0.60</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>99.96</td>
<td>91.21</td>
<td></td>
<td>0.70</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>93.41</td>
<td>69.45</td>
<td></td>
<td>0.80</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>49.40</td>
<td>36.82</td>
<td></td>
<td>0.90</td>
<td>99.80</td>
<td>93.88</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>22.05</td>
<td>21.24</td>
<td></td>
<td>0.95</td>
<td>78.92</td>
<td>71.90</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>6.72</td>
<td>5.93</td>
<td></td>
<td>0.99</td>
<td>11.83</td>
<td>10.52</td>
</tr>
<tr>
<td>50</td>
<td>0.50</td>
<td>100.00</td>
<td>100.00</td>
<td>500</td>
<td>0.50</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>100.00</td>
<td>100.00</td>
<td></td>
<td>0.60</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>99.90</td>
<td>96.54</td>
<td></td>
<td>0.70</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>95.91</td>
<td>80.92</td>
<td></td>
<td>0.80</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>57.96</td>
<td>45.82</td>
<td></td>
<td>0.90</td>
<td>100.00</td>
<td>97.17</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>24.69</td>
<td>12.80</td>
<td></td>
<td>0.95</td>
<td>84.30</td>
<td>78.91</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>7.34</td>
<td>5.37</td>
<td></td>
<td>0.99</td>
<td>12.12</td>
<td>9.20</td>
</tr>
</tbody>
</table>
Table 6.7: Estimated power of likelihood ratio tests $\Lambda_{(n),c}$ and $\Lambda^*_{(n),c}$ for different sample sizes and $P = 0.05$
(with 30% censoring)

<table>
<thead>
<tr>
<th>sample size</th>
<th>$\nu$</th>
<th>$\text{power(%)}$</th>
<th>$\Lambda_{(n),c}$</th>
<th>$\Lambda^*_{(n),c}$</th>
<th>sample size</th>
<th>$\nu$</th>
<th>$\text{power(%)}$</th>
<th>$\Lambda_{(n),c}$</th>
<th>$\Lambda^*_{(n),c}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.50</td>
<td>89.32</td>
<td>78.93</td>
<td></td>
<td>100</td>
<td>0.50</td>
<td>100.00</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>75.89</td>
<td>60.42</td>
<td></td>
<td></td>
<td>0.60</td>
<td>100.00</td>
<td>98.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>70.40</td>
<td>54.68</td>
<td></td>
<td></td>
<td>0.70</td>
<td>95.45</td>
<td>91.80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>40.63</td>
<td>23.44</td>
<td></td>
<td></td>
<td>0.80</td>
<td>89.87</td>
<td>83.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>19.80</td>
<td>16.73</td>
<td></td>
<td></td>
<td>0.90</td>
<td>76.80</td>
<td>59.44</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>12.06</td>
<td>7.52</td>
<td></td>
<td></td>
<td>0.95</td>
<td>31.40</td>
<td>28.91</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>5.46</td>
<td>5.10</td>
<td></td>
<td></td>
<td>0.99</td>
<td>6.78</td>
<td>9.80</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.50</td>
<td>92.67</td>
<td>85.58</td>
<td></td>
<td>200</td>
<td>0.50</td>
<td>100.00</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>84.12</td>
<td>76.22</td>
<td></td>
<td></td>
<td>0.60</td>
<td>100.00</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>69.89</td>
<td>51.64</td>
<td></td>
<td></td>
<td>0.70</td>
<td>100.00</td>
<td>93.89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>53.66</td>
<td>48.84</td>
<td></td>
<td></td>
<td>0.80</td>
<td>90.55</td>
<td>89.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>23.13</td>
<td>17.11</td>
<td></td>
<td></td>
<td>0.90</td>
<td>79.49</td>
<td>68.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>12.12</td>
<td>9.20</td>
<td></td>
<td></td>
<td>0.95</td>
<td>52.30</td>
<td>47.55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>6.50</td>
<td>5.48</td>
<td></td>
<td></td>
<td>0.99</td>
<td>8.78</td>
<td>6.87</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.50</td>
<td>98.90</td>
<td>90.69</td>
<td></td>
<td>300</td>
<td>0.50</td>
<td>100.00</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>90.76</td>
<td>85.17</td>
<td></td>
<td></td>
<td>0.60</td>
<td>100.00</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>79.80</td>
<td>69.20</td>
<td></td>
<td></td>
<td>0.70</td>
<td>100.00</td>
<td>98.90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>65.06</td>
<td>51.73</td>
<td></td>
<td></td>
<td>0.80</td>
<td>93.45</td>
<td>91.22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>37.22</td>
<td>30.15</td>
<td></td>
<td></td>
<td>0.90</td>
<td>88.33</td>
<td>83.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>14.98</td>
<td>12.41</td>
<td></td>
<td></td>
<td>0.95</td>
<td>60.36</td>
<td>48.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>6.81</td>
<td>6.01</td>
<td></td>
<td></td>
<td>0.99</td>
<td>11.02</td>
<td>10.60</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.50</td>
<td>99.96</td>
<td>91.49</td>
<td></td>
<td>400</td>
<td>0.50</td>
<td>100.00</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>90.87</td>
<td>87.69</td>
<td></td>
<td></td>
<td>0.60</td>
<td>100.00</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>82.45</td>
<td>78.42</td>
<td></td>
<td></td>
<td>0.70</td>
<td>100.00</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>79.01</td>
<td>60.66</td>
<td></td>
<td></td>
<td>0.80</td>
<td>98.30</td>
<td>92.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>43.23</td>
<td>30.25</td>
<td></td>
<td></td>
<td>0.90</td>
<td>91.78</td>
<td>82.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>18.87</td>
<td>13.92</td>
<td></td>
<td></td>
<td>0.95</td>
<td>59.73</td>
<td>47.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>6.60</td>
<td>5.88</td>
<td></td>
<td></td>
<td>0.99</td>
<td>12.40</td>
<td>11.40</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.50</td>
<td>100.00</td>
<td>98.89</td>
<td></td>
<td>500</td>
<td>0.50</td>
<td>100.00</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>92.05</td>
<td>89.60</td>
<td></td>
<td></td>
<td>0.60</td>
<td>100.00</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>84.98</td>
<td>80.52</td>
<td></td>
<td></td>
<td>0.70</td>
<td>100.00</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>70.32</td>
<td>63.33</td>
<td></td>
<td></td>
<td>0.80</td>
<td>100.00</td>
<td>99.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>45.02</td>
<td>38.81</td>
<td></td>
<td></td>
<td>0.90</td>
<td>90.33</td>
<td>89.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>23.23</td>
<td>22.91</td>
<td></td>
<td></td>
<td>0.95</td>
<td>65.13</td>
<td>52.58</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>7.02</td>
<td>6.70</td>
<td></td>
<td></td>
<td>0.99</td>
<td>13.84</td>
<td>13.51</td>
<td></td>
</tr>
</tbody>
</table>
Table 6.8: Estimated power of LR $\Lambda_{(n)}$ and $\Lambda^*_{(n)}$ when data are from a GW model and $P = 0.05$

<table>
<thead>
<tr>
<th>sample size</th>
<th>$\nu$</th>
<th>$\Lambda_{(n)}$</th>
<th>$\Lambda^*_{(n)}$</th>
<th>sample size</th>
<th>$\nu$</th>
<th>$\Lambda_{(n)}$</th>
<th>$\Lambda^*_{(n)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>1.0</td>
<td>99.52</td>
<td>79.27</td>
<td>100</td>
<td>1.0</td>
<td>100.00</td>
<td>99.72</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>87.92</td>
<td>43.56</td>
<td></td>
<td>2.0</td>
<td>100.00</td>
<td>92.11</td>
</tr>
<tr>
<td></td>
<td>4.0</td>
<td>50.14</td>
<td>21.63</td>
<td></td>
<td>4.0</td>
<td>97.30</td>
<td>48.43</td>
</tr>
<tr>
<td></td>
<td>8.0</td>
<td>24.95</td>
<td>13.90</td>
<td></td>
<td>8.0</td>
<td>63.48</td>
<td>24.38</td>
</tr>
<tr>
<td></td>
<td>16.0</td>
<td>13.43</td>
<td>10.19</td>
<td></td>
<td>16.0</td>
<td>24.75</td>
<td>16.28</td>
</tr>
<tr>
<td></td>
<td>50.0</td>
<td>5.85</td>
<td>8.12</td>
<td></td>
<td>50.0</td>
<td>9.07</td>
<td>13.05</td>
</tr>
<tr>
<td>50</td>
<td>1.0</td>
<td>100.00</td>
<td>98.14</td>
<td>500</td>
<td>1.0</td>
<td>100.00</td>
<td>99.87</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>99.65</td>
<td>74.08</td>
<td></td>
<td>2.0</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>4.0</td>
<td>83.19</td>
<td>32.21</td>
<td></td>
<td>4.0</td>
<td>100.00</td>
<td>97.43</td>
</tr>
<tr>
<td></td>
<td>8.0</td>
<td>41.13</td>
<td>16.22</td>
<td></td>
<td>8.0</td>
<td>99.24</td>
<td>63.59</td>
</tr>
<tr>
<td></td>
<td>16.0</td>
<td>19.61</td>
<td>11.05</td>
<td></td>
<td>16.0</td>
<td>67.17</td>
<td>42.20</td>
</tr>
<tr>
<td></td>
<td>50.0</td>
<td>7.44</td>
<td>7.31</td>
<td></td>
<td>50.0</td>
<td>15.05</td>
<td>25.53</td>
</tr>
</tbody>
</table>

geneity arises from the GW distribution. The LR tests are relatively insensitive to model misspecification. In addition, when there are nuisance parameters, the power of the LR test for detecting heterogeneity based on the GW model is higher than the corresponding power of the score test.

### 6.4 Extension

#### 6.4.1 Multivariate case

The LR tests can be easily extended to the $p$-variate case, where $p > 2$. The LR statistic in the case in which there are no nuisance parameters is

$$\Lambda_{(n)} = 2\left\{l_n(\hat{\nu}) - \sum_{j=1}^{p} l_{n,j}(\xi_j, \phi_j)\right\},$$

where $l_n(.)$ is the log-likelihood function of the $p$-variate PSW model, $l_{n,j}(\xi_j, \phi_j)$ is the Weibull log-likelihood function of the $j$th variate with known parameters $\xi_j, \phi_j$, and $\hat{\nu}$ is the unrestricted ML estimate of $\nu$ when $\xi_j$ and $\phi_j$ ($j = 1, \ldots, p$) are known.
Let $\theta = (\xi_1, \phi_1, \ldots, \xi_p, \phi_p)$, the LR statistic in the case in which there are nuisance parameters is

$$
\Lambda^*_{(n)} = 2\left\{ l_n(\hat{\nu}, \hat{\theta}) - \sum_{j=1}^{p} l_{n,j}(\hat{\xi}_j, \hat{\phi}_j) \right\},
$$

where $\hat{\nu}$ and $\hat{\theta}$ are the unrestricted ML estimates of $\nu$ and $\theta$, $\hat{\xi}_j$ and $\hat{\phi}_j$ ($j = 1, \ldots, p$) are the ML estimates of Weibull parameters $\xi_j$ and $\phi_j$ for the $j$th variate.

It is obvious that the asymptotic properties in the bivariate case carry over to the $p$-variate case.

### 6.4.2 Case with covariates

To investigate the effect of covariate on the LR tests, we performed simulation experiments to examine the rate of convergence of the LR statistics $\Lambda_{(n)}$ and $\Lambda^*_{(n)}$ when there is a covariate in the observations. The generation of the random variable $(Y_1, Y_2)$ and the covariate $x$ were done in the same way as described in Section 5.4.2. Each simulation was based on 1000 replications. The results listed in Tables 6.9 and 6.10 indicate that the presence of a covariate does not affect the large sample null properties of $\Lambda_{(n)}$ and $\Lambda^*_{(n)}$. The rates of convergence of $\Lambda_{(n)}$ and $\Lambda^*_{(n)}$ in the case with covariates are the similar to those in the case without covariates.

### 6.5 Comparison with the Score test

The obvious advantage of using score tests is that there is no need to estimate the PSW parameters, which makes computation work easier. However, to apply the score tests to data, one has to use special tables of critical values or carry out the necessary simulations. Moreover, in the application of score tests on multivariate data with $p > 2$, additional tables of critical values are needed, which are not yet available.

However, the asymptotic results based on the LR statistics are reasonable for the
Table 6.9: Estimated critical values of likelihood ratio test statistic $\Lambda_n$
(with a covariate, marginal parameters known)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Percent(%) ($\Lambda_{(n)} = 0$)</th>
<th>Quantiles(s.d.) conditional on $\Lambda_{(n)} &gt; 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$p (\Lambda_{(n)} :::: 0)$</td>
<td>0.90</td>
</tr>
<tr>
<td>100</td>
<td>57.3</td>
<td>2.67(0.22)</td>
</tr>
<tr>
<td>500</td>
<td>53.4</td>
<td>2.64(0.15)</td>
</tr>
<tr>
<td>1000</td>
<td>52.6</td>
<td>2.85(0.09)</td>
</tr>
<tr>
<td>5000</td>
<td>51.4</td>
<td>2.77(0.17)</td>
</tr>
<tr>
<td>$\infty$</td>
<td>51.0</td>
<td>2.71</td>
</tr>
</tbody>
</table>

Table 6.10: Estimated critical values of likelihood ratio test statistic $\Lambda_n^*$
(with a covariate, marginal parameters unknown)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Percent(%) ($\Lambda_{(n)} = 0$)</th>
<th>Quantiles(s.d.) conditional on $\Lambda_{(n)} &gt; 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$p (\Lambda_{(n)} :::: 0)$</td>
<td>0.90</td>
</tr>
<tr>
<td>100</td>
<td>55.9</td>
<td>2.85(0.29)</td>
</tr>
<tr>
<td>500</td>
<td>56.1</td>
<td>2.65(0.18)</td>
</tr>
<tr>
<td>1000</td>
<td>57.6</td>
<td>2.78(0.09)</td>
</tr>
<tr>
<td>5000</td>
<td>55.0</td>
<td>2.72(0.10)</td>
</tr>
<tr>
<td>$\infty$</td>
<td>51.0</td>
<td>2.71</td>
</tr>
</tbody>
</table>
use, even if the sample size is $50 \leq n \leq 100$. Furthermore, in the multivariate case with $p > 2$, the asymptotic results obtained in the bivariate case still hold. Hence, there is no need to obtain critical values for each individual multivariate case. Computational problems might arise in the maximum likelihood estimation of the PSW parameters. However, this should be eased by the two stage marginal estimation method discussed in Chapter 4.

The score and LR tests both have good power to detect heterogeneity when the model is misspecified. They are both able to accommodate the covariate information. The loss of power in the right censoring case is similar for the two tests.

### 6.6 Examples

#### 6.6.1 Infant nutrition data

The LR test was performed on the infant nutrition data as an alternative method to detect heterogeneity based on the PSW model. The maximised log-likelihood under $H_1: 0 < \nu < 1$, as shown in Table 4.13, is $l_n(\nu_1, \xi_1, \phi_1, \xi_2, \phi_2) = -241.837$; the marginal log-likelihoods under $H_0: \nu = 1$ are $-114.34$ and $-132.475$. Therefore $\Lambda^*_n = 2 \times (-241.837 + 114.34 + 132.475) = 9.950$. The critical value of the LR test at significance level $P = 0.05$ obtained from Table 6.5 is $C_{LR} = 2.61$, while the corresponding $\chi^2$ value is 2.71. Therefore, the LR test gives evidence against the null hypothesis. This result agrees with the result obtained from the score test.

#### 6.6.2 Repeated endurance exercise tests data

The dose of the drug given to the patients is considered as a covariate while we calculate the LR statistic $\Lambda^*_n$ to detect the heterogeneity in the exercise time data (1 hour and 3 hours after drug treatment). The LR statistic is $\Lambda^*_n = 2 \times (-199.01 + 101.72 + 115.47) = 36.76$. At significance level $P = 0.05$, the LR test
gives strong evidence against the null hypothesis. This result also agrees with the result obtained from the score test.

### 6.6.3 Fibre failure strength data

Similarly, the LR test was calculated on the fibre failure strength data to detect heterogeneity. The LR statistics are calculated with the combinations of each pair. They are $\Lambda^{*}_{(n)} = 40.70$ for the pair $(Y_1, Y_2)$, $\Lambda^{*}_{(n)} = 40.72$ for the pair $(Y_1, Y_3)$, $\Lambda^{*}_{(n)} = 12.92$ for the pair $(Y_1, Y_4)$, $\Lambda^{*}_{(n)} = 37.95$ for the pair $(Y_2, Y_3)$, $\Lambda^{*}_{(n)} = 5.86$ for the pair $(Y_2, Y_4)$, $\Lambda^{*}_{(n)} = 9.48$ for the pair $(Y_3, Y_4)$. All the pairs have greater $\Lambda^{*}_{(n)}$ values than $C_{LR}$ at $P = 0.05$. Therefore, heterogeneity exists in all the pairs including $(Y_2, Y_4)$, for which heterogeneity was not detected heterogeneity by the score test.
Chapter 7

Further diagnostic methods for detecting heterogeneity in the PSW model

In this chapter, we discuss some other diagnostic approaches to detecting heterogeneity in the PSW model. First, we propose a simple test which is motivated by an important aspect of the score statistic, and discuss its properties. In Section 7.2, we introduce a diagnostic method based on component-wise minima which is usually unaffected by censoring and which does not require parameter estimation. In Section 7.3, a concordance test is proposed, and its properties are discussed. Examples are given in Section 7.4 to illustrate these diagnostic methods. Finally, a discussion of the methods presented is given in Section 7.5.
7.1 A Simple Test for Frailty

7.1.1 Distribution of statistic \( S \)

Suppose that \( Y_1 \) and \( Y_2 \) have joint survivor function

\[
S(y_1, y_2) = \exp(-s^\nu),
\]

where \( s = \xi_1 y_1^{\phi_1} + \xi_2 y_2^{\phi_2} \). Then the joint density function of \( Y_1 \) and \( Y_2 \) is

\[
f(y_1, y_2) = \exp(-s^\nu) \nu s^{\nu-2} (\nu s^\nu - \nu + 1) \prod_{j=1}^{2} \xi_j \phi_j y_j^{\phi_j - 1}.
\]

Let \( S \) be the random variable corresponding to \( s \) given above. It will be useful to obtain the density of \( S \). Take a simple transformation, that is,

\[
\begin{align*}
t_1 &= \xi_1 y_1^{\phi_1}, \\
t_2 &= \xi_1 y_1^{\phi_1} + \xi_2 y_2^{\phi_2},
\end{align*}
\]

then,

\[
\begin{align*}
y_1 &= (\frac{t_2}{t_1})^{1/\phi_1}, \\
y_2 &= (\frac{t_2 - t_1}{\xi_2})^{1/\phi_2}.
\end{align*}
\]

The Jacobian is

\[
J = (\frac{t_1}{\xi_1})^{1/\phi_1} (\frac{t_2 - t_1}{\xi_2})^{1/\phi_2} \prod_{j=1}^{2} (\frac{1}{\xi_j \phi_j}).
\]

Hence, the joint distribution function of \((T_1, T_2)\) is given by

\[
g(t_1, t_2) = \exp(-t_2^\nu) [\nu^2 t_2^{2\nu-2} - \nu(\nu - 1)t_2^{\nu-2}],
\]

where \( 0 < t_1 < t_2 \). Integrating out \( t_1 \) from \( g(t_1, t_2) \) yields the marginal density of \( T_2 \), which is the density of the statistic \( S \). Therefore, the density of \( S \) is given by

\[
f(s) = \exp(-s^\nu) \nu s^{\nu-1} (\nu s^\nu - \nu + 1).
\]

Furthermore, the distribution function of \( S \) is

\[
F(s) = 1 - \exp(-s^\nu)(1 + \nu s^\nu).
\]

Note that when \( \nu = 1 \), \( S \) has a gamma distribution with shape parameter 2 and scale parameter 1, as expected, since in this case \( S \) is just the sum of two independent cumulative hazard functions, each of which has unit exponential distribution.
7.1.2 The case with no nuisance parameters

Suppose that $Y_1, \ldots, Y_n$ is a bivariate random sample of size $n$ from a population with joint survivor function (7.1). To test for the presence of heterogeneity is equivalent to testing the null hypothesis $H_0 : \nu = 1$ against the alternative hypothesis $H_1 : 0 < \nu < 1$. To fix ideas, suppose that the underlying Weibull parameters $\xi_j$ and $\phi_j$ are known so that there are no nuisance parameters.

Let $S_i$ denote the random variable $S$ for bivariate observation $i$. It is shown in Chapter 5 that an important quantity in testing for heterogeneity in this situation is $\sum_{i=1}^{n} S_i^{-1}$, a large value of which tends to indicate the presence of heterogeneity. Unfortunately, the behaviour of this quantity is non-regular since it has infinite variance under the null hypothesis. Typically, when $\sum_{i=1}^{n} S_i^{-1}$ is large it is because a small number of $S_i$ values are very small. This in turn leads to our proposing the sample minimum of the $S_i$,

$$M = \min_{i} S_i,$$

as a simple test statistic for detecting heterogeneity. Small values of $M$ supply evidence against $H_0$.

Under the null hypothesis the $S_i$ are independent, gamma random variables, each with shape parameter 2 and scale parameter 1. Consequently, it follows that the null distribution function of $M$ is given by

$$F_0(m) = 1 - \exp(-nm)(1 + m)^n.$$ 

Therefore, for a test with significance level $P$, critical value $c_P$ satisfies

$$\log(1 + c_P) - c_P = \frac{1}{n} \log(1 - P). \quad (7.4)$$

Equation (7.4) may be solved numerically for $c_P$. However, in the spirit of simplicity, an excellent approximation may be found as follows. Since in practice both $c_P$ and $P$ are small, expansion of the log terms in (7.4) gives, to a good approximation,

$$c_P \approx \sqrt{2P/n}. \quad (7.5)$$

145
Some exact critical values derived from equation (7.4) for selected values of sample size $n$ and for $P = 0.05$ are given in the first column of Table 7.1. The approximate critical values estimated from equation (7.5) are also listed in this table (second line in the first column) to show how good the approximation is. The approximate critical values are slightly smaller than the exact values.

Using (7.3), it is straightforward to show that, under the PSW model,

$$P(M > m) = P(S_1 > m, S_2 > m, \ldots, S_n > m)$$

$$= (1 + \nu m^\nu)^n \exp(-nm^\nu).$$

Hence the power of the test at significance level $P$ is given by

$$\Pi_P(\nu) = 1 - (1 + \nu c_P^\nu)^n \exp(-nc_P^\nu). \quad (7.6)$$

Values of $\Pi_P(\nu)$ for $P = 0.05$ and selected values of $\nu$ are given in Table 7.2. Note that $\Pi_P(\nu) \to 1$ as $n \to \infty$ for all values of $\nu$ in the $H_1$ region. Also, $\Pi_P(\nu) \to 1 - \exp(-n)$ as $\nu$ tends to zero for fixed $n$.

### 7.1.3 The case with nuisance parameters

In practice the Weibull parameters $\xi_j$, $\phi_j$, ($j = 1,2$) are often unknown. To cope with this situation, we propose a modified statistic $M^*$, which is of the same form as $M$ but with the $\xi_j$ and $\phi_j$ replaced by their respective maximum likelihood estimators under the null model. That is

$$M^* = \min_{i} S_i^*,$$

where $S_i^* = \hat{\xi}_1 y_{i1}^{\hat{\phi}_1} + \hat{\xi}_2 y_{i2}^{\hat{\phi}_2}$ and $\hat{\xi}_j, \hat{\phi}_j$ are the maximum likelihood estimators of $\xi_j$, $\phi_j$ under $H_0$. Thus, the simplicity of the test statistic remains since only standard maximum likelihood estimation under the Weibull model is required.

To investigate the properties of the modified test statistic, $M^*$, a simulation study was carried out. First, for various values of $n$, the null behaviour of $M^*$ was estimated on the basis of 10,000 simulated samples for each $n$. The estimated
critical values for $M^*$ at significance level $P = 0.05$ are given in Table 7.1 for illustration. These results indicate that for $n \geq 20$ the critical values for $M$ may safely be used for a test based on $M^*$. Thus, lengthy tabulations of critical values are not required in practice since approximation (7.5) may be used.

Secondly, using critical values from (7.4), the power of the test based on $M^*$ was estimated for various values of $\nu$ and $n$. The estimated power of the test is given for various combinations of $\nu$ and $n$ with a significance level $P = 0.05$. Clearly, the presence of nuisance parameters has reduced the power relative to the case with no nuisance parameters, as one might expect. However, the test based on $M^*$ is certainly a useful and simple diagnostic for the presence of heterogeneity.

As an intermediate situation that sometimes happens in practice, consider the case in which only the $\xi_j$ are unknown. For example, this might occur if the underlying Weibull distribution were thought to be of the exponential ($\phi_j = 1$) or Rayleigh ($\phi_j = 2$) type. To investigate this situation the critical values and power calculations for $M^*$ were repeated for this case as above. Selected results are included in Tables 7.1 and 7.2. Once again, the critical values for the case with no nuisance parameters may safely be used and the power in this situation is comparable with that of the case of no nuisance parameters. This in turn suggests that the loss in power of the test when all the Weibull parameters have to be estimated, is due to the maximum likelihood estimates of the Weibull shape parameters, the $\phi_j$, being most seriously affected when the independence model is fitted in the presence of dependence inducing heterogeneity.

### 7.1.4 The effect of censoring

A common feature of lifetime data is the presence of right censored observations. So it is reasonable to ask how $M$ and $M^*$ behave in the presence of right censoring. A simulation experiment was run to investigate the effect of relatively simple right censoring where only relatively large observations are censored. Estimated critical values are given in Table 7.1 for the three cases discussed above (no nuisance para-
meters, Weibull scale parameters unknown, all Weibull parameters unknown) for selected values of \( n \) and for significance probability 0.05. The censoring mechanism used was the same as that mentioned in the previous chapters. The right censored observations were above a fixed threshold in each margin, where the thresholds were chosen so that, on average, 30% of bivariate observations have at least one censored component. As before, each critical value listed in Table 7.1 shows that the critical values derived from (7.4), or approximated by (7.5), may still be used safely for \( n \geq 20 \) even with 30% right censoring.

Corresponding power calculations were also carried out and the results for 30% censoring are given in Table 7.2. Once again, for each of the three cases discussed (no nuisance parameters, Weibull scale parameters unknown, all Weibull parameters unknown), the power figures for each case show little difference between the situation with no censoring and that with 30% censoring. This is perhaps not surprising because information lost in the upper tail due to right censoring is relatively unimportant for our purposes where the lower tail is of concern.

### 7.2 A diagnostic based on component-wise minima

Consider the case in which \( Y_1, Y_2, \ldots, Y_n \) is a bivariate random sample of size \( n \) from a population with survival function (7.1), where \( Y_i = (Y_{i1}, Y_{i2}) \). Let \( W_1 = \min_i \{Y_{i1}\} \) and \( W_2 = \min_i \{Y_{i2}\} \). Thus, \( W_1 \) and \( W_2 \) are the component-wise minima.

In the same spirit as the previous section, we note that \( S_i \) will certainly be small if both cumulative hazard components of observation \( i \) are small. In particular we focus attention on the situation in which both components of a single bivariate observation are the component-wise minima. Let \( A \) be the event \( \exists i \), such that \( (Y_{i1}, Y_{i2}) = (W_1, W_2) \). In other words \( A \) corresponds to there being an actual bivariate observation whose components are the component-wise minima. The
Table 7.1: Critical values of the test $M$ and $M^*$ for $P = 0.05$

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>$C_1$</th>
<th>$C_2$</th>
<th>$C_3$</th>
<th>$C_4$</th>
<th>$C_5$</th>
<th>$C_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.0733</td>
<td>0.0731</td>
<td>0.0770</td>
<td>0.0735</td>
<td>0.0778</td>
<td>0.0761</td>
</tr>
<tr>
<td></td>
<td>0.0707</td>
<td>(0.0017)</td>
<td>(0.0017)</td>
<td>(0.0017)</td>
<td>(0.0016)</td>
<td>(0.0015)</td>
</tr>
<tr>
<td>30</td>
<td>0.0596</td>
<td>0.0597</td>
<td>0.0615</td>
<td>0.0600</td>
<td>0.0627</td>
<td>0.0611</td>
</tr>
<tr>
<td></td>
<td>0.0577</td>
<td>(0.0021)</td>
<td>(0.0014)</td>
<td>(0.0013)</td>
<td>(0.0011)</td>
<td>(0.0013)</td>
</tr>
<tr>
<td>40</td>
<td>0.0515</td>
<td>0.0518</td>
<td>0.0529</td>
<td>0.0512</td>
<td>0.0535</td>
<td>0.0527</td>
</tr>
<tr>
<td></td>
<td>0.0500</td>
<td>(0.0012)</td>
<td>(0.0012)</td>
<td>(0.0012)</td>
<td>(0.0013)</td>
<td>(0.0015)</td>
</tr>
<tr>
<td>50</td>
<td>0.0460</td>
<td>0.0467</td>
<td>0.0468</td>
<td>0.0461</td>
<td>0.0483</td>
<td>0.0470</td>
</tr>
<tr>
<td></td>
<td>0.0447</td>
<td>(0.0009)</td>
<td>(0.0012)</td>
<td>(0.0012)</td>
<td>(0.0011)</td>
<td>(0.0011)</td>
</tr>
<tr>
<td>70</td>
<td>0.0388</td>
<td>0.0391</td>
<td>0.0393</td>
<td>0.0386</td>
<td>0.0402</td>
<td>0.0392</td>
</tr>
<tr>
<td></td>
<td>0.0378</td>
<td>(0.0004)</td>
<td>(0.0008)</td>
<td>(0.0009)</td>
<td>(0.0009)</td>
<td>(0.0009)</td>
</tr>
<tr>
<td>100</td>
<td>0.0324</td>
<td>0.0328</td>
<td>0.0325</td>
<td>0.0323</td>
<td>0.0331</td>
<td>0.0325</td>
</tr>
<tr>
<td></td>
<td>0.0316</td>
<td>(0.0008)</td>
<td>(0.0008)</td>
<td>(0.0009)</td>
<td>(0.0009)</td>
<td>(0.0007)</td>
</tr>
</tbody>
</table>

- $C_1$: no nuisance parameters, no censoring
- $C_2$: all parameters unknown, no censoring
- $C_3$: only parameters $\xi_j$ (j=1,2) unknown, no censoring
- $C_4$: no nuisance parameters, 30% censoring
- $C_5$: all parameters unknown, 30% censoring
- $C_6$: only parameters $\xi_j$ (j=1,2) unknown, 30% censoring

Note that the $C_1$ figures on the first lines are exact, using equation (7.4). The $C_1$ figures on the second lines are approximations, using equation (7.5). Values in parentheses are estimated standard errors.
Table 7.2: Estimated power(%) of the tests based on $M$ and $M^*$ for $P = 0.05$

<table>
<thead>
<tr>
<th>sample size</th>
<th>$\nu$</th>
<th>$P_1$</th>
<th>$P_2$</th>
<th>$P_3$</th>
<th>$P_4$</th>
<th>$P_5$</th>
<th>$P_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.2</td>
<td>100.00</td>
<td>57.52</td>
<td>100.00</td>
<td>100.00</td>
<td>56.41</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>99.16</td>
<td>51.23</td>
<td>100.00</td>
<td>99.23</td>
<td>50.73</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>84.56</td>
<td>40.67</td>
<td>90.82</td>
<td>84.31</td>
<td>40.44</td>
<td>85.73</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>44.24</td>
<td>25.81</td>
<td>46.43</td>
<td>43.86</td>
<td>24.98</td>
<td>41.55</td>
</tr>
<tr>
<td>30</td>
<td>0.2</td>
<td>100.00</td>
<td>63.20</td>
<td>100.00</td>
<td>100.00</td>
<td>62.35</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>100.00</td>
<td>59.12</td>
<td>100.00</td>
<td>100.00</td>
<td>57.74</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>91.87</td>
<td>48.46</td>
<td>95.51</td>
<td>91.94</td>
<td>46.12</td>
<td>92.72</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>52.65</td>
<td>29.61</td>
<td>55.03</td>
<td>52.08</td>
<td>29.14</td>
<td>49.44</td>
</tr>
<tr>
<td>40</td>
<td>0.2</td>
<td>100.00</td>
<td>68.07</td>
<td>100.00</td>
<td>100.00</td>
<td>67.27</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>99.99</td>
<td>62.33</td>
<td>100.00</td>
<td>100.00</td>
<td>61.72</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>94.65</td>
<td>51.86</td>
<td>98.25</td>
<td>95.35</td>
<td>51.98</td>
<td>96.52</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>57.28</td>
<td>34.33</td>
<td>62.73</td>
<td>57.47</td>
<td>33.92</td>
<td>55.00</td>
</tr>
<tr>
<td>50</td>
<td>0.2</td>
<td>100.00</td>
<td>72.75</td>
<td>100.00</td>
<td>100.00</td>
<td>71.34</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>100.00</td>
<td>67.46</td>
<td>99.93</td>
<td>100.00</td>
<td>66.82</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>97.68</td>
<td>55.38</td>
<td>99.11</td>
<td>97.25</td>
<td>55.42</td>
<td>98.64</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>62.52</td>
<td>37.32</td>
<td>66.48</td>
<td>62.46</td>
<td>35.54</td>
<td>59.52</td>
</tr>
<tr>
<td>70</td>
<td>0.2</td>
<td>100.00</td>
<td>78.48</td>
<td>100.00</td>
<td>100.00</td>
<td>77.67</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>100.00</td>
<td>71.46</td>
<td>100.00</td>
<td>100.00</td>
<td>72.56</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>99.43</td>
<td>61.37</td>
<td>99.96</td>
<td>98.16</td>
<td>60.08</td>
<td>99.17</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>69.78</td>
<td>40.16</td>
<td>73.86</td>
<td>69.24</td>
<td>39.34</td>
<td>66.17</td>
</tr>
<tr>
<td>100</td>
<td>0.2</td>
<td>100.00</td>
<td>82.77</td>
<td>100.00</td>
<td>100.00</td>
<td>83.95</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>100.00</td>
<td>79.78</td>
<td>100.00</td>
<td>100.00</td>
<td>78.51</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>99.97</td>
<td>66.42</td>
<td>100.00</td>
<td>99.91</td>
<td>67.41</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>76.82</td>
<td>47.22</td>
<td>79.44</td>
<td>75.43</td>
<td>47.66</td>
<td>73.80</td>
</tr>
</tbody>
</table>

$P_1$: no nuisance parameters, no censoring
$P_2$: all parameters unknown, no censoring
$P_3$: only parameters $\xi_j$ ($j=1,2$) unknown, no censoring
$P_4$: no nuisance parameters, 30% censoring
$P_5$: all parameters unknown, 30% censoring
$P_6$: only parameters $\xi_j$ ($j=1,2$) unknown, 30% censoring

Note that the $P_1$ figures are exact, using equation (7.6).
probability of $A$ may be derived as follows:

$$P(A) = n P(Y_{11} = W_1, Y_{12} = W_2)$$

$$= n \int_0^\infty \int_0^\infty P(Y_{21} > y_1, \ldots, Y_{n1} > y_1, Y_{22} > y_2, \ldots, Y_{n2} > y_2$$

$$\mid Y_{11} = y_1, Y_{12} = y_2) f(y_1, y_2) \, dy_1 \, dy_2$$

$$= n \int_0^\infty \int_0^\infty [S(y_1, y_2)]^{n-1} f(y_1, y_2) \, dy_1 \, dy_2$$

$$= n \int_0^\infty \int_0^\infty \exp\{-n z^\nu\} \{\nu z^{2(\nu-1)} - \nu(\nu - 1) z^{\nu-2}\} \, dy_1 \, dy_2$$

$$= \frac{1}{n} + \frac{(n-1)(1-\nu)}{n} \quad (7.7)$$

where $z = \xi_1 y_1^{\phi_1} + \xi_2 y_2^{\phi_2}$. Thus, $P(A) = \frac{1}{n} \{1 + (n-1)(1-\nu)\}$.

Under $H_0 : \nu = 1$, where the two components are completely independent, $P(A) = \frac{1}{n}$, as expected. Also, when $\nu = 0$, so that the two components are completely dependent, $P(A) = 1$, as expected. Further, $P(A) > \frac{1}{2}$ whatever the value of $n$ when $\nu \leq \frac{1}{2}$. Thus, event $A$ is more likely to occur than not when there is strong dependence between components.

This provides us with a straightforward and useful diagnostic method for detecting the presence of heterogeneity when $n$ is not small. If $A$ does not occur, then little information about heterogeneity can be inferred. However, providing $n$ is not small ($n \geq 20$, say, so that the significance probability is no larger than 0.05), if $A$ does occur, then this is strong evidence against the null hypothesis.

The beauty of this diagnostic is that it is immediate from an eye-balling of the data, no parameter estimation is required and it is unaffected by simple Type I or Type II censoring. Even with more complex censoring schemes many data configurations will allow occurrence or non-occurrence of $A$ to be decided upon.
7.3 A concordance test

7.3.1 Definition of the statistic U

Suppose that a random sample \( Y_1, \ldots, Y_n \) of size \( n \) is from a bivariate PSW model, where \( Y_i = (Y_{i1}, Y_{i2}) \). For \( 1 \leq i < j \leq n \),

\[
Z_{ij} = \begin{cases} 
1 & \text{if the pair } (i, j) \text{ is concordant} \\
0 & \text{otherwise}
\end{cases},
\]

(7.8)

where concordance of the pair \((i, j)\) corresponds to the events \( A_1 \cup A_2 \), where \( A_1 = \{Y_{i1} < Y_{j1}, Y_{i2} < Y_{j2}\} \) and \( A_2 = \{Y_{i1} > Y_{j1}, Y_{i2} > Y_{j2}\} \). The statistic \( U \) is defined as

\[
U = \frac{1}{C_n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Z_{ij}.
\]

We propose a concordance test based on the statistic \( U \) to detect the heterogeneity in the PSW model, where \( U \) is analogous to Kendall’s tau (Kendall, 1938). It is apparent that \( U \) is invariant under monotone transformations. Furthermore, as with Kendall’s tau, it is a \( U \)-statistic (Hoeffding, 1948) so that it is asymptotically normal. In terms of these features, it is natural that \( U \) is considered as an alternative measure of dependence between random variables.

7.3.2 Properties of the statistic \( U \)

The concordance of a pair of observations \( Y_i \) and \( Y_j \) is equivalent to the case in which either \( Y_i \) or \( Y_j \) is the component-wise minimum. By taking \( n = 2 \) in equation (7.7), we have

\[
P(Z_{ij} = 1) = 1 - \frac{\nu}{2},
\]

and

\[
P(Z_{ij} = 0) = \frac{\nu}{2}.
\]

Therefore, the expectation and variance of random variable \( Z_{ij} \) are

\[
E(Z_{ij}) = 1 - \frac{\nu}{2},
\]

(7.9)
The expectation of $U$ is easily obtained thereafter, which is

$$E(U) = \frac{1}{C_n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E(Z_{ij}) = E(Z_{ij}) = 1 - \frac{\nu}{2}.$$  

We now consider the calculation of the variance of $U$. Note that $Z_{ij}$ and $Z_{kl}$ are independent if they share no common subscript. Therefore, we only need to consider $Cov(Z_{ij}, Z_{kl})$ for the pairs of $(Z_{ij}, Z_{kl})$ with one common subscript. Since $Z_{ij}$ represents the same event as $Z_{ji}$, to fix the idea, we denote the covariance of $(Z_{ij}, Z_{kl})$ with one common subscript as $Cov(Z_{ij}, Z_{il})$, where $1 \leq i < j \leq n$, $1 \leq k < l \leq n$, and $j \neq l$.

For $1 \leq i < j \leq n$, the total number of all the possible pairs of $(i, j)$ is $C_n^2 = \frac{n(n-1)}{2}$. Hence, the total number of all the possible pairs of $(Z_{ij}, Z_{kl})$ is $C_n^2 = \frac{n(n-1)}{2}$, where $1 \leq i < j \leq n$, $1 \leq k < l \leq n$, and $(i, j) \neq (k, l)$. For $1 \leq i, j, k, l \leq n$, and $i \neq j \neq k \neq l$, the total number of all the combinations of $(i, j, k, l)$ is $C_n^4$ so that the total number of all the pairs of $(Z_{ij}, Z_{kl})$ with no common subscript is $3C_n^4$.

Let $N_0$ be the number of all the pairs with one common subscript. It follows that

$$N_0 = \text{number of all the possible pairs of } (Z_{ij}, Z_{kl}) - \text{number of all the pairs with no common subscript}$$

$$= C_n^2 \frac{n(n-1)}{2} - 3C_n^4$$

$$= \frac{1}{2} n(n-1)(n-2).$$

Therefore, the variance of $U$ is given by

$$Var(U) = \left(\frac{1}{C_n^2}\right)^2 Var\left(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Z_{ij}\right)$$

$$= \frac{1}{C_n^2} Var(Z_{ij}) + \left(\frac{1}{C_n^2}\right)^2 \times N_0 \times 2Cov(Z_{ij}, Z_{il})$$

$$= \frac{2}{n(n-1)}[Var(Z_{ij}) + 2(n-2)Cov(Z_{ij}, Z_{il})]$$

$$= \frac{2}{n(n-1)}\{Var(Z_{ij}) + 2(n-2)[E(Z_{ij}Z_{il}) - E(Z_{ij})E(Z_{il})]\}. \quad (7.11)$$
To evaluate $E(Z_{ij}Z_{il})$, let

$E_1 = P(\text{ith observation is the component-wise maximum of the } i, j, l\text{th observations})$;

$E_2 = P(\text{ith observation is the component-wise minimum of the } i, j, l\text{th observations})$;

$E_3 = P(\text{ith observation is the component-wise median of the } i, j, l\text{th observations})$.

$E_1, E_2$ and $E_3$ can be obtained through the corresponding joint and marginal survivor functions.

$$
E_1 = \int_0^\infty \int_0^\infty F^2(y_1, y_2)f(y_1, y_2)dy_1dy_2,
$$

$$
E_2 = \int_0^\infty \int_0^\infty S^2(y_1, y_2)f(y_1, y_2)dy_1dy_2,
$$

$$
E_3 = 2\int_0^\infty \int_0^\infty F(y_1, y_2)S(y_1, y_2)f(y_1, y_2)dy_1dy_2,
$$

where $F(y_1, y_2) = 1 - S(y_1) - S(y_2) + S(y_1)S(y_2)$. Therefore, we have

$$
E(Z_{ij}Z_{il}) = P(Z_{ij}Z_{il} = 1) - P(z_{ij} = 1 \& z_{il} = 1)
$$

$$
= E_1 + E_2 + E_3
$$

$$
= \int_0^\infty \int_0^\infty \{1 + S_1^2(y_1) + S_2^2(y_2) + 4S^2(y_1, y_2) - 2S_1(y_1) - 2S_2(y_2)
$$

$$
+4S(y_1, y_2) + 2S_1(y_1)S_2(y_2) - 4S_1(y_1)S_2(y_1, y_2)
$$

$$
-4S_2(y_2)S(y_1, y_2)\}f(y_1, y_2)dy_1dy_2
$$

(7.12)

where $S_1(y_1)$ and $S_2(y_2)$ are the marginal survivor functions. They may be taken as random variables with uniform distribution in $[0, 1]$. Hence, their corresponding first and second moments are

$$
E(S_j(y_j)) = \int_0^\infty \int_0^\infty S_j(y_j)f(y_1, y_2)dy_1dy_2 = \frac{1}{2},
$$

$$
E(S_j^2(y_j)) = \int_0^\infty \int_0^\infty S_j^2(y_j)f(y_1, y_2)dy_1dy_2 = \frac{1}{3},
$$

where $j = 1, 2$.

Under the null hypothesis, the two components are independent. Therefore

$$
S(y_1, y_2) = S_1(y_1)S_2(y_2),
$$
so that we may easily obtain

\[ E(Z_{ij}Z_{il}) = 1 + 2E(S_I^2(y_i)) + 4E^2(S_I^2(y_i)) - 4E(S_I^2(y_i)) + 6E^2(S_I(y_i)) - 8E(S_I^2(y_i))E(S_I(y_i)) = \frac{5}{18} \]  

(7.13)

Therefore, the expectation and variance of the statistic \( U \) under \( H_0 \) are

\[ E(U) = \frac{1}{2}, \]  

(7.14)

\[ Var(U) = \frac{2n + 5}{18n(n - 1)}. \]  

(7.15)

When \( Y_1 \) and \( Y_2 \) are independent, we have asymptotically

\[ \frac{U - \frac{1}{2}}{\sqrt{\frac{2n + 5}{18n(n - 1)}}} \rightarrow N(0,1) \quad \text{as} \quad n \rightarrow \infty. \]

At significance level \( P \), the approximate critical value \( C_P \) of the test is

\[ C_P = z_P \sqrt{\frac{2n + 5}{18n(n - 1)}} + \frac{1}{2}, \]  

(7.16)

where \( \Phi(z_P) = 1 - P \), and \( \Phi(.) \) is the standard normal distribution function.

Under the alternative hypothesis, where \( Y_1 \) and \( Y_2 \) have bivariate PSW distribution, the marginal survivor function of \( Y_j \), \( j = 1, 2 \), is

\[ S_j(y_j) = \exp(-\xi_j^{\nu}y_j^{\phi_j^{\nu}}). \]

Without loss of generality, we may take \( \xi_j = \phi_j = 1 \). Under \( H_1 \), we have

\[ E[S(y_1, y_2)] = \int_0^\infty \int_0^\infty S(y_1, y_2)f(y_1, y_2)dy_1dy_2 = \frac{1}{2} - \frac{1}{4}^{\nu}, \]

and

\[ E[S^2(y_1, y_2)] = \int_0^\infty \int_0^\infty S^2(y_1, y_2)f(y_1, y_2)dy_1dy_2 = \frac{1}{3} - \frac{2}{9}^{\nu}. \]

It seems impossible to derive analytic expressions for the remaining unknown terms in the equation (7.12), which are, for \( j = 1, 2 \),

\[ E[S_j(y_j)S(y_1, y_2)] = \int_0^\infty \int_0^\infty S_j(y_j)S(y_1, y_2)f(y_1, y_2)dy_1dy_2, \]  

(7.17)
and

$$E[S_1(y_1)S_2(y_2)] = \int_0^\infty \int_0^\infty S_1(y_1)S_2(y_2)f(y_1,y_2)dy_1dy_2. \quad (7.18)$$

However, the above integrals can be simplified and reduced to single integrals via variable transformation as follows:

$$E[S_3(y_3)S(y_1,y_2)] = \int_0^\infty \int_0^\infty \exp(-y_1^\nu) \exp(-2(y_1 + y_2)^\nu)[\nu^2(y_1 + y_2)^{2\nu-2} - \nu(\nu - 1)(y_1 + y_2)^{\nu-2}]dy_1dy_2$$

$$= I_1 - I_2,$$

and

$$E[S_1(y_1)S_2(y_2)] = \int_0^\infty \int_0^\infty \exp(-y_1^\nu) \exp(-y_2^\nu) \exp[(-(y_1 + y_2)^\nu)[\nu^2(y_1 + y_2)^{2\nu-2} - \nu(\nu - 1)(y_1 + y_2)^{\nu-2}]dy_1dy_2$$

$$= J_1 - J_2,$$

where

$$I_1 = \nu^2 \int_0^\infty \int_0^\infty \exp[-y_1^\nu - 2(y_1 + y_2)^\nu](y_1 + y_2)^{2\nu-2}dy_1dy_2$$

$$= \nu^2 \int_0^1 \int_0^\infty \exp[-r^\nu(1 + 2s^\nu)]r^{2\nu-2}drds$$

$$= \nu \int_0^1 (\int_0^\infty \exp[-t(s^\nu + 1)]ttdt)ds$$

$$= \nu \int_0^1 \frac{1}{(s^\nu + 1)^2}ds,$$

$$I_2 = \nu(\nu - 1) \int_0^\infty \int_0^\infty \exp[-y_1^\nu - 2(y_1 + y_2)^\nu](y_1 + y_2)^{\nu-2}dy_1dy_2$$

$$= \nu(\nu - 1) \int_0^1 \int_0^\infty \exp[-r^\nu(1 + 2s^\nu)]r^{\nu-2}drds$$

$$= (\nu - 1) \int_0^1 (\int_0^\infty \exp[-t(s^\nu + 1)]ttdt)ds$$

$$= \nu \int_0^1 \frac{1}{s^\nu + 2}ds,$$

$$J_1 = \nu^2 \int_0^\infty \int_0^\infty \exp[-y_1^\nu + y_2^\nu - (y_1 + y_2)^\nu](y_1 + y_2)^{2\nu-2}dy_1dy_2$$

$$= \nu^2 \int_0^1 \int_0^\infty \exp[-r^\nu(1 + s^\nu)]r^{2\nu-2}drds$$

$$= \nu \int_0^1 (\int_0^\infty \exp[-t(1 + s^\nu)]ttdt)ds$$

$$= v_1 - v_2,$$

$$J_2 = \nu(\nu - 1) \int_0^\infty \int_0^\infty \exp[-y_1^\nu + y_2^\nu - (y_1 + y_2)^\nu](y_1 + y_2)^{\nu-2}dy_1dy_2$$

$$= \nu(\nu - 1) \int_0^1 \int_0^\infty \exp[-r^\nu(1 + s^\nu)]r^{\nu-2}drds$$

$$= (\nu - 1) \int_0^1 (\int_0^\infty \exp[-t(1 + s^\nu)]ttdt)ds$$

$$= \nu \int_0^1 \frac{1}{s^\nu + 2}ds.$$
\[
\begin{align*}
J_2 &= \nu(\nu - 1) \int_0^1 \int_0^\infty \exp[-y_1^\nu - y_2^\nu - (y_1 + y_2)^\nu] (y_1 + y_2)^{\nu - 2} dy_1 dy_2 \\
&= \nu(\nu - 1) \int_0^1 \int_0^\infty \exp[-t(1 + s^\nu + (1 - s)^\nu)] t^{\nu - 1} dt ds \\
&= (\nu - 1) \int_0^1 \left( \int_0^\infty \exp[-t(1 + s^\nu + (1 - s)^\nu)] dt \right) ds
\end{align*}
\]

Numerical evaluation of the above simplified single integrals can be achieved by calling NAG subroutine D01AHF. According to the evaluated results, quadratic curves are fitted in terms of \( \nu \), which represent good approximations of the above integrals. Hence we have

\[
E[S_j(y_j)S(y_1, y_2)] \approx 0.335 - 0.126\nu - 0.043\nu^2,
\]

and

\[
E[S_1(y_1)S_2(y_2)] \approx 0.336 - 0.03\nu - 0.060\nu^2.
\]

Figure 7.1 displays the evaluated integrals for various \( \nu \) values and the fitted quadratic curves. The approximations are clearly satisfactory. For the special case \( \nu = 1 \), the fitted models give approximations \( E[S_j(y_j)S(y_1, y_2)] = 0.166 \), and \( E[S_1(y_1)S_2(y_2)] = 0.246 \), while exact results under \( H_0 \) give \( E[S_j(y_j)S(y_1, y_2)] = \frac{1}{6} \), and \( E[S_1(y_1)S_2(y_2)] = \frac{1}{4} \).

Therefore, under the alternative hypothesis, we have approximately

\[
E[Z_{ij}Z_{id}] = 0.992 - 0.941\nu + 0.224\nu^2.
\]  

(7.19)

According to the above approximation, \( E[Z_{ij}Z_{id}] \) is approximately equal to 0.275 when \( \nu = 1 \). Its exact value shown in (7.13) is \( \frac{5}{18} = 0.2778 \). Furthermore, \( Var(U) \) may be evaluated from

\[
Var(U) = \frac{2}{n(n - 1)} \left\{ \frac{1}{2} \nu(1 - \frac{\nu}{2}) - 2(n - 2)(0.026\nu^2 - 0.059\nu + 0.008) \right\}
\]  

(7.20)
In terms of the asymptotic normality of $U$ under $H_1$, the power of the $U$ test under the significance level $P$ can be obtained. That is

$$\Pi_P(\nu) = P(U > C_P)$$

$$= P\left(\frac{U - 1 + \frac{\nu}{2}}{\sqrt{\text{Var}(U)}} > \frac{C_P - 1 + \frac{\nu}{2}}{\sqrt{\text{Var}(U)}}\right)$$

$$= \Phi\left(\frac{C_P - 1 + \frac{\nu}{2}}{\sqrt{\text{Var}(U)}}\right)$$

where $C_P$ is the critical value which may be obtained from (7.16). Numerical evaluation of $\text{Var}(U)$ under $H_1$ may be obtained through the expression (7.20).

The estimated powers of the concordance test based on $U$ in the no censoring case for various values of sample size and $\nu$ are given in Table 7.4, where the significance level $P$ is chosen as 0.05, critical values are obtained from (7.16). $P_1$ in Table 7.4 shows higher powers than those of the test based on $M^*$.

### 7.3.3 Comparison with Kendall’s tau

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a continuous bivariate distribution, where $Y_i = (Y_{i1}, Y_{i2})$, Kendall (1938) considers the statistic $\tau$ (Kendall’s tau) as
a measure of the association between the components of the bivariate random variable \( Y \). For \( 1 \leq i < j \leq n \), let

\[
Z_{ij}' = \begin{cases} 
1 & \text{if } (Y_{i1} - Y_{j1})(Y_{i2} - Y_{j2}) > 0 \\
-1 & \text{if } (Y_{i1} - Y_{j1})(Y_{i2} - Y_{j2}) < 0 
\end{cases} 
\]  

(7.21)

The statistic \( T \) is defined as

\[
T = \frac{1}{C_n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Z_{ij}'. 
\]

We can see that Kendall’s tau is the overall average of the number of concordants minus the number of discordants, which is analogous to our definition of \( U \). Since the sum of the number of concordants and the number of discordants is \( C_n^2 \), the total number of pairs of comparisons which can be made, the relationship between \( \tau \) and \( U \) is

\[
\tau = 2U - 1. 
\]  

(7.22)

Kendall’s tau is a U-statistic so that it is asymptotically normal, see Hoeffding (1948). Under \( H_0 \), the two components are independently Weibull distributed. The expectation and variance of \( \tau \) can be obtained through the equation (7.22).

\[
E(\tau) = 2E(U) - 1 = 1 - \nu, 
\]

and

\[
Var(\tau) = 4Var(U) = \frac{2(2n + 5)}{9n(n - 1)}. 
\]

The expectation of \( \tau \) obtained agrees with the result from Chapter 3, Section 3.3.2, and the variance of \( \tau \) obtained agrees with the general result from Hoeffding (1948). Therefore, for significance level \( P \), the critical value \( C'_P \) of the test based on Kendall’s tau is

\[
C'_P = z_P \sqrt{\frac{2(2n + 5)}{9n(n - 1)}}, 
\]  

(7.23)

where \( \Phi(z_P) = P \), and \( \Phi(.) \) is the standard normal distribution function.

Following the asymptotic normality of the statistic \( \tau \), similar as the \( U \) test, the power of the test based on Kendall’s tau is obtained as

\[
\Pi_P(\nu) = \Psi \left( \frac{C'_P - 1 + \nu}{2\sqrt{Var(U)}} \right), 
\]
where \( C_p' \) is the critical value form \((7.23)\), \( Var(U) \) under \( H_1 \) can be evaluated from \((7.20)\).

Since there is a linear relationship between \( U \) and \( \tau \), the power of the test based on Kendall’s tau is the same as the power of the test based on the \( U \) statistic.

### 7.3.4 Censoring

So far, we have only considered the situation in which there is no censoring. To accommodate testing for heterogeneity in bivariate data subject to censoring, a modification of the statistic \( U \) is given as follows. Suppose that a bivariate sample \((Y_{i1}, Y_{i2}) (i = 1, \ldots, n)\) may be censored on the right, that is \( Y_{ik} = \min\{Y_{ik}^t, Y_{ik}^c\} \), where \( Y_{ik}^t \) denotes \( Y_{ik} \) is observed; \( Y_{ik}^c \) denotes \( Y_{ik} \) is censored; \( k = 1, 2 \). For \( 1 \leq i < j \leq n \),

\[
Z_{ij} = \begin{cases} 
1 & \text{if the pair } (i, j) \text{ is definitely concordant} \\
0 & \text{otherwise}
\end{cases}, \quad (7.24)
\]

where the pair \((i, j)\) is definitely concordant means that \( A^c_1 \cup A^c_2 \) occurs where

\[
A^c_1 = \{Y_{i1}^d < Y_{j1}, Y_{i2}^d < Y_{j2}\} \quad \text{and} \quad A^c_2 = \{Y_{i1}^d > Y_{j1}, Y_{i2}^d > Y_{j2}\}.
\]

The symbols \(<^{d}\) and \(>^{d}\) mean “definitely less than” and “definitely greater than” respectively. \( Y_{ik}^d > Y_{jk}^d \) occurs when both \( Y_{ik} \) and \( Y_{jk} \) are uncensored and \( Y_{ik}^t > Y_{jk}^t \), or when \( Y_{ik} \) is censored at \( Y_{ik}^c \) and \( Y_{jk} \) is observed with \( Y_{ik}^c > Y_{jk}^t \). Similarly, \( Y_{ik}^d < Y_{jk}^d \) occurs when both \( Y_{ik} \) and \( Y_{jk} \) are uncensored and \( Y_{ik}^t < Y_{jk}^t \), or when \( Y_{ik} \) is observed and \( Y_{jk} \) is censored at \( Y_{jk}^c \) with \( Y_{ik}^t < Y_{jk}^c \). The test statistic \( U \) is defined as

\[
U = \frac{1}{C_n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Z_{ij}.
\]

This is an extended version of the \( U \) statistic in Section 7.3.1.

As an extension of the test based on Kendall’s tau, Oakes (1982) has proposed a modified test for heterogeneity in bivariate censored survival times. His definition of extended Kendall’s tau is the same as our definition above. It is shown that the asymptotic normality of the extended Kendall’s tau still follows from the results of
Hoeffding (1948). However, as one might expect, this extended test should not be used when there is very heavy censoring because of loss of too much information. Similar work was done by Weier & Basu (1980). They gave several modifications of Kendall’s tau to accommodate testing for independence in bivariate data subject to censoring. Oakes’ (1982b) extended version of Kendall’s tau is the same as the “simple adjusted” tau referred to in Weier & Basu’s work.

We performed a simulation experiment to investigate the performance of the $U$ statistic under $H_0$ and the power of the $U$ test in the case with censoring. The simulation study was carried out for various values of sample size and for the cases with 15%, 30% and 50% censoring of either or both components of observations. The censoring mechanism used is the same as stated in previous chapters.

On the basis of 10,000 simulation, we estimated the mean and variance of statistic $U$ under $H_0$ for various sample sizes and censoring situations, see Table 7.3. For each sample size, the bottom line in Table 7.3 represents the mean and variance of $U$ corresponding to the no censoring case, which are obtained from (7.14) and (7.15). Simulation results also confirm that, for various censoring situations, $U$ is asymptotically normal, as $n \to \infty$. The more censored cases there are, the further away from 0.5 is the mean of $U$, and the higher the variance of $U$.

Using the results from Table 7.3 and the asymptotic normal property, the powers of the concordance test based on $U$ in the censoring case were estimated at significance level $P = 0.05$. $P_2$, $P_3$ and $P_4$ listed in table 7.4 correspond to estimated powers in the cases with 15%, 30% and 50% right censoring for various sample sizes. The results indicate that the presence of censoring has reduced the power of the test relative to the case with no censoring. As one might expect, the more censored cases there are, the less information we have to detect the heterogeneity, and the less powerful is the test. The power of the concordance test based on $U$ when there is 30% censoring is almost the same as the power of the $M^*$ test in the same censoring situation.
Table 7.3: Estimated mean and variance of statistic $U$ under $H_0$

<table>
<thead>
<tr>
<th>sample size</th>
<th>censoring (%)</th>
<th>mean</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>50.0</td>
<td>0.4137</td>
<td>0.008045</td>
</tr>
<tr>
<td></td>
<td>30.0</td>
<td>0.4741</td>
<td>0.006845</td>
</tr>
<tr>
<td></td>
<td>15.0</td>
<td>0.4930</td>
<td>0.006523</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.5000</td>
<td>0.006579</td>
</tr>
<tr>
<td>50</td>
<td>50.0</td>
<td>0.4137</td>
<td>0.003034</td>
</tr>
<tr>
<td></td>
<td>30.0</td>
<td>0.4734</td>
<td>0.002498</td>
</tr>
<tr>
<td></td>
<td>15.0</td>
<td>0.4942</td>
<td>0.002381</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.5000</td>
<td>0.002381</td>
</tr>
<tr>
<td>100</td>
<td>50.0</td>
<td>0.4134</td>
<td>0.001514</td>
</tr>
<tr>
<td></td>
<td>30.0</td>
<td>0.4732</td>
<td>0.001194</td>
</tr>
<tr>
<td></td>
<td>15.0</td>
<td>0.4945</td>
<td>0.001152</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.5000</td>
<td>0.001150</td>
</tr>
</tbody>
</table>

Table 7.4: Estimated power(%) of the concordance test based on $U$ statistic

<table>
<thead>
<tr>
<th>sample size</th>
<th>$\nu$</th>
<th>$P_1$</th>
<th>$P_2$</th>
<th>$P_3$</th>
<th>$P_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.5</td>
<td>97.07</td>
<td>75.21</td>
<td>42.12</td>
<td>18.81</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>84.03</td>
<td>57.44</td>
<td>33.43</td>
<td>15.02</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>59.40</td>
<td>43.83</td>
<td>23.92</td>
<td>9.75</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>32.86</td>
<td>26.27</td>
<td>15.78</td>
<td>7.51</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>13.90</td>
<td>13.52</td>
<td>10.75</td>
<td>6.19</td>
</tr>
<tr>
<td>50</td>
<td>0.5</td>
<td>100.00</td>
<td>97.37</td>
<td>73.44</td>
<td>22.33</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>99.87</td>
<td>89.87</td>
<td>57.82</td>
<td>16.54</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>95.06</td>
<td>70.48</td>
<td>40.07</td>
<td>13.77</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>67.39</td>
<td>48.43</td>
<td>25.71</td>
<td>8.95</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>25.55</td>
<td>20.45</td>
<td>13.13</td>
<td>6.78</td>
</tr>
<tr>
<td>100</td>
<td>0.5</td>
<td>100.00</td>
<td>100.00</td>
<td>91.45</td>
<td>30.62</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>100.00</td>
<td>99.14</td>
<td>78.74</td>
<td>21.43</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>99.94</td>
<td>92.99</td>
<td>58.45</td>
<td>15.88</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>92.59</td>
<td>71.45</td>
<td>36.36</td>
<td>9.55</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>42.95</td>
<td>31.62</td>
<td>16.71</td>
<td>6.83</td>
</tr>
</tbody>
</table>

$P_1$: estimated power of the test in the case without censoring
$P_2$: estimated power of the test in the case with 15% censoring
$P_3$: estimated power of the test in the case with 30% censoring
$P_4$: estimated power of the test in the case with 50% censoring
7.4 Examples

7.4.1 Infant nutrition data

The infant nutrition data are used to illustrate the application of these diagnostic methods. On the basis of the null maximum likelihood estimates of Weibull parameters obtained in Chapter 1, the statistic $M^*$ is evaluated as $M^* = 0.0707$. Using the approximation in (7.5), we have that the critical value for a significance level $P = 0.05$ is 0.043. Thus the result does not give strong evidence against the null hypothesis on the basis of this test.

From Table 1.1, we may see that the earliest age of introduction of fish and egg is 4 months, and there is an observation whose components form the component-wise minima. Hence, the diagnostic based on the component-wise minima indicates the presence of frailty with a significance probability of $\frac{1}{55}$. However, if we have a further look at the data, we find that actually the vast majority of the age of introduction have been given to the nearest month. This probably has little effect on the parameter estimates. But, if the data had greater precision, the event $A$ might not have occurred. For example, the pairs of data $(4, 4)$ and $(4, 6)$ might have been $(3.8, 4.1)$ and $(3.6, 5.9)$. On the other hand, if the pair $(4, 4)$ was $(3.5, 3.5)$, then the value of $M^*$ would be 0.042, which is significant at $P = 0.05$.

For this data set $U = 0.4721$. For significance level $P = 0.05$ and sample size $n = 55$, the critical value evaluated from (7.16) is 0.5761. Therefore, the result from the concordance $U$ test gives no evidence of the presence of heterogeneity. Again, given the coarse rounding of the data, the result should be treated with caution.

Therefore, this data set is also an example to show that proper use of these diagnostic methods relies on the data having been recorded precisely enough.
7.4.2 Repeated endurance exercise tests data

The association between the exercise times 1 hour and 3 hours after drug treatment are of interest here. First, under the null hypothesis, the covariate, dose, is included into the Weibull models. The observed value of $S_j^*$ is obtained as $S_j^* = \hat{\xi}_{i1} y_{i1}^{\hat{\beta}_1} + \hat{\xi}_{i2} y_{i2}^{\hat{\beta}_2}$, where $\log \xi_{ij} = \hat{\beta}_{i0} + \hat{\beta}_{i1}\text{dose}(i)$, and $\hat{\beta}_{ij}, \hat{\phi}_j$ are the null maximum likelihood estimates. Hence, $M^* = \min_j S_j^* = 0.1541$. From the approximation (7.5), the critical value for $n = 21$ and $P = 0.05$ is 0.069. This test gives no evidence against the null model.

The minima of exercise times 1 hour and 3 hours after treatment occur in different observations. Thus there is no evidence for the presence of heterogeneity on the basis of the component-wise minima. However, the estimated value of $U$ is 0.7604. Since 30% of 21 observations are censored in this data set, the mean and variance of the $U$ under null hypothesis is 0.4741 and 0.006845, respectively, see Table 7.3. Thus, the corresponding critical value is 0.6098, while the critical value for no censoring is 0.6293. Therefore, at the significance level $P = 0.05$, the concordance test based on $U$ gives evidence of heterogeneity.

7.4.3 Fibre failure strength data

To illustrate the methods, we focus on the bivariate fibre failure strength data with fibre section of length 5mm and 12mm. The observed value of $M^*$ is obtained based on the null maximum likelihood estimates given in Chapter 1, which is $M^* = 0.1927$. The critical value at $P = 0.05$ is approximately 0.049. Hence, the result from the test based on the $M^*$ test suggests that there is no strong evidence to reject null model.

However, the component-wise minimum of the sample exists in the 28th observation in this data set, which is (2.65, 2.01). The null probability of this event is only 0.024. From this point of view, there is evidence of the presence of heterogeneity.
The observed concordance statistic from the data is $U = 0.7619$, while the critical value at $P = 0.05$ and $n = 42$ evaluated from (7.16) is 0.5879. Hence, the concordance test based on $U$ gives strong evidence for the presence of heterogeneity.

### 7.5 Discussion

Three simple methods for detecting the presence of heterogeneity are proposed in the context of the bivariate PSW model. The component-wise minima diagnostic and the statistic $U$ can be obtained easily without estimating any parameters. Extensive tables of critical values are not required for these approaches. The methods may also be used in the presence of right censoring.

The concordance test based on the statistic $U$ is only applicable to the bivariate PSW model in particular. For the more general multivariate situation with $p > 2$, the statistic $M$ and $M^*$ could easily be modified to include $p$ cumulative hazard terms in place of the two terms needed in the bivariate case. The null distribution of $M$ would then be gamma with shape parameter $p$ and scale parameter 1. Once again, nothing more computationally complicated than maximum likelihood estimation of univariate Weibull distributions would be needed. For the component-wise minima diagnostic, the presence of an observation consisting of all $p$ component-wise minima would supply strong evidence of the presence of heterogeneity (the significance probability would be $n^{-(p-1)}$) when $n$ is not small. But it would be unlikely to occur except in cases of very high dependence. Non-occurrence of such a point would be non-informative.

The statistic $M^*$ may be generalised to cope with covariate information, as in the example given above. The component-wise minima diagnostic could only be modified to cope with covariate information by introducing parameter estimation, in which its immediacy and simplicity would be lost.
Chapter 8

Conclusion and further work

8.1 Conclusion

The focus of this thesis has been the investigation of statistical methods for Weibull based random effects models. The main features of the three Weibull based random effects models, i.e. the gamma mixture of Weibulls (GW) model, the positive stable mixture of Weibulls (PSW) model and the Inverse Gaussian mixture of Weibulls (IGW) model are highlighted. The dependence structures of the three models are explored by employing various association measures. A two stage marginal estimation method is suggested to estimate the parameters of the PSW model. A variety of methods for detecting heterogeneity for the Weibull based random effects models are proposed, and their properties are examined. Conclusions of this work are given here.

The models

The three Weibull based random effects models are generalisations of the Weibull model that accommodate random effects representing heterogeneity in individuals or groups. Because of the closed forms of their survivor functions, it is easy to fit these distributions by maximum likelihood with or without right censored observations. The closed forms of the conditional and marginal distributions are
also convenient for deriving properties.

A justification for the GW model is that the gamma form is flexible in shape, which leads to flexible modelling of survival data. The PSW model is particularly attractive because its marginals are Weibull distributed. The IGW model has a more complex form in general than the GW and PWS models. In addition, the IGW model is a special case of the generalised PSW model.

**Dependence structure**

The dependence structure between pairs of components can be measured in various ways. The product moment correlation coefficient $\rho$ and Kendall’s $\tau$ measure the global association between variables. The explicit relationship between Kendall’s $\tau$ and the parameters $\nu$ (in the GW and the PSW models) and $\theta$ (in the IGW model) shows that $\nu$ or $\theta$ plays an important role as an association parameter in the corresponding Weibull based random effects model.

Local dependence structure may be measured by contour plots, the cross ratio, the time-dependent correlation coefficient, the conditional expected residual life and by conditional probability. Each local measure describes the changes of strength of association with regards to the changes of lifetimes, but the choice of an appropriate measure depends on the questions of interest. All these measures confirm that the three Weibull based random effects models have an equi-correlated structure with positive association between components. The generalised form of the PSW model also admits the possibility of negative association.

**Estimation methods**

The two stage marginal estimation method discussed in Chapter 4 provides an easier way to estimate the multivariate ($p \geq 2$) PSW model. Asymptotic results show that, in the bivariate case without censoring, the variances of the estimators from the two stage marginal method are very close to those of the corresponding maximum likelihood estimators. A simulation study for the bivariate case confirms that, for finite sample size, the estimators from the two stage marginal estimation method are highly efficient in the case without censoring. The estimated efficiency
is also good when the censoring is not too heavy (≤ 30% censoring).

Further results from the simulation study indicate that the two stage marginal estimators are still highly efficient when \( p = 3 \) and \( p = 4 \).

**Diagnostic methods**

Large sample properties of the score test for heterogeneity have been investigated in Chapter 5. The score statistic has infinite null variance. However, after an appropriate normalisation, its asymptotic null distribution is normal. In addition, simulation results show that the rate of convergence of the score to its asymptotic distribution is very slow. Therefore, critical values of the score test for small samples in the case with nuisance parameters are estimated using simulation.

In Chapter 6, study of the asymptotic properties of the likelihood ratio (LR) test for heterogeneity indicates that, when sample size \( n \to \infty \), the null distribution of the LR statistic has a \( \chi^2_{(1)} \) distribution with probability 0.5 and is zero with probability 0.5. The simulation study shows that the rate of the convergence to this limit is very reasonable, and much better than the rate of the convergence to the normal of the score statistic for both with and without nuisance parameters cases. For sample sizes \( n \geq 50 \), it is suggested that appropriate \( \chi^2_{(1)} \) values may be used as critical values for the LR test. For sample sizes \( n < 50 \), critical values are estimated by simulation.

The score and LR tests have comparable good power of detecting the presence of heterogeneity in the cases with and without censoring. The calculation of the score test is easier because it only involves maximum likelihood estimation under the null model, whereas the calculation of the LR test involves estimation under both the null and full models. This difference becomes more important when \( p > 2 \). However, the complication of the non-null model estimation for the PSW is eased by the two stage marginal estimation method discussed in Chapter 4.

The other three diagnostic approaches discussed in Chapter 7 provide simpler ways of detecting the presence of heterogeneity in the context of the bivariate PSW model, although these diagnostic methods have less power than the score
and LR tests. The calculation of the statistic $M^*$ only involves maximum likelihood estimation of the null Weibull distributions. Critical values for the test based on $M^*$ can be obtained by a simple approximation. Therefore, no lengthy table critical values is required. The diagnostic based on the component-wise minima may be easily obtained from an immediate eye-balling of the data, and is not affected by simple Type I and Type II censoring. The convenience of this approach is obvious although the diagnostic is uninformative in certain situations. The concordance statistic $U$ may also be calculated easily without estimating any parameters. The asymptotic normality of the $U$ statistic provides a test based on the $U$ with reasonable power of detecting heterogeneity.

8.2 Further work

There are a few directions in which this work could be taken further. Some general statements in this regard are given as follows.

In the context of detecting the presence of heterogeneity, the tests discussed in this thesis may be further developed in some aspects. First, the asymptotic behaviour of the score and LR tests discussed in this work only involve the cases without censoring and with right censoring at fixed points. In practice, lifetime data might be censored in more complicated forms. For example, observation might cease after a pre-specified number of failures, which is the case of Type II censoring. Therefore, the asymptotic properties of the score and LR tests in the case of more complex censoring mechanisms need to be explored.

Secondly, the asymptotic behaviour of the score and LR statistic in the case with nuisance parameters and with covariate information has been explored only by considering very simple situations in the simulation study. More intensive investigation of the score and LR statistics when in the presence of covariate information using asymptotic theory and simulation may be informative.

Finally, although the general forms of the score and LR tests in the multivariate
(\(p > 2\)) case are given in Chapter 5 and 6, and null properties of the score test in the tri-variate case are also discussed, the general behaviour of the score and LR statistics are not explored in the thesis. Therefore, further work to investigate the null and non-null properties of the score and LR tests in the dimensional (\(p > 2\)) situations may be worthwhile.

The three simple diagnostic approaches have been discussed in Chapter 7 without considering covariate information. Therefore, another area of interest is to investigate the null and non-null properties of the tests based on \(M\), \(M^*\) and \(U\) statistics. Similar to the study of the score and LR tests, further investigation of the diagnostic methods based on \(M^*\) and \(U\) under more complex censoring mechanism may be pursued. In Chapter 7, the asymptotic normality of the \(U\) statistic is obtained. However, the convergence to the normal is not confirmed by simulation. Hence, a simulation study to examine the rate of convergence of the \(U\) statistic in the cases with and without censoring may be useful though one would anticipate that the convergence rate would be good.

Since the generalised PSW model allows both positive and negative association between the lifetimes, it is a more flexible model to accommodate heterogeneity between individuals or groups. However, the involvement of the parameter \(\kappa\) makes the estimation and inference based on the model more complicated. It is worth pursuing further work in this direction. For example, heterogeneity tests based on this model should be explored.
References


Genest, C. and Mackay, R.J. (1986), Archimedean copulas and families of bi-dimensional laws for which the marginals are given, Canadian Journal of Statistics, 14, 2, 145-159


Hougaard, P. (1986a), Survival models for heterogeneous populations derived from stable distributions, Biometrika, 73, 387-396.

Hougaard, P. (1986b), A class of multivariate failure time distributions, Biometrika, 73, 671-678.


Kendall, M.G. (1938), A new measure of rank correlation, Biometrika, 30, 81-93.

Kimber, A.C. (1985), Tests for the exponential, Weibull and Gumbel distributions based on the stabilised probability plot, Biometrika, 72, 661-663.


Appendix A: Program

Simulation on the estimation of parameters of the PSW model

```
C******************************************************
PROGRAM MLECOMP
C--------------------------------------------------------
C PROGRAM GENERATES RANDOM VARIABLES (x1, x2), i=1,...N FROM
C A POSITIVE STABLE MIXTURE OF WEIBULL DISTRIBUTION AND
C COMPARE THE ESTIMATION FROM TWO STAGE MARGINAL METHOD WITH MLE
C--------------------------------------------------------
C MAIN PROGRAM
IMPLICIT NONE
C m IS THE SIMULATION TIMES
C nn IS THE SAMPLE SIZE OF THE VARIABLES
DOUBLE PRECISION y1(10000), y2(10000), alpha(10000)
double precision x1(10000), x2(10000)
double precision xi, phi, xi2, phi2, nu, a, b, effi
double precision alpahaj, betaaj, alphaj2, beta2j, nuj
double precision alpahm, betalm, alpham2, beta2m, num
double precision est(5,2), mean(5,2), sd(5,2), ss1(5,2), ss2(5,2)
INTEGER i, j, k
C SET UP INITIAL VALUES
C xi=alpha^(1/nu), phi=beta/nu
nn=100
m=1000
nu=0.5
xi1=1.0d0
phi1=1.0/nu
xi2=1.0d0
phi2=1.0/nu
DO 10, i=1, 5
DO 20, j=1, 2
```

175
ss1(i,j)=0
ss2(i,j)=0
20 continue
10 continue

C THE SIMULATION START HERE
CALL G05CCF
do 30 k=1, m
nmodel=1
call weibull
nmodel=2
call weibull

C GENERATE POSITIVE STABLE SAMPLE{alpha(i)}
call psi(nu)
C GENERATE WEIBULL-POSITIVE STABLE MIXTURE SAMPLE(x1i, x2i)
do 40, i=1, nn
x1(i)=y1(i)/alpha(i)**(1/phi1)
x2(i)=y2(i)/alpha(i)**(1/phi2)
40 continue

C MAXIMUM LIKELIHOOD ESTIMATES VIA JOINT METHOD
nmodel=4
call mlepswj(alpha1j, beta1j, alpha2j, beta2j, nuj)
est(1,1)=alpha1j
est(2,1)=beta1j
est(3,1)=alpha2j
est(4,1)=beta2j
est(5,1)=nuj

C STEP 1 OF MLE ON MARGINAL METHOD
C MLE OF MARGINAL WEIBULL PARAMETERS
nmodel=1
call mleweibull(a, b)
alphalm=a
betalm=b
est(1,2)=alphalm
est(2,2)=betalm

C MLE OF MARGINAL WEIBULL PARAMETERS
nmodel=2
call mleweibull(a, b)
alphabetam=a
betam=b
est(3,2)=alphabetam
est(4,2)=betam

C STEP 2 OF MLE ON MARGINAL METHOD
C MLE ON ASSOCIATION PARAMETER NU BY USING MARGINAL EST OF WEIBULL

176
nmodel=3
call mlepswm(num)
est(5,2)=num

C FOR THE CALCULATION OF MEAN AND S.D.
do 50 i=1,5
   do 60 j=1,2
      ssl(i,j)=ssl(i,j)+est(i,j)
      ss2(i,j)=ss2(i,j)+est(i,j)*est(i,j)
   60 continue
50 continue

C END UP THE SIMULATION
30 continue

C CALCULATION OF MEAN AND S.D.
do 70 i=1,5
   do 80 j=1,2
      mean(i,j)=ssl(i,j)/dble(m)
      sd(i,j)=ss2(i,j)/dble(m)-mean(i,j)*mean(i,j)
   80 continue
70 continue
write(*, 100) mean(i, l), sd(i, l), mean(i, 2), sd(i, 2), effi
100 format( lx, 2(fl0.5, "("), fl0.5")"), fl0.5)
end

C******************************************************************************
C******************************************************************************
C******************************************************************************
subroutine weibull
C******************************************************************************
This subroutine generates weibull random variables with shape and
scale parameters xi and phi, and sample size nn
C******************************************************************************
implicit none
double precision yl(10000), y2(10000), alpha(10000)
double precision x1(10000), x2(10000)
double precision xi1, phi1, xi2, phi2
double precision G05DPF, aw1, aw2, bw1, bw2
integer i, ifail, nn, nmodel
common nn/observed/yl, y2, x1, x2, alpha/model/nmodel
common/parameter1/xi1, phi1, xi2, phi2
ifail=0
aw1=phi1
bw1=1/xi1
aw2=phi2
bw2=1/xi2
if (nmodel.eq.1) then
   do 10 i=1,nn


```fortran
C subroutine psi(nu)
C This function computes the random variable alpha
C which comes from a positive stable law
C
C double precision nu, v, w, a, alpha(10000), yl(10000), y2(10000)
C double precision x1(10000), x2(10000)
C integer i, nn
C common nn/observed/yl, y2, xl, x2, alpha
C external functions
C double precision g05daf, g05dbf
C do 10 i=1, nn
C v = g05daf(0.0d+00, 1.0d+00)
C v = v*3.141592D+00
C w = g05dbf(1.0d+00)
C a = (1.0d+00 - nu)/nu
C alpha(i) = (dsin(nu*v)/((dsin(v))**(1.0d+00/nu)))*
C + (sin((1.0d+00 - nu)*v))/w)**(a)
10 continue
C
C subroutine mlepswj(xisl, phisl, xis2, phis2, mlenu)
C maximum likelihood estimation for the weibull mixture sample
C
C double precision yl(10000), y2(10000), alpha(10000)
C double precision xisl, phisl, xis2, phis2, mlenu
C double precision x1(10000), x2(10000)
C integer nn
C common nn/observed/yl, y2, x1, x2, alpha/model/nmodel
C
C NAG VARIABLES
C double precision bl(5), bu(5), xx(5), w(70), f
C integer iw(7), ifail, ibound, n, liw, lw
```
C SET UP NAG VARIABLES IN E04JAF

n=5
ibound=0
bl(1)=1.0d-6
bl(2)=1.0d-6
bl(3)=1.0d-6
bl(4)=1.0d-6
bl(5)=1.0d-6
bu(1)=1.0d6
bu(2)=1.0d6
bu(3)=1.0d6
bu(4)=1.0d6
bu(5)=1.0d0
xx(1)=1.0d0
xx(2)=1.0d0
xx(3)=1.0d0
xx(4)=1.0d0
xx(5)=0.5d0
liw=7
lw=70
ifail=-1

C CALL THE SUBROUTINE E04JAF

call E04JAF(n, ibound, bl, bu, xx, f, liw, lw, ifail)
xisl=xx(1)
phisl=xx(2)
xis2=xx(3)
phis2=xx(4)
mnenu=xx(5)
return
end

C******************************************************************************

C******************************************************************************

subroutine mleweibull(a, b)
  C miximum likelihood estimation for the
  C weibull marginal distributed sample

  double precision y1(10000), y2(10000), alpha(10000)
  double precision x1(10000), x2(10000)
  double precision a, b
  integer nn
  common nn/observed/y1, y2, x1, x2, alpha/model/nmodel

C NAG VARIABLES
  double precision bl(2), bu(2), xx(2), w(25), f
  integer           liw(4), ifail, ibound, n, liw, lw

C SET UP NAG VARIABLES IN E04JAF
n=2
ibound=0
bl(1)=1.0d-6
bl(2)=1.0d-6
bu(1)=1.0d6
bu(2)=1.0d6
xx(1)=1.0d0
xx(2)=1.0d0
liw=4
lw=25
ifail=-1
C CALL THE SUBROUTINE E04JAF
call E04JAF(n, ibound, bl, bu, xx, f, iw, liw, lw, ifail)
a=xx(1)
b=xx(2)
return
end
C *******************************************************************
C ******************************************************************************
C ****************************************************************************
C subroutine mlepswm(num)
C -----------------------------------------------------------------
C maximum likelihood estimation for the weibull distributed sample
C ------------------------------------------------------------------
C double precision yl(10000), y2(10000), alpha(10000)
C double precision alphalm, betalm, alpha2m, beta2m, num
C double precision x1(10000), x2(10000)
C integer nn, count(4)
C common nn/observed/yl, y2, x1, x2, alpha/model/nmodel
C common /parameter2/alphalm, betalm, alpha2m, beta2m
C common /indicator/count
C NAG VARIABLES
C double precision bl(1), bu(1), xx(1), w(13), f
C integer iw(3), ifail, ibound, n, liw, lw
C SET UP NAG VARIABLES IN E04JAF
n=1
ibound=0
bl(1)=1.0d-5
bu(1)=1.0d0
xx(1)=0.5d0
liw=3
lw=13
ifail=-1
C CALL THE SUBROUTINE E04JAF
call E04JAF(n, ibound, bl, bu, xx, f, iw, liw, lw, ifail)
num=xx(1)
return
end

******************************************************************************

******************************************************************************

subroutine funct1(n,xc,fc)
******************************************************************************

This subroutine calculates the negative log likelihood for a set of
nn iid sample under various conditions

The observed values are passed via the common block
******************************************************************************

implicit none

integer nn,n,i,nmodel,count(4)
DOUBLE PRECISION yl(10000),y2(10000),alpha(10000)
double precision xl(10000),x2(10000)
double precision alphalm,betalm,alpha2m,beta2m
common nn/observed/yl,y2,x1,x2,alpha/model/nmodel
common /parameter2/alphalm,betalm,alpha2m,beta2m
double precision xc(n),fc,g,s,ss
g=0.0d0

** keeps track of the neg log likelihood as i loops over the data
if (nmodel.eq.1) then
  g=-nn*(log(xc(1))+log(xc(2))
do 10, i=1, nn
  g=g+xc(1)*xl(i)**xc(2)-(xc(2)-1)*log(xl(i))
10  continue

elseif (nmodel.eq.2) then
  g=-nn*(log(xc(1))+log(xc(2))
do 20, i=1, nn
  g=g+xc(1)*x2(i)**xc(2)-(xc(2)-1)*log(x2(i))
20  continue

elseif (nmodel.eq.3) then
  g=-nn*(1/xc(1)**log(alphalm)+log(alpha2m)+log(betalm)+log(beta2m)
* -2*log(xc(1))
do 30 i=1, nn
  s=alphalm**(1/xc(1))*x1(i)**(betalm/xc(1))
* +alpha2m**(1/xc(1))*x2(i)**(beta2m/xc(1))
  ss=xc(1)**s**2*xc(1)-xc(1)**s*ss(1)**ss(1)-ss(1)**ss(1)
  g=g+s**xc(1)-(betalm/xc(1)-1)*log(x1(i))
* -(beta2m/xc(1)-1)*log(x2(i))-log(ss)
30  continue

elseif (nmodel.eq.4) then
  g=-nn*(1/xc(5)**log(xc(1))+log(xc(3))+log(xc(2))+log(xc(4))
* -2*log(xc(5))
do 40 i=1, nn
  s=(xc(1)**(1/xc(5))*x1(i)**(xc(2)/xc(5))+xc(3)**(1/xc(5))*x2(i)
* **(xc(4)/xc(5))

181
ss=xc(5)*xc(5)*s**((2*xc(5)-2)-xc(5)-(xc(5)-1)*s**((xc(5)-2)
g=g+s**xc(5)-(xc(2)/xc(5)-1)*log(xi(i))-(xc(4)/xc(5)-1)*log(x2(i))
* -log(ss)
40 continue
endif
fc=g
return
end
C
************************************************************************************************************

Simulation on the score test

C
************************************************************************************************************
 PROGRAM SCOREM
C
-----------------------------------------------------------------------------------
C PROGRAM GENERATES RANDOM VARIABLES (y1, y2), i=1,...N FROM
C TWO INDEPENDENT WEIBULL DISTRIBUTION, with xi=1 and phi=1
C UNDER HO: nu=1, CALCULATES SCORE STATISTIC IN THE CASE WITHOUT
C CENSORING AND PARAMETERS ARE UNKNOW(ESTIMATED FROM ML)
C -----------------------------------------------------------------------------------
C MAIN PROGRAM
IMPLICIT NONE
Cn IS THE SAMPLE SIZE
Cm IS THE SIMULATION TIMES
double precision yl(100000), y2(100000), ss, score(100000)
double precision xil, phil, xi2, phi2, qz(9)
common/observed/yl, y2, nn/model/nmodel
INTEGER nn, m, i, j, nmodel
nn=100
m=10000
write (*,*) 'observation number = ', nn
write (*,*) 'simulation times = ', m
CALL G05CCF
do 10 i=1, m
C GENERATE WEIBULL SAMPLE AND GENERATE CENSORED DATA IN THE SAMPLE
nmodel=1
call weibull
C MAXIMUM LIKELIHOOD ESTIMATE OF THE PARAMETERS
call mleweibull(xil, phil)
C GENERATE WEIBULL SAMPLE AND GENERATE CENSORED DATA IN THE SAMPLE
nmodel=2
call weibull
C MAXIMUM LIKELIHOOD ESTIMATE OF THE PARAMETERS
call mleweibull(xi2, phi2)
C CALCULATE THE SCORE STATISTIC

182
score(i)=0  
do 30 j=1,nn  
ss=xil*y1(j)**phi1+xi2*y2(j)**phi2  
score(i)=score(i)+(2*(l+log(ss))-l/ss-ss*log(ss))  
30 continue  
10 continue  
call sort(score,m)  
call quantile(score,m,qz)  
do 50 i=1,9  
qz(i)=qz(i)/sqrt(0.5*double(nn)*log(double(nn)))  
50 continue  
write (*,200) nn, (qz(i), i=1,9)  
200 format(I5,ix,9(F5.2,ix))  
end

C     ************************************************************
C     ************************************************************
C       subroutine weibull
C     ************************************************************
C       This subroutine generates weibull random variables with shape and
C       scale parameters 1.0, and sample size nn
C     ************************************************************
C       implicit none
C       double precision yl(100000), y2(100000), G05DPF, a, b
C       common/observed/yl, y2, nn/model/nmodel
C       integer i, ifail, nn, nmodel
C       ifail=0
C       a=1.0d0
C       b=1.0d0
C       if (nmodel.eq.1) then
C       do 10 i=1,nn
C       y1(i)=G05DPF(a,b,ifail)
C       10 continue
C       elseif (nmodel.eq.2) then
C       do 20 i=1,nn
C       y2(i)=G05DPF(a,b,ifail)
C       20 continue
C       endif
C       return
C       end

C     ************************************************************
C     ************************************************************
C       subroutine mleweibull(xi, phi)
C     ************************************************************
C       maximum likelihood estimation for the weibull distributed sample
double precision yl(100000), y2(100000), xi, phi
integer nmodel, nn
common/observed/yl, y2, nn/model/nmodel

NAG VARIABLES
double precision bl(2), bu(2), xx(2), w(25), f
integer iw(4), ifail, ibound, n, liw, lw

SET UP NAG VARIABLES IN E04JAF
n=2
ibound=0
bl(1)=0.0d0
bl(2)=0.0d0
bu(1)=1.0d6
bu(2)=1.0d6
xx(1)=0.5d0
xx(2)=0.5d0
liw=4
lw=25
ifail=-1
CALL THE SUBROUTINE E04JAF
call E04JAF(n, ibound, bl, bu, xx, f, iw, liw, lw, ifail)
xi=xx(1)
phi=xx(2)
return
end

*****************************************************************

subroutine functl(n, xc, fc)
*****************************************************************
This subroutine calculates the negative log likelihood for a set of
n iid points from a weibull distribution
The observed values are passed via the common block
*****************************************************************
implicit none
integer nn, n, i, nmodel
double precision xc(n), fc, g, yl(100000), y2(100000)
common/observed/y1, y2, nn/model/nmodel

g=0.0d0

g keeps track of the neg log likelihood as i loops over the data
if (nmodel.eq.1) then
  g=0.0d0+0
  do 10, i=1, nn
     g=g+xc(1)*yl(i)**xc(2)-((xc(2)-1)*log(yl(i))+log(xc(1))+log(xc(2)))
  10 continue
elseif (nmodel.eq.2) then
  g=0.0d0+0
  do 20, i=1, nn

g=g+xc(1)*y2(i)**xc(2)-((xc(2)-1)*log(y2(i))+log(xc(1))+log(xc(2)))
continue
endif
fc=g
return
end
C
******************************************************************************
C
******************************************************************************
SUBROUTINE SORT(Z,M)
C THIS ROUTINE SORT THE DATA
C
******************************************************************************
IMPLICIT NONE
double precision z(10000)
integer m,ifail
ifail=0
call M01CAF(z,1,m,'ASCENDING',ifail)
return
end
C
******************************************************************************
C
******************************************************************************
SUBROUTINE QUANTILE(Z,M,QZ)
C THIS ROUTINE CALCULATE QUANTINE
C
******************************************************************************
IMPLICIT NONE
double precision z(10000),qz(9)
real q(9)
real index(9)
integer i,k,m
data q(1),q(2),q(3),q(4),q(5),q(6),q(7),q(8),q(9)
* /0.01,0.025,0.05,0.10,0.50,0.90,0.95,0.975,0.99/
do 20, i=1, 9
index(i)=q(i)*m
k=int(index(i))
if (k.eq.index(i)) then
qz(i)=z(k)
else
qz(i)=(z(k)+z(k+1))/2
endif
20 continue
return
end
C
******************************************************************************
Appendix B: List of Publications

Published papers


Conference presentation
