# Location-Scale Bivariate Weibull Distributions For Bivariate Lifetime Modeling

Except where reference is made to the work of others, the work described in this thesis is my own or was done in collaboration with my advisory committee. This thesis does not include proprietary or classified information.

Yi Han

Certificate of Approval:

Asheber Abebe
Assistant Professor
Mathematics and Statistics

Mark Carpenter, Chair
Associate Professor
Mathematics and Statistics

Nedret Billor
Associate Professor
Mathematics and Statistics

Stephen L. McFarland
Acting Dean
Graduate School

# Location-Scale Bivariate Weibull Distributions For Bivariate Lifetime Modeling

Yi Han

A Thesis

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Master of Science

Auburn, Alabama
May 11, 2006

Location-Scale Bivariate Weibull Distributions For Bivariate Lifetime Modeling

Yi Han



Signature of Author

Date of Graduation

# Vita

Yi Han, son of Mulan Zhu and Yujian Han, was born in Qing Chuan, Si Chuan Province, China on twenty eighth of November, 1973. He completed his Bachelors in Engineering Economics and Masters in Management Engineering from Central-South University of Technology, Changsha, China, in July, 1996, and May, 1999, respectively. After graduation, he worked as a Financial Analyst in China Copper Lead Zinc Co. Ltd., Beijing, China. In May, 2003, he earned his Masters in Industrial and Systems Engineering from Auburn University, Alabama, U.S.A.

## Thesis Abstract

## Location-Scale Bivariate Weibull Distributions For Bivariate Lifetime Modeling

Yi Han

Master of Science, May 11, 2006
(M.I.S.E., Auburn University, 2003)
(M.M., Central South University of Technology, 1999)
(B.E., Central South University of Technology, 1996)

114 Typed Pages

Directed by Mark Carpenter

Much research has been conducted over the last thirty years in the development and characterization of bivariate survival distributions. Typically, the multivariate distribution is derived assuming that the marginal distributions are of some specified lifetime family. In this thesis, we examine various bivariate Weibull models. In addition, a location-scale bivariate Weibull model is proposed. Bivariate parameter estimation, with and without censoring, is developed and applied to real and simulated data. Examples are drawn from biomedical research.

## Acknowledgments

I wish to express sincere appreciation to my academic advisor, Dr. Mark Carpenter for his direction, assistance, advice and patience during my studies at Auburn University. My appreciation also goes to Dr. Asheber Abebe and Dr. Nedret Billor for serving on my advisory committee and their constructive comments. I appreciate the great help I received from Mr. Norou Diawara. I would like to extend my special thanks and great appreciation to my family and my fiance for their love, support, and constant encouragement.

In loving memory of my father.

Style manual or journal used Journal of Approximation Theory (together with the style known as "aums"). Bibliograpy follows van Leunen's *A Handbook for Scholars.*

Computer software used The document preparation package $\TeX$ (specifically $\LaTeX$) together with the departmental style-file `aums.sty`.

## Table of Contents

## List of Tables

CHAPTER 1

INTRODUCTION

## 1.1 Univariate Lifetime Distributions

The term lifetime generally refers to the time to some events such as death or failure from a certain starting point. We define lifetime or survival analysis as the collection of statistical models and methodologies to analyze lifetime data of various types. In applications in engineering and biomedical sciences, failure time and survival time are often used synonymously as lifetime, thus we have terms failure time distributions and survival time distributions respectively. Lifetime distributions generally have positive support, i.e., lifetime data can take on only non-negative real values.

For a nonnegative continuous random variable $X$, the cumulative distribution function (cdf) $F_X(\cdot)$ and survivor distribution $S_X(\cdot)$ are defined as

$$F_X(x) = P(X \leq x) = \int_0^x f(t)\,dt, \tag{1.1}$$

$$S_X(x) = P(X \geq x) = \int_x^\infty f(t)\,dt. \tag{1.2}$$

and the hazard function is given by

$$h(x) = -\frac{\lim_{\triangle x \to 0} \left[\frac{S(x+\triangle x)-S(x)}{S(x)}\right]}{S(x)} = \frac{f(x)}{S(x)}. \tag{1.3}$$

The hazard function gives the instantaneous rate of failure at time $x$, given that the individual survives up to $x$, and carries important information concerning the risk of failure versus time. It is often desirable to model lifetime distributions through the hazard function if factors affecting an individual's lifetime are time-dependent, or vary over time.

### 1.1.1 Univariate Weibull Distribution

In his 1951 paper, "A Statistical Distribution Function of Wide Applicability", the Swedish Professor Waloddi Weibull introduced the Weibull Distribution and stated "Experience has shown that, in many cases, it fits the observations better than other known distribution functions". Eventually, the Weibull Distribution became the most useful tool in reliability due to its unique characteristics and wide range of applicability, especially so when it pertains to describing the underlying distribution of time to failure (TTF) of mechanical or electrical components or systems.

Professor Weibull defined his original cumulative distribution function as

$$F(x) = 1 - \exp\left[-\frac{(x - x_\mu)^m}{x_0^m}\right], \tag{1.4}$$

where $x_u$, $m$, and $x_0$ correspond to the more universal notations used herein of the location parameter $\delta$, the shape parameter $\beta$, and the scale parameter $\theta$ as in (1.5).

We define the three-parameter Weibull probability density function (pdf) as

$$f(x) = \frac{\beta}{\theta}\left(\frac{x-\delta}{\theta}\right)^{\beta-1} \exp\left[-\left(\frac{x-\delta}{\theta}\right)^{\beta}\right], x \geq \delta \geq 0,\ \theta,\ \beta > 0. \tag{1.5}$$

The two-parameter Weibull probability density function (pdf) is given as

$$f(x) = \frac{\beta}{\theta}\left(\frac{x}{\theta}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\theta}\right)^{\beta}\right], x \geq 0,\ \theta,\ \beta > 0. \tag{1.6}$$

Note that (1.6) is a special case of (1.5) where $\delta = 0$. When $\beta = 1$, the Weibull pdf becomes an exponential pdf.

The corresponding cdfs of the two and three-parameter Weibull distribution are given as

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\theta}\right)^{\beta}\right], \tag{1.7}$$

and

$$F(x) = 1 - \exp\left[-\left(\frac{x-\delta}{\theta}\right)^{\beta}\right]. \tag{1.8}$$

Note that all Weibull cdf's cross at one point where the cdf is valued at approximately 0.63, and the corresponding $x$ value is 1, which is the value of the scale parameter $\theta$. We define such a value of $x$ as the characteristic life of the Weibull distribution,which is the time at which the value of the cdf is exactly equal to $1-e^{-1}$. In other words, 63.212% of the population fails by the time of characteristic life no matter what the values of the other parameters are. For two parameter Weibull distribution, the characteristic life is equal to $\theta$, and $\theta + \delta$ for the three parameter Weibull.

The survival functions, the probability that an individual survives at least time $x$, of the two and three-parameter Weibull distribution are given as

$$S(x) = \exp\left[-\left(\frac{x}{\theta}\right)^{\beta}\right], \tag{1.9}$$

and

$$S(x) = \exp\left[-\left(\frac{x-\delta}{\theta}\right)^{\beta}\right]. \tag{1.10}$$

The hazard function of the three-parameter Weibull distribution is given as

$$h(x) = \frac{f(x)}{S(x)} = \frac{\beta}{\theta}\left(\frac{x-\delta}{\theta}\right)^{\beta-1}; \tag{1.11}$$

and the corresponding cumulative hazard function is given as

$$H(x) = \int_{0}^{x} h(t)dt = \left(\frac{x-\delta}{\theta}\right)^{\beta} = -\ln\left[S(x)\right] \tag{1.12}$$

The Weibull distribution can have increasing, decreasing and constant hazard rates, which in term reflects the versatileness of the Weibull distribution in lifetime or survival analysis.

The above (1.5), (1.9) and (1.12) show the relationships among the pdf, the survivor function, the hazard function and the cumulative hazard function of the Weibull distribution, i.e., given any one of them, the others follow.

### 1.1.2 Distributional Properties for Univariate Weibull Distribution

**The Moments of Weibull Distribution**

The general noncentral moments of the two parameter Weibull Distribution is given by

$$E\left(X^n\right) = \theta^n \Gamma\left(1 + \frac{n}{\beta}\right), \tag{1.13}$$

for any integer n, and $\Gamma\left(cdot\right)$ is a gamma function defined as

$$\Gamma\left(k\right) = \int_0^\infty x^{k-1} \exp\left(-x\right) dx,\ \ k > 0.$$

The general moments of the three parameter (location-scale) Weibull distribution is more complicated, but can be derived from that of the two parameter Weibull. Let the continuous positive random variable $X$ follow a two parameter Weibull distribution as defined in (1.6), then $Y = X + \delta$ follows the three parameter Weibull distribution as in (1.5). The general moment is thus given by

$$E\left(Y^n\right) = E\left[\left(X + \delta\right)^n\right]. \tag{1.14}$$

By applying the binomial theorem that for positive integers n,

$$(x+a)^n = \sum_{k=0}^{n} \binom{n}{k} x^k a^{n-k}, \tag{1.15}$$

(1.14) becomes

$$\begin{aligned} E\left(Y^n\right) &= E\left[(X+\delta)^n\right] = E\left[\sum_{k=0}^{n} \binom{n}{k} X^k \delta^{n-k}\right] \\ &= \sum_{k=0}^{n} E\left[\binom{n}{k} X^k \delta^{n-k}\right] = \sum_{k=0}^{n} \left[\binom{n}{k} \delta^{n-k} E\left(X^k\right)\right] \\ &= \sum_{k=0}^{n} \left[\binom{n}{k} \theta^k \delta^{n-k} \Gamma\left(1+\frac{k}{\beta}\right)\right] \end{aligned} \tag{1.16}$$

**The Mean, Median, and Variance of Weibull Distribution**

The mean and the variance of the Weibull distribution can be derived from the general moment. The means of the two and three parameter Weibull are given by

$$\begin{aligned} E\left(X\right) &= \theta\Gamma\left(1+\frac{1}{\beta}\right), \ and \\ E\left(Y\right) &= \delta+\theta\Gamma\left(1+\frac{1}{\beta}\right). \end{aligned} \tag{1.17}$$

The variance of the two parameter Weibull distribution is given by

$$\begin{aligned} Var(X) &= E\left(X^2\right) - E^2(X) \\ &= \theta^2\Gamma\left(1+\frac{2}{\beta}\right) - \theta^2\Gamma^2\left(1+\frac{1}{\beta}\right) \\ &= \theta^2\left[\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma^2\left(1+\frac{1}{\beta}\right)\right], \end{aligned} \tag{1.18}$$

and the variance of the three parameter Weibull distribution is given by

$$\begin{aligned} Var(X) &= E\left(X^2\right) - E^2(X) \\ &= \sum_{k=0}^{2}\left[\binom{2}{k}\theta^k\delta^{2-k}\Gamma\left(1+\frac{k}{\beta}\right)\right] - \left[\theta\Gamma\left(1+\frac{1}{\beta}\right)+\delta\right]^2 \\ &= \left[\delta^2 + 2\theta\delta\Gamma\left(1+\frac{1}{\beta}\right) + \theta^2\Gamma\left(1+\frac{2}{\beta}\right)\right] \\ &\quad - \left[\theta^2\Gamma^2\left(1+\frac{1}{\beta}\right) + 2\theta\delta\Gamma\left(1+\frac{1}{\beta}\right) + \delta^2\right] \\ &= \theta^2\left[\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma^2\left(1+\frac{1}{\beta}\right)\right], \end{aligned} \tag{1.19}$$

which is the same as the variance of the two parameter Weibull distribution.

The medians of the two and three parameter Weibull can be solved from the cdfs or the survival functions, and are given as

$$\begin{aligned} M(X) &= \theta\left(-\ln 0.5\right)^{1/\beta},\ and \\ M(X) &= \theta\left(-\ln 0.5\right)^{1/\beta} + \delta. \end{aligned} \tag{1.20}$$

### 1.1.3 Univariate Weibull Distribution Parameter Estimation

There are several methods for estimating the parameters of the Weibull distribution, i.e., probability plotting, hazard plotting, and maximum likelihood. The method of maximum likelihood (ML) is a commonly used procedure because it has very desirable properties that when the sample size n is large. Under certain regularity conditions, the maximum likelihood estimator of any parameter is almost unbiased and has a variance that is nearly as small as can be achieved by any estimator, and its sampling distribution (or pdf) approaches normality [9, Devore (2000)].

Let $X_1$, $X_2$,..., $X_n$ be a random sample from a two-parameter Weibull distribution, and $x_1$, $x_2$,..., $x_n$ be the corresponding observed values, then the likelihood function ($LF$) is given by

$$L(\theta, \beta) = \prod_{i=1}^{n} \left( \frac{\beta}{\theta^{\beta}} x_i^{\beta-1} e^{-\left(\frac{x_i}{\theta}\right)^{\beta}} dx_i \right) \tag{1.21}$$

Since the logarithm transform is a monotone increasing one, maximizing the natural logarithm of the likelihood function is equivalent to maximizing the likelihood function itself. Taking the natural logarithm of the $LF$, and setting both derivatives to zero yields two sets of score equations that do not give closed-form solution for the maximum likelihood estimates (mle). Instead, for each sample set, the equations can be solved using an iterative numerical procedure which is quite tedious without the aid of computers. In most instances, however, a simple trial and error approach also works [7, Cohen (1965)].

The corresponding loglikelihood function (LLF) for two parameter Weibull distribution is given by

$$l(\theta, \beta) = \sum_{i=1}^{n} \left[ \ln\left(\frac{\beta}{\theta}\right) + (\beta - 1)\ln\left(\frac{x_i}{\theta}\right) - \left(\frac{x_i}{\theta}\right)^{\beta} \right]$$

### 1.1.4 Asymptotic Normality and Confidence Intervals of MLE Weibull Parameters

It is well known that the sampling distributions ($SMD$) of maximum likelihood estimators for Weibull parameters approach normality asymptotically. For example, [35, Miller (1984)] measured the degree of Normality for the MLE of $\beta$ using Chi-square goodness-of-fit. He found that when the sample size is around 170, the MLE of $\beta$ is approximately normally distributed. For small or medium sample sizes, distributions of parameters are clearly skewed. Moreover, [30, Liu (1997)] suggests that for 20 or less observed failures, two-parameter Weibull distribution should be a preferred choice for more stable and more conservative results.

We can construct asymptotic confidence intervals for the Weibull parameters estimated by maximum likelihood method when the sample size is large. Because the characteristic life, $\theta$, and the minimum life, $\delta$, are the-larger-the-better ($LTB$) type of parameters, it is reasonable to construct lower one-sided confidence intervals for $\theta$ and $\delta$, and two-sided confidence interval (CI) for $\beta$.

In order to calculate the asymptotic confidence intervals, we first need to estimate standard errors of the parameters. Information Matrix and Bootstrapping method can be utilized to better estimate standard errors. The conservative Bonferroni confidence interval is also derived to address the problem of correlations between Weibull parameters.

### 1.1.5 Information Matrix and Variance-Covariance Matrix of MLE Weibull Parameters

The information matrix I can be constructed from the logarithm of the likelihood function, where its $ij^{th}$ element is

$$I_{ij} = E\left[-\frac{\partial^2 L(\theta; X)}{\partial\theta_i \partial\theta_j}\right] \tag{1.22}$$

The inverse of the information matrix, $I^{-1}$, is the variance-covariance matrix, where the diagonal elements are variances of parameters and elements elsewhere are covariances. However, applying the expectation operator to the above equations in order to obtain exact results is often too complicated to accomplish, though asymptotic information matrix and variance-covariance matrix can be constructed as the sample size increases. One option is to use simulation, such as parametric Bootstrapping as proposed by [10, Efron (1985)], with the MLEs of $\theta$ and $\beta$ as seeds.

### 1.1.6 Bonferroni Simultaneous Confidence Intervals for the 2-Parameter Weibull Model

It is clearly shown from the information matrix that there are correlations between Weibull parameters. The CIs obtained by Variance-Covariance and Bootstrapping ignore such factors. If the CIs are independent, then the joint confidence coefficient for a joint CI would be the product of all the confidence coefficients of the parameter CIs, i.e.

$$(1-\alpha_{joint}) = \prod_{i=1}^{m}(1-\alpha_i) \tag{1.23}$$

where m is the number of parameters. The intervals, however, are not independent for Weibull parameters. It can be shown that the overall error rate, $\alpha_{joint}$, is no more than the summation of all the individual error rates, or, $\alpha_{joint} \le \sum_{i=1}^{m}\alpha_i$, which implies that when a joint confidence region is to be constructed with overall error rate $\alpha_{joint}$, the individual error rates should be set at around $\frac{\alpha_{joint}}{m}$, or, if different individual error rates are desired, set them such that $\sum_{i=1}^{m}\alpha_i \approx \alpha_{joint}$.

So, in order to obtain a simultaneous rectangular Bonferroni $(1-\alpha_{joint})$CI region for Weibull parameters $\theta$ and $\beta$, we should set the individual confidence coefficients for the CIs of $\theta$ and $\beta$ both at $(1-\frac{\alpha_{joint}}{2})$. Therefore, the Bonferroni $(1-\alpha_{joint})$CI region for $\theta$ and $\beta$ are as follows:

The lower $(1-\frac{\alpha_{joint}}{2})$CI for $\theta$

$$\hat{\theta} - Z_{\frac{\alpha_{joint}}{2}} * se(\hat{\theta}) \tag{1.24}$$

The two-sided $(1-\frac{\alpha_{joint}}{2})$CI for $\beta$

$$\hat{\beta} \pm Z_{\frac{\alpha_{joint}}{4}} * se(\hat{\beta}) \tag{1.25}$$

## 1.2 Multivariate Lifetime Distributions

Literature is abundant on multivariate lifetime data and distributions. [18, Hougaard (2000)] and [37, Murthy, etc. (2004)] provide comprehensive and updated literature reviews. However as mentioned in [26, Lawless (2002)], gaps exist in some areas. A related issue is the introduction of covariates in multivariate survivor analysis, we analyze and investigate this for the bivariate case.

### 1.2.1 Multivariate Distribution Functions

Multivariate lifetime data arise when multiple events occur for each subject in the study. The problem addressed hereby involves continuous nonnegative random variables of lifetime, $X_1$, $X_2, ...,$ $X_n$, with joint probability density function as $f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)$.

A function $f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)$ is a bivariate pdf if

1. $f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) \geq 0 \ \forall \ x_i, \ i = 1, 2, ..., n$;

2. $\underset{\Re^n}{\int\int ... \int} f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n = 1$.

The multivariate distribution and survivor functions are defined as

$$F_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n) = P\left(X_i \leq x_i, \forall x_i, i = 1,2,...,n\right) \tag{1.26}$$

$$S_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n) = P\left(X_i \geq x_i, \forall x_i, i = 1,2,...,n\right) \tag{1.27}$$

and the marginal and joint hazard functions are given by

$$\begin{aligned} \lambda_j\left(x\right) &= \frac{-\partial S_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n)/\partial x_j}{S_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n)} \\ \lambda\left(x_1,x_2,...,x_n\right) &= \frac{f_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n)}{S_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n)} \end{aligned} \tag{1.28}$$

The joint hazard function describes the instantenuous probability that all subjects experience an event given the subjects have survived up to a time $x$.

[23, Joe (1997)] summaries the following properties of a multivariate distribution function.

1. $\lim_{x_j \to \infty} S\left(x_1,x_2,...,x_n\right) = 0 \;\; j = 1,2,...,n;$

2. $\lim_{x_j \to \infty \forall j} F\left(x_1,x_2,...,x_n\right) = 1, \;\; j = 1,2,...,n;$

3. For all $(a_1,...,a_n)$, $(b_1,...,b_n)$ with $a_j < b_j$, $j = 1,2,...,n$,

$$\sum_{i_1=1}^{2} ... \sum_{i_n=1}^{2} (-1)^{i_1+...+i_n} F\left(x_{1i_1},...,x_{ni_n}\right) \geq 0, \qquad x_{j1} = a_j, \;\; x_{j2} = b_j \tag{Rectangle Inequality}$$

If $F$ has $n^{th}$-order derivatives, the above property is equivalent to $\partial^n F/\partial x_1 \partial x_2 ... \partial x_n \geq 0$.

Unlike procedures in the univariate settings, it is difficult to show if a function $F$ is a proper multivariate cdf by the above properties.

### 1.2.2 Dependence Structure and Types

An independence assumption of the bivariate covariates simplifies the question and [26, Lawless (2002)], page 502 refers to this assumption as "working independence". There is a great need, however, in modeling multicomponent systems that are not independent. The independence assumption is impractical in many models such as the time of first and second occurrence of cancer tumors, a breakdown of dual generators, or the survivor times of paired organ system (for example lungs and kidneys in the human body).

Much research has been done in this direction. [32, Marshall and Olkin (1967)] presented a derivation of the multivariate exponential and Weibull distributions with a shock model such that the components in the system have simultaneous failure time with a positive probability. [33, Marshall and Olkin (1988)] presented another method without discussing inference procedures.

[6, Conway (1983)] and [20, Huang and Kotz (1984)] developed the idea of Farlie - Gumbel - Morgenstern (FGM) families of bivariate distributions, but the statistical procedures do not necessarily fit in the statistical estimations. [17, Hougaard (1986)] derived the model for a bivariate Weibull distributions as a mixture. [31, Lu and Battacharrya (1990)] considered modeling the failure behavior of a two component system through the construction of a new unifying bivariate family of lifetime distributions with absolute continuity including positive and negative quadrant dependence, and a bivariate Weibull model is obtained as a special case.

[21, Iyer and Manjunath (2002)] and [22, Iyer and Manjunath (2004)] derived lifetime distributions assuming a linear relationship between the two variables of interest. They presented bivariate distributions that have specified exponential marginal distributions and motivate the linear structured relationship between two variables and $X_2$ in two parts: the measurement model that gives the data $x = (x_1; x_2)$ and the structured equation part that explains the relationship via a latent random variable $Z$ that is independent of $X_1$. The variables are then related as

$$X_2 = aX_1 + Z, \quad a \geq 0$$

Once the covariates $X_1$ and $X_2$ are specified, $Z$ is determined. When $X_1$ and $X_2$ each follow normal distributions, $Z$ also follows a normal distribution. In fact, the normal case is the only one when $Z$ has the same distribution as $X_1$ and $X_2$. Moreover, the result cannot be extended from normal to the exponential. The distribution of $Z$ is not exponential given $X_1$ and $X_2$ are exponentially distributed unless we assume independence between them. We would like to develop similar procedures for non-normal distributions, look at the properties of the model, and investigate statistical inferences.

The physical meaning of the random variable $Z$ is that it allows the model to have the effect of fatal shock. Note that the linear relation described above should not be confused with the linear regression model. Indeed, the regression model is expressed as

$$X_2 = \beta X_1 + \alpha + \epsilon, \quad \epsilon \sim N\left(0, \sigma^2\right)$$

where the $\beta$ and $\alpha$ are unknown constants to be estimated from the relationship between the pair $(X_1, X_2)$. In the later part of the thesis, we will discuss a location-scale regression model for bivariate data.

# Chapter 2

## Some Bivariate Failure Time Distributions

Denote the bivariate joint probability density function (p.d.f.) of nonnegative lifetime variables $T_1$, $T_2$ as $f_{T_1,T_2}(t_1,t_2)$, and the survivor function as

$$S_{T_1,T_2}(t_1,t_2) = P(T_1 \geqslant t_1, T_2 \geqslant t_2) \tag{2.1}$$

and respectively the marginal survivor function as

$$S_j(t_j) = P(T_j \geqslant t_j) \quad j = 1,2$$

Given the lifetime variable $T_j$ is continuous, the joint p.d.f. is given by

$$f_{T_1,T_2}(t_1,t_2) = \frac{(-1)^2 \partial^2 S(t_1,t_2)}{\partial t_1 \partial t_2} \tag{2.2}$$

and by [26, Lawless (2002)], the hazard functions, which specify the joint distribution of $T_1$ and $T_2$, are denoted by

$$\begin{aligned} \lambda_j(t) &= \frac{-\partial S(t_1,t_2)/\partial t_j}{S(t_1,t_2)}|_{t_j=t_i=t}, \\ \lambda_{ij}(t_i|t_j) &= \frac{-\partial^2 S(t_1,t_2)/\partial t_i \partial t_j}{\partial S(t_1,t_2)/\partial t_j}, \qquad t_i > t_j,\ i,\ j=1,2,\ j \neq i \end{aligned} \tag{2.3}$$

The lifetimes $T_1, T_2$ are not in general independent, such as in the case of lifetimes of a pair of twins. Literature is abundant with methods of modeling bivariate distributions. For models with specified continuous marginal distributions, the joint survivor function can be represented by a parametric family of copulas such as models considered by [5, Clayton (1978)]. Extensive work have been done on the construction of bivariate exponential models as in [15, Gumbel (1960)], [13, Freund (1961)] and [40, Sarkar (1987)]. [32, Marshall and Olkin (1967)] and [27, Lee (1979)] constructed bivariate Weibull models by power transformation of the marginal of a bivariate exponential. [33, Marshall and Olkin (1988)] derived general families of bivariate distributions from mixture models by transformation. [17, Hougaard (1986)] discussed another common approach through random effects which will be introduced in following sections.

## 2.1 Linearly Associated Bivariate Failure Time Distributions

Let $X_1$ and $X_2$ be fixed marginally as exponential random variables with hazard rates $\lambda_1$ and $\lambda_2$, respectively. Then by introducing a latent variable, $Z$, statistically independent of $X_1$, a linear relationship is formed between $X_1$ and $X_2$ by setting

$$X_2 = aX_1 + Z, \tag{2.4}$$

for $a > 0$. [21, Iyer, Manjunath and Manivasakan (2002)] and [22, Iyer and Manjunanth (2004)] show through Laplace transforms, the distribution of the latent variable

$Z$ can be completely and uniquely characterized as the product of a Bernoulli random variable with $P(Z = 0) = a\lambda_2/\lambda_1$ and a continuous random variable having the same distribution as $X_2$. Therefore, $Z$ is distributed as mixture of a point mass at zero and an exponential with hazard rate $\lambda_2$. Note that when $Z = 0$ then $X_2$ is proportional to $X_1$ with proportionality constant $a$, which is fixed and known.

For the special case of $a = 1$ in (2.4), there is a positive probability for simultaneously events, i.e., $P(X_2 = X_1) > 0$. This phenomenon is often referred to as a "fatal shock" in reference to the now famous bivariate exponential proposed by [32, Marshall and Olkin (1967)]. Most bivariate exponential and Weibull models proposed in the literature share this property, including, for example, the multivariate Weibull proposed by [16, Hanagal (1996)]. In system reliability theory, [39, Rausand and Hoyland (2004)] refers to this situation as "common cause failures" or as "cascading failures" when the failure of one component is initiated by the failure of another in a system. There are many realistic applications of this model in the physical and biological sciences, such as, in medical research where simultaneous failure can occur in pairs of organs (kidneys, livers and eyes), in engineering where a random shock to a system of components may cause simultaneous failures, or in animal chemoprevention studies where several tumors may become palpable on the same day. Since $a > 0$, the model driven by (2.4) is less restrictive in that it includes the possibility for simultaneous failure ($a = 1$) and proportional failure times with proportionality constant $a$, i.e.,$P(X_2 = aX_1) = P(Z = 0) = p > 0$.

### 2.1.1 Linearly Associated Bivariate Exponential (BVE)

Suppose the continuous random variable $X_i$ has an exponential pdf with hazard $\lambda_i$,

$$f_{X_i}(x) = \lambda_i e^{-\lambda_i x} I(x > 0), \tag{2.5}$$

$i = 1, 2$ and $I(\cdot)$ is the indicator function.

Based on the linear structure given in (2.4) and the fact that $Z$ has a point mass at zero, we see that

$$X_2 = aX_1 + Z = \begin{cases} aX_1 & \text{if } Z = 0 \\ aX_1 + Z & \text{if } Z \neq 0 \end{cases},$$

where $P(Z = 0) = p$, $P(Z \neq 0) = 1 - p$, $p = a\lambda_2/\lambda_1$, and $Z$ is independent of $X_1$. Since $Z$ is a mixture of discrete and continuous distributions with a point mass at 0, i.e., $P(Z = 0) = p$, we can use the Direc delta to express the distribution of $Z$ as the following integrable density function,

$$f_Z(z) = p\delta(z) + (1 - p) f_{X_2}(z) I(z > 0), \tag{2.6}$$

where $I(\cdot)$ is an indicator function and $\delta(\cdot)$ is *Dirac delta function.* More details and applications of the $\delta$-function can be found in [1, Au and Tam (1999)] and [25, Khuri

(2004)]. Here, we define the $\delta$-function through its following mathematical properties:

$$\delta(t) = 0, \text{ if } t \neq 0, \text{and} \int_{-\infty}^{\infty} \delta(t)dt = 1 \tag{2.7}$$

If $h(t)$ is a real function with simple roots $t_1, \ldots, t_n$ and is differentiable at each root with $h'(t_i) \neq 0$, $i = 1, \ldots, n$, then

$$\delta(h(t)) = \sum_{i=1}^{n} \frac{\delta(t - t_i)}{|h'(t_i)|} \Rightarrow \delta(ct) = \frac{1}{|c|}\delta(t), c \neq 0, \Rightarrow \delta(-t) = \delta(t) \tag{2.8}$$

$$\int_{-\infty}^{\infty} g(t)\delta(t - t_0)dt = g(t_0) \Rightarrow \int_{-\infty}^{\infty} \delta(t - t_0)dt = 1 \tag{2.9}$$

[25, Khuri (2004)] demonstrates how the $\delta$-function can be used to generalize distribution theory and provides a unified approach in finding transformations, without regard to whether the transformation is one-to-one and without the computation of the Jacobian matrix. This property proves quite useful in the distribution derivations in this paper, since each is discontinuous on a line-transect, i.e., $P(X_2 = aX_1) > 0$.

Since $X_1$ is independent of $Z$, from (2.5) and (2.6), we can write the joint pdf as:

$$f_{X_1,Z}(x, z) = f_{X_1}(x)f_Z(z) = pf_{X_1}(x)\delta(z) + (1 - p)f_{X_1}(x)f_{X_2}(z)I(z > 0). \tag{2.10}$$

Notice that (2.6) and (2.10) are stated for any positive support distributions rather than the specific exponentials given in 2.5. Theorem 1 similarly expresses the joint density in terms of general positive support distributions. After the proof of Theorem 1, we give the resulting bivariate exponential.

**Theorem 2.1** *Let $X_1$ be a positive support random variable with marginal density $f_{X_1}(x_1)$. Let $Z$ be a random variable with density function given in 2.6 and let $X_2 = aX_1 + Z$, $a > 0$. Denote the variance of $X_i$ as $\sigma_i^2, i = 1, 2$, assume $\sigma_2 > a\sigma_1$ and let $p = a\sigma_1/\sigma_2$. Then the joint pdf of $X_1$ and $X_2$ is given by*

$$f_{X_1X_2}(x_1, x_2) = pf_{X_1}(x_1)\delta(x_2-ax_1)+(1-p)f_{X_1}(x_1)f_{X_2}(x_2-ax_1)I(x_2 > ax_1), x_2 \geq ax_1, \tag{2.11}$$

*and the variance/covariance matrix, $\Sigma$, and correlation matrix, $\rho$, are given as*

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & a\sigma_1^2 \\ a\sigma_1^2 & \sigma_2^2 \end{bmatrix} \text{ and } \boldsymbol{\rho} = \begin{bmatrix} 1 & a\sigma_1/\sigma_2 \\ a\sigma_1/\sigma_2 & 1 \end{bmatrix}, \tag{2.12}$$

*and $\Sigma$ positive definite.*

**Proof.** Following [25, Khuri (2004)], we find the joint density, $f(x_1, x_2)$ as

$$\begin{aligned} f(x_1, x_2) &= \int_{-\infty}^{\infty} f_{X_1,Z}(x_1, z)\,\delta(ax_1 + z - x_2)\,dz \\ &= \int_{-\infty}^{\infty} \{pf_{X_1}(x_1)\,\delta(z)\,\delta(ax_1 + z - x_2) \\ &\quad + (1-p)\,f_{X_1}(x_1)\,f_{X_2}(x_2)\,\delta(ax_1 + z - x_2)\}dz \end{aligned}$$

from 2.10. Note that 2.7 implies $\delta(z)\delta(ax_1+z-x_2)=0$ when $z\neq 0$ and $\delta(ax_1+z-x_2)=\delta(z-(x_2-ax_1))$, so

$$\begin{aligned} f(x_1,x_2) &= \int_{-\infty}^{\infty}\{pf_{X_1}(x_1)\delta(z)\delta(x_2-ax_1) \\ &\quad +(1-p)f_{X_1}(x_1)f_{X_2}(z)\delta(z-(x_2-ax_1))\}dz \\ &= pf_{X_1}(x_1)\delta(x_2-ax_1)\int_{-\infty}^{\infty}\delta(z)dz \\ &\quad +(1-p)f_{X_1}(x_1)\int_{-\infty}^{\infty}f_{X_2}(z)\delta(z-(x_2-ax_1))dz \\ &= pf_{X_1}(x_1)\delta(x_2-ax_1)+(1-p)f_{X_1}(x_1)f_{X_2}(x_2-ax_1)I(x_2>ax_1), \end{aligned}$$

where the first integral evaluation is from 2.7 and the second from 2.9. The covariance can be found by noting that

$$\mathrm{cov}(X_1,X_2)=\mathrm{cov}(X_1,aX_1-Z)=\mathrm{cov}(X_1,aX_1)+\mathrm{cov}(X_1,Z)=a\sigma_1^2,$$

using the fact that $X_1$ and $Z$ are independent. ■

From Theorem 3, if we let $X_i$ be an exponential given in 2.5, then 2.11 gives a bivariate exponential, henceforth referred to as the BVE$(\lambda_1,\lambda_2,a)$,

$$f(x_1,x_2)=p\lambda_1 e^{-\lambda_1 x_1}\delta(x_2-ax_1)+(1-p)\lambda_1\lambda_2 e^{-\lambda_2 x_2}e^{-(\lambda_1-a\lambda_2)x_1}I(x_2>ax_1), \quad (2.13)$$

where $p=a\sigma_1/\sigma_2=a\lambda_2/\lambda_1=\mathrm{corr}(X_1,X_2)$. Note that $\mathrm{cov}(X_1,X_2)=a/\lambda_1^2$. The density given in 2.13 differs from the one presented in [21, Iyer, Manjunath, and

Manivasakan (2002)] and [22, Iyer and Manjunath (2004)] due to an error in their derivation.

The joint cumulative distribution function (JCDF) for the exponential given in 2.13 can be written as

$$\begin{aligned}
F(x_1, x_2) &= p\lambda_1 \int_0^{x_1} \int_0^{x_2} e^{-\lambda_1 u} \delta(v - au) dv du \\
&\quad +(1-p)\lambda_1\lambda_2 \int_0^{x_1} \int_{au}^{x_2} e^{-\lambda_2 v} e^{-(\lambda_1 - a\lambda_2)u} dv du \\
&= p\lambda_1 \int_0^{x_1} \left\{ e^{-\lambda_1 u} \int_0^{x_2} \delta(v - au) dv \right\} du \\
&\quad +(1-p)\lambda_1 \int_0^{x_1} e^{-(\lambda_1 - a\lambda_2)u} (e^{-a\lambda_2 u} - e^{-\lambda_2 x_2}) du \\
&= p\lambda_1 \int_0^{x_1} e^{-\lambda_1 u} du \\
&\quad +(1-p)\lambda_1 \int_0^{x_1} e^{-\lambda_1 u} du - (1-p)\frac{\lambda_1}{(\lambda_1 - a\lambda_2)} e^{-\lambda_2 x_2} (1 - e^{-(\lambda_1 - a\lambda_2)x_1}).
\end{aligned}$$

Recalling that $p = a\lambda_2/\lambda_1$, we get

$$F(x_1, x_2) = (1 - e^{-\lambda_1 x_1}) + e^{-\lambda_2 x_2}(e^{-(\lambda_1 - a\lambda_2)x_1} - 1), x_2 \geq ax_1, \tag{2.14}$$

which we see is discontinuous at $y = ax$. Note that the expression in (11) is the exact expression given in [22, Iyer et al. (2004)]. Similarly, since $S(x_1, x_2) = 1 + F(x_1, x_2) - F(\infty, x_2) - F(x_1, \infty)$, the joint survival function (JSF) is given as

$$S(x_1, x_2) = e^{-\lambda_2 x_2} e^{-(\lambda_1 - a\lambda_2)x_1}. \tag{2.15}$$

If we let $X_{(1)} = \min\{X_1, X_2\}$, then from 2.15 we see that

$$P(X_{(1)} > t) = P(X_1 > t, X_2 > t) = S(t,t) = e^{-\lambda_2 t}e^{-(\lambda_1 - a\lambda_2)t} = e^{-(\lambda_1 - a\lambda_2 + \lambda_2)t}$$

which is the survival function for an exponential$(\lambda_1 - a\lambda_2 + \lambda_2)$. Further, with minor adaptation, $X_1$ and $X_2$ are said to have a joint distribution with Weibull minimums after arbitrary scaling, as defined in [27, Lee (1979)].

In Theorem 4, below, we give the maximum likelihood estimators of $\lambda_1$ and $\lambda_2$ based on the joint likelihood expression. We refer to these estimators as $\hat{\lambda}_1$ and $\hat{\lambda}_2$ and these will be compared to the marginal MLE's, denoted as $\hat{\lambda}_1^*$ and $\hat{\lambda}_2^*$, which we give immediately following the proof of Theorem 2. We define the marginal MLE's as those estimators that maximize the univariate marginal likelihood functions separately for $\lambda_1$ and $\lambda_2$. [26, Lawless (2002)] refers to the analysis of the marginal MLE's as assuming "working independence". As we will show later, assuming working independence comes at a cost in terms of mean-squared-error.

**Theorem 2.2** *For a given random sample of size n, $(x_{1i}, x_{2i}), i = 1, \ldots, n,$, from a bivariate exponential($\lambda_1, \lambda_2$), the joint maximum likelihood estimators of $(\lambda_1, \lambda_2)$ is $(\hat{\lambda}_1, \hat{\lambda}_2)$, where*

$$\hat{\lambda}_1 = \frac{a}{\bar{x}_2} + \frac{(n-k)}{n\bar{x}_1} \text{ and } \hat{\lambda}_2 = \frac{1}{\bar{x}_2} \tag{2.16}$$

*$k = \sum_{i=1}^{n} I(x_2 - ax_1 = 0)$, $\bar{x}_1 = \sum x_1/n$ and $\bar{x}_2 = \sum x_2/n$. Also, $\hat{\lambda}' = (\hat{\lambda}_1, \hat{\lambda}_2)'$ is approximately bivariate normal with mean vector $\lambda$ and variance/covariance matrix*

$\Sigma$, *where*

$$\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \ \ and\ \Sigma = \frac{1}{n}\begin{pmatrix} \lambda_1(\lambda_1 - a\lambda_2) + a^2\lambda_2^2 & a\lambda_2^2 \\ a\lambda_2^2 & \lambda_2^2 \end{pmatrix} \tag{2.17}$$

*and* $Corr(\hat{\lambda}_1, \hat{\lambda}_2) = ((1-p)/p^2 + 1)^{-1/2}$.

**Proof.** For a given pair of bivariate exponential random variables, $(x_{1i}, x_{2i}), i = 1, \ldots, n$, it is easy to see that maximum likelihood estimation under the likelihood function $L(\lambda_1, \lambda_2 | x_{1i}, x_{2i})$ is equivalent to the maximum likelihood estimation under $L(\lambda_1, \lambda_2 | x_{1i}, z_i)$, since $x_{2i} = ax_{i1} + z_i$. From 2.13, the likelihood function for a random sample of size $n$ of pairs $(x_{1i}, z_i)$ for $1 \le i \le n$ is given by

$$\begin{aligned} L(\lambda_1, \lambda_2) &= \prod_{i=1}^{n} \left(p\lambda_1 e^{-\lambda_1 x_{1i}}\right)^{r_i} \left((1-p)\lambda_1\lambda_2 e^{-\lambda_1 x_{1i} - \lambda_2 z_i}\right)^{1-r_i} \\ &= \prod_{i=1}^{n} \left[a\lambda_2 e^{-\lambda_1 x_{1i}}\right]^{r_i} \left[(\lambda_1 - a\lambda_2)\lambda_2 e^{-\lambda_1 x_{1i}} e^{-\lambda_2 z_i}\right]^{1-r_i} \\ &= (a\lambda_2)^{\sum_i r_i} e^{-\lambda_1 \sum_i r_i x_{1i}} \left[(\lambda_1 - a\lambda_2)\lambda_2\right]^{\sum_i (1-r_i)} \\ &\quad \times e^{-\lambda_1 \sum_i (1-r_i) x_{1i}} e^{-\lambda_2 \sum_i (1-r_i) z_i}. \end{aligned}$$

where $r_i = I(z_i = 0)$. The log-likelihood is given as

$$\begin{aligned}
LL(\lambda_1, \lambda_2) &= \log(a\lambda_2) \sum_i r_i - \lambda_1 \sum_i r_i x_{1i} \\
&\quad + \log\left[(\lambda_1 - a\lambda_2)\lambda_2\right] \sum_i (1 - r_i) - \lambda_1 \sum_i (1 - r_i) x_{1i} - \lambda_2 \sum_i (1 - r_i) z_i \\
&= \log(a) \sum_i r_i + n \log(\lambda_2) - \lambda_1 \sum_i r_i x_{1i} \\
&\quad + \log(\lambda_1 - a\lambda_2) \sum_i (1 - r_i) - \lambda_1 \sum_i (1 - r_i) x_{1i} - \lambda_2 \sum_i z_i,
\end{aligned}$$

since $\sum (1 - r_i) z_i = \sum z_i$. The partial derivative with respect to $\lambda_1$ is given as

$$\frac{\partial LL}{\partial \lambda_1} = -\sum_i r_i x_{1i} + \frac{\sum_i (1 - r_i)}{\lambda_1 - a\lambda_2} - \sum_i (1 - r_i) x_{1i} = \frac{\sum_i (1 - r_i)}{\lambda_1 - a\lambda_2} - \sum_i x_{1i}.$$

Setting $\partial LL / \partial \lambda_1 = 0$ gives the following likelihood equation

$$\frac{\lambda_1 - a\lambda_2}{\sum_i (1 - r_i)} = \frac{1}{\sum_i x_{1i}}. \tag{2.18}$$

The partial derivative with respect to $\lambda_2$ is given by

$$\frac{\partial LL}{\partial \lambda_2} = \frac{n}{\lambda_2} - \frac{a \sum_i (1 - r_i)}{\lambda_1 - a\lambda_2} - \sum_i z_i = \frac{n}{\lambda_2} - a \sum_i x_{1i} - \sum_i z_i,$$

by substituting the right hand side of 2.18 for the left hand side.

Setting $\partial LL/\partial\lambda_2 = 0$ we get the second likelihood equation

$$\frac{n}{\lambda_2} = a\sum_i x_{1i} + \sum_i z_i \tag{2.19}$$

Solving for $\lambda_2$ in 2.19, we have

$$\lambda_2 = \frac{n}{a\sum_i x_{1i} + \sum_i z_i} = \frac{1}{a\frac{\sum_i x_{1i}}{n} + \frac{1}{n}\sum_i z_i} = \frac{1}{a\bar{x}_1 + \bar{z}}.$$

Solving for $\lambda_1$ in 2.18 and substituting the above value of $\lambda_2$ gives

$$\lambda_1 = \frac{a}{a\bar{x}_1 + \bar{z}} + \frac{\sum_i(1-r_i)}{\sum_i x_{1i}} = \frac{a}{a\bar{x}_1 + \bar{z}} + \frac{\sum_i(1-r_i)}{n\bar{x}_1}.$$

Since $\bar{x}_2 = a\bar{x}_1 + \bar{z}$ and $k = \sum(1-r_i)$, it follows that the estimators in 2.16 are the solutions to the likelihood equations. It easily follows that the Hessian matrix is given by

$$H(\lambda_1,\lambda_2) = \left(\frac{\partial^2 LL(\lambda_1,\lambda_2)}{\partial\lambda_i\partial\lambda_j}\right) = \begin{bmatrix} -\frac{\sum_i(1-r_i)}{(\lambda_1-a\lambda_2)^2} & a\frac{\sum_i(1-r_i)}{(\lambda_1-a\lambda_2)^2} \\ a\frac{\sum_i(1-r_i)}{(\lambda_1-a\lambda_2)^2} & -\frac{n}{\lambda_2^2} - a^2\frac{\sum_i(1-r_i)}{(\lambda_1-a\lambda_2)^2} \end{bmatrix}.$$

and

$$det(H(\hat{\lambda}_1,\hat{\lambda}_2)) = \frac{n(n-k)}{\hat{\lambda}_2^2(\hat{\lambda}_1 - a\hat{\lambda}_2)} > 0,$$

since $\hat{\lambda}_1 - a\hat{\lambda}_2 = (n-k)/(n\bar{x}_1) > 0$. Therefore, estimators in 2.16 are the maximum likelihood solutions. Since $k$ is distributed as a Binomial$(n, 1-p)$, i.e., $E(k) = n(1-p)$, where $p = a\lambda_2/\lambda_1$ and $\lambda_1 - a\lambda_2 = \lambda_1(1 - a\lambda_2/\lambda_1)$, Fisher's Information (see page 546, [26, Lawless (2002)]) is given as

$$I(\lambda_1, \lambda_2) = E\left(-\frac{\partial^2 LL(\lambda_1, \lambda_2)}{\partial\lambda_i \partial\lambda_j}\right) = \begin{bmatrix} \frac{n}{\lambda_1(\lambda_1 - a\lambda_2)} & -\frac{a \cdot n}{\lambda_1(\lambda_1 - a\lambda_2)} \\ -\frac{a \cdot n}{\lambda_1(\lambda_1 - a\lambda_2)} & \frac{n}{\lambda_2^2} + \frac{a^2 \cdot n}{\lambda_1(\lambda_1 - a\lambda_2)} \end{bmatrix}.$$

Therefore, from [26, Lawless (2002)], $\hat{\lambda}$ is approximately normal with mean $\lambda' = (\lambda_1, \lambda_2)'$ and variance/covariance matrix $\Sigma = I^{-1}(\lambda_1, \lambda_2)$. ■

As mentioned previously, alternatives to the joint MLE's given in 2.16 can by found by maximizing the marginal likelihood expressions separately for $\lambda_1$ and $\lambda_2$. These marginal MLE's are well-known (see page 54 of [26, Lawless (2002)]) and are given as

$$\hat{\lambda}_1^* = \frac{1}{\bar{x}_1} \qquad \text{and} \qquad \hat{\lambda}_2^* = \frac{1}{\bar{x}_2}. \tag{2.20}$$

We observe that the marginal and joint MLE's for $\lambda_2$ are identical, i.e., $\hat{\lambda}_2^* \equiv \hat{\lambda}_2$, but for $\lambda_1$ the MLE's are quite different for this model.

We can adapt the likelihood for censored data. Suppose $(X_{1i}, X_{2i})$, $i = 1, \ldots, n$ represents a random sample of size $n$ from a population with joint survival function given in (11). If the observations are subject to right censoring with potential censoring times $C_{1i}$ and $C_{2i}$, then the data will come in the form of $(x_{1i}, x_{2i})$, $x_{1i} = \min(X_{1i}, C_{1i})$ and $x_{2i} = \min(X_{2i}, C_{2i})$, $i = 1, \ldots, n$. Then, following [26, Lawless (2002)], the likelihood function for a given observation can be expressed as

$$L^*(\lambda_1, \lambda_2) = L(\lambda_1, \lambda_2)^{\gamma_{1i}\gamma_{2i}} \left[\frac{-\partial S}{\partial x_{1i}}\right]^{\gamma_{1i}(1-\gamma_{2i})} \left[\frac{-\partial S}{\partial x_{2i}}\right]^{(1-\gamma_{1i})\gamma_{2i}} \cdot S^{(1-\gamma_{1i})(1-\gamma_{2i})},$$

where $S = S(x_{1i}, x_{2i})$ is defined in 2.14, $L(\lambda_1, \lambda_2)$ is defined in the proof of Theorem 2, and $\gamma_{ji} = 1$ if the $ij^{th}$ data value is not censored and zero otherwise.

### 2.1.2 Linearly Associated Bivariate Weibull

If $X_i$ is exponential random variable with hazard rate $\lambda_i$, $i = 1, 2$ and pdf given in 2.5, then for a fixed $\beta > 0$, it is well known that $Y_i = X_i^{1/\beta}$ is distributed as Weibull$(\lambda_i, \beta)$ random variable with pdf given as,

$$f_{Y_i}(y) = \beta\lambda_i y^{\beta-1} e^{-\lambda_i y^\beta} I(y > 0), i = 1, 2 \tag{2.21}$$

From this one could derive a joint Weibull$(\lambda_1, \lambda_2, \beta)$ with mariginals, Weibull$(\lambda_i, \beta), i = 1, 2$, with the following linear relationship

$$Y_2^\beta = aY_1^\beta + Z^\beta \tag{2.22}$$

Since the marginal are Weibull with pdf given in 2.21, we know that (see page 18, Lawless (2002))

$$EY_i^r = \lambda_i^{-r/\beta}\Gamma(1 + r/\beta) \text{ and } \sigma_{y_i}^2 = \frac{1}{\lambda_i^{2/\beta}}\big[\Gamma(1 + \frac{2}{\beta}) - \Gamma^2(1 + \frac{1}{\beta})\big], \qquad i = 1, 2. \quad (2.23)$$

In Theorem 3, below, we give the derived bivariate Weibull, based on the structure in 2.22, along with its covariance structure. The proof of this theorem involves the $\delta$-function and its properties given in (2.7), (2.8) and (2.9).

**Theorem 2.3** *Suppose $(X_1, X_2)$ has the joint distribution given in 2.5, with exponential marginal given in 2.5. Let $Y_i = X_i^{1/\beta}, i = 1, 2,\ \beta > 0$. Then the joint density of $(Y_1, Y_2)$ is given as*

$$f(y_1, y_1) = p\lambda_1\beta y_1^{\beta-1}e^{-\lambda_1 y_1^\beta}\delta(y_1, y_2) + (1-p)\lambda_1\lambda_2\beta^2 y_1^{\beta-1}y_2^{\beta-1}e^{-\lambda_2 y_2^\beta}e^{-(\lambda_1 - a\lambda_2)y_1^\beta}I(y_1, y_2), \quad (2.24)$$

*where $\delta(y_1, y_2) = \delta(y_2 - a^{1/\beta}y_1)$ and $I(y_1, y_2) = I(y_2 > a^{1/\beta}y_1)$. The marginal distribution of $Y_i$ is given in 2.21, $i = 1, 2$ and*

$$\begin{aligned}\mathrm{Cov}(Y_1, Y_2) &= \frac{a^{1+\frac{1}{\beta}}\lambda_2\Gamma(1 + 2/\beta)}{\lambda_1^{1+2/\beta}} + \frac{\lambda_2}{(\lambda_1 - a\lambda_2)^{\frac{1}{\beta}}}\int_0^\infty IG(g(y_2), 1 + 1/\beta)f_{Y_2}(y_2)dy_2 \\ &\quad - \frac{\Gamma^2(1 + 1/\beta)}{(\lambda_1\lambda_2)^{1/\beta}}\end{aligned}$$

*where $g(y_2) = (\lambda_1 - a\lambda_2)y_2^\beta/a$ and $IG(x, k) = \int_0^x t^{k-1}e^{-t}dt$ as defined on page 25 of Lawless (2003).*

**Proof.** To derive the density in 2.24, we follow the transformation approach given in [25, Khuri (2004)], by using the appropriate $\delta$-functions as follows,

$$\begin{aligned}
f(y_1, y_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2)\delta(x_1^{1/\beta} - y_1)\delta(x_2^{1/\beta} - y_2)dx_2dx_1 \\
&= p\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1}(x_1)\delta(x_2 - ax_1)\delta(x_1^{1/\beta} - y_1)\delta(x_2^{1/\beta} - y_2)dx_2dx_1 \\
&\quad +(1-p)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1}(x_1)f_{X_2}(x_2 - ax_1)\delta(x_1^{1/\beta} - y_1)\delta(x_2^{1/\beta} - y_2)dx_2dx_1 \\
&= p \text{ Part } 1 + (1-p) \text{ Part } 2
\end{aligned}$$

Since $\delta(x_2 - ax_1)\delta(x_2^{1/\beta} - y_2) = \delta(x_2 - ax_1)\delta((ax_1)^{1/\beta} - y_2)$, we have

$$\begin{aligned}
\text{Part } 1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1}(x_1)\delta(x_1^{1/\beta} - y_1)\delta(x_2 - ax_1)\delta((ax_1)^{1/\beta} - y_2)dx_2dx_1 \\
&= \int_{-\infty}^{\infty}\left\{\delta(x_1^{1/\beta} - y_1)\delta(x_2 - ax_1)\delta((ax_1)^{1/\beta} - y_2)f_{X_1}(x_1)\int_{-\infty}^{\infty}\delta(x_2 - ax_1)dx_2\right\}dx_1 \\
&= \int_{-\infty}^{\infty}\delta(x_1^{1/\beta} - y_1)\delta((ax_1)^{1/\beta} - y_2)f_{X_1}(x_1)dx_1 \\
&= \int_{-\infty}^{\infty}\delta(x_1^{1/\beta} - y_1)\delta(a^{1/\beta}y_1 - y_2)f_{X_1}(x_1)dx_1 \\
&= \delta(a^{1/\beta}y_1 - y_2)\int_{-\infty}^{\infty}\beta y_1^{\beta-1}f_{X_1}(x_1)\delta(x_1 - y_1^{\beta})dx_1 \\
&= \beta y_1^{\beta-1}f_{X_1}(y_1^{\beta})\delta(y_2 - a^{1/\beta}y_1).
\end{aligned}$$

Similarly,

$$\begin{aligned}\text{Part 2} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1}(x_1) f_{X_2}(x_2 - a x_1)\beta^2 y_1^{\beta-1} y_2^{\beta-1}\delta(x_1 - y_1^{\beta})\delta(x_2 - y_2^{\beta}) dx_2 dx_1 \\ &= \beta^2 y_1^{\beta-1} y_2^{\beta-1}\int_{-\infty}^{\infty} f_{X_1}(x_1) f_{X_2}(y_2^{\beta} - a x_1)\delta(x_1 - y_1^{\beta}) dx_1 \\ &= \beta^2 y_1^{\beta-1} y_2^{\beta-1} f_{X_1}(y_1^{\beta}) f_{X_2}(y_2^{\beta} - a y_1^{\beta}).\end{aligned}$$

Putting together Parts 1 and 2 above we get

$$\begin{aligned}f(y_1, y_2) &= p\beta y_1^{\beta-1} f_{X_1}(y_1^{\beta})\delta(y_2 - a^{1/\beta} y_1) \\ &\quad + (1-p)\,\beta^2 y_1^{\beta-1} y_2^{\beta-1} f_{X_1}(y_1^{\beta}) f_{X_2}(y_2^{\beta} - a y_1^{\beta}) I(\cdot)\end{aligned} \tag{2.25}$$

$I(\cdot)$ is an indicator function for all $\{(y_1, y_2) : y_2 \geq a^{1/\beta} y_1\}$. Since $X_i$, $i = 1, 2$, have densities given in 2.5, the joint density given in 2.24 follows directly from 2.25 above. Given that $X_1$ and $X_2$ have the density given in 2.5, the result follows directly.

Now, to derive the covariance expression we must find

$$E(Y_1 \cdot Y_2) = \int_0^{\infty}\int_0^{\infty} y_1 y_2 f_{Y_1,Y_2}(y_1, y_2) dy_2 dy_1 = \text{Part A} + \text{Part B}$$

where

$$
\begin{aligned}
\text{Part A} &= p\int_0^{\infty}\int_0^{\infty}\lambda_1\beta y_1 y_2 y_1^{\beta-1}e^{-\lambda_1 y_1^{\beta}}\delta(y_1,y_2)dy_2dy_1 \\
&= \int_0^{\infty} y_1\cdot(a^{1/\beta}y_1)\cdot p\beta\lambda_1 y_1^{\beta-1}e^{-\lambda_1 y_1^{\beta}}dy_1 \qquad \text{(by } \delta\text{-function property (6))} \\
&= p\beta\lambda_1 a^{1/\beta}\int_0^{\infty} y_1^2 y_1^{\beta-1}e^{-\lambda_1 y_1^{\beta}}dy_1 \\
&= pa^{1/\beta}E(Y_2^2) = a^{1+1/\beta}\lambda_1^{-2/\beta-1}\lambda_2\Gamma(1+2/\beta) \qquad \text{(by (20) and } p = a\lambda_2/\lambda_1)
\end{aligned}
$$

and

$$
\begin{aligned}
\text{Part B} &= \int_0^{\infty}\int_0^{\infty} y_1 y_2 f_{Y_1,Y_2}(y_1,y_2)I_(y_2^{\beta} > ay_1^{\beta})dy_1dy_2 \\
&= (1-p)\beta\lambda_1\lambda_2\int_0^{\infty} y_2^{\beta}e^{-\lambda_2 y_2^{\beta}}dy_2\int_0^{y_2/a^{1/\beta}}\beta y_1^{\beta}e^{-(\lambda_1-a\lambda_2)y_1^{\beta}}dy_1 \\
&= \frac{(1-p)\lambda_1\lambda_2}{\lambda_1-a\lambda_2}\int_0^{\infty}\beta y_2^{\beta}e^{-\lambda_2 y_2^{\beta}}dy_2\int_0^{y_2/a^{1/\beta}}\beta(\lambda_1-a\lambda_2)y_1 y_1^{\beta-1}e^{-(\lambda_1-a\lambda_2)y_1^{\beta}}dy_1 \\
&= \frac{(1-p)\lambda_1\lambda_2}{\lambda_1-a\lambda_2}\int_0^{\infty}\beta y_2^{\beta}e^{-\lambda_2 y_2^{\beta}}dy_2\frac{1}{(\lambda_1-a\lambda_2)^{1/\beta}}\int_0^{\frac{(\lambda_1-a\lambda_2)}{a}y_2^{\beta}} t^{1/\beta}e^{-t}dt \\
&= \frac{(1-p)\lambda_1\lambda_2}{(\lambda_1-a\lambda_2)^{1+\frac{1}{\beta}}}\int_0^{\infty}\beta y_2^{\beta}e^{-\lambda_2 y_2^{\beta}}dy_2\int_0^{\frac{(\lambda_1-a\lambda_2)}{a}y_2^{\beta}} t^{1/\beta}e^{-t}dt.
\end{aligned}
$$

Substituting Parts A and B into the $E(Y_1 \cdot Y_2)$ expression above and subtracting $E(Y_1)\cdot E(Y_2)$ from 2.23, the covariance expression follows. □ ■

Henceforth, we will refer to the bivariate Weibull density given in 2.23 as a BVW$(\lambda_1,\lambda_2,\beta,a)$. The covariance structure given in Theorem 3 for the proposed BVW$(\lambda_1,\lambda_2,\beta,a)$ is quite complicated and not in closed-form. Similarly complicated structures are found with the multivariate Weibull proposed by [17, Hougaard (1986)] and [31, Lu and Bhattacharyya (1990)], as well as, the bivariate Weibull derived from [32, Marshall and Olkin (1967)] by taking the identical transformation. The correlation between $Y_1$ and $Y_2$, denoted $\rho_{(Y_1,Y_2)}$ is found as

$$\begin{aligned}\rho_{(Y_1,Y_2)} &= \frac{\mathrm{Cov}(Y_1,Y_2)}{\sqrt{\mathrm{Var}(Y_1)}\sqrt{\mathrm{Var}(Y_2)}} \\ &= \left(\frac{a\lambda_2}{\lambda_1}\right)^{1+\frac{1}{\beta}} \frac{\Gamma(1+\frac{2}{\beta})}{\Gamma(1+\frac{2}{\beta})-\Gamma^2(1+\frac{1}{\beta})} - \frac{\Gamma^2(1+\frac{1}{\beta})}{\Gamma(1+\frac{2}{\beta})-\Gamma^2(1+\frac{1}{\beta})} \\ &\quad + \frac{\Gamma(1+\frac{1}{\beta})\lambda_2}{\Gamma(1+\frac{2}{\beta})-\Gamma^2(1+\frac{1}{\beta})}\left[\frac{\lambda_1\lambda_2}{(\lambda_1-a\lambda_2)}\right]^{\frac{1}{\beta}} \int_0^\infty \beta y_2^\beta e^{-\lambda_2 y_2^\beta} IG(g(y_2),1+1/\beta) f_{Y_2}(y_2)dy_2.\end{aligned}$$

Similar to the likelihood for the exponential, we see that the likelihood for a given pair $(y_1, y_2)$ is given as

$$L(\lambda_1,\lambda_2,\beta) = \left(a\lambda_2\beta y_1^{\beta-1}e^{-\lambda_1 y_1^\beta}\right)^r \left((\lambda_1-a\lambda_2)\lambda_2\beta^2 y_1^{\beta-1}y_2^{\beta-1}e^{-\lambda_2 y_2^\beta}e^{-(\lambda_1-a\lambda_2)y_1^\beta}\right)^{1-r}, \tag{2.26}$$

where $r = I(y_2 - a^{1/\beta}y_1 = 0)$. The likelihood in (23) will lead to joint maximum likelihood estimators of $(\lambda_1, \lambda_2, \beta)$ given as $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\beta})$, where

$$\hat{\lambda}_1 = \frac{a}{\bar{y}_2^*} + \frac{(n-k)}{n\bar{y}_1^*}, \hat{\lambda}_2 = \frac{1}{\bar{y}_2^*}, \text{ and } \hat{\beta} \tag{2.27}$$

where $\bar{y}_j^* = \frac{1}{n}\sum_{i=1}^n y_{ji}^{\hat{\beta}}, i = 1, 2, k = \sum_{i=1}^n I(y_2^{\hat{\beta}} - ay_1^{\hat{\beta}} = 0)$. The estimator $\hat{\beta}$ represents the solution to the third likelihood equation found by differentiating the log-likelihood with respect to $\beta$, plugging the above estimators $\hat{\lambda}_1$ and $\hat{\lambda}_2$ and numerically solving for $\beta$.

The JCDF can be found directly from (11) as:

$$F(y_1, y_2) = F_{X_1,X_2}(y_1^\beta, y_2^\beta) = (1 - e^{-\lambda_1 y_1^\beta}) + e^{-\lambda_2 y_2^\beta}(e^{-(\lambda_1 - a\lambda_2)y_1^\beta} - 1), \qquad (2.28)$$

for $y_2 \geq a^\beta y_1$. From (25), it is easy to verify that the marginal have the Weibull distribution given in (18) by noting that:

$$\begin{aligned} F_{Y_1}(y_1) &= \lim_{y_2 \to \infty} F(y_1, y_2) = (1 - e^{-\lambda_1 y_1^\beta}) \\ F_{Y_2}(y_2) &= \lim_{y_1 \to \infty} F(y_1, y_2) = F(y_2/a^{1/\beta}, y_2) = (1 - e^{-\lambda_2 y_2^\beta}) \end{aligned}$$

Similarly, the JSF is given as:

$$S(y_1, y_2) = e^{-\lambda_2 y_2^\beta} e^{-(\lambda_1 - a\lambda_2)y_1^\beta} \qquad (2.29)$$

Note that if we let $Y_{(1)} = \min\{Y_1, Y_2\}$, then (26) gives

$$P(Y_{(1)} > t) = P(Y_1 > t, Y_2 > t) = S(t, t) = e^{-\lambda_2 t^\beta} e^{-(\lambda_1 - a\lambda_2)t^\beta} = e^{-(\lambda_1 - a\lambda_2 + \lambda_2)t^\beta}$$

which is the survival function for an Weibull($\lambda_1 - a\lambda_2 + \lambda_2, \beta$). It is easy to show that $Y_1$ and $Y_2$ have a joint distribution with Weibull minimums after arbitrary scaling, as defined in Lee (1979).

## 2.2 Bivariate Lifetime Distributions Based on Random Hazards

[17, Hougaard (1986)] provides a general method of constructing bivariate failure time distributions where both components in a system are affected by random hazards.

**Lemma 2.4** *Let $\lambda_1(t)$ and $\lambda_2(t)$ be two arbitrary hazard functions, and $\Lambda_1(t)$ and $\Lambda_2(t)$ be the corresponding cumulative hazard functions, where $\Lambda_j(t) = \int_0^t \lambda_j(x)dx$. Let $T_j$, $j = 1, 2$, be conditionally independent lifetimes given a specific quantity $Z$. The marginal hazard of $T_j$ is $Z\lambda_j$, and its cumulative hazard function is $Z\Lambda_j$. Then the conditional bivariate survivor function $P(T_1 \geq t_1, T_2 \geq t_2 | Z = z) = \exp(-z\Lambda.)$, where $\Lambda. = \Lambda_1(t) + \Lambda_2(t)$.*

**Proof.** The conditional joint survivor function of $T_j's$ given $Z$ is

$$\begin{aligned} S(t_1, t_2 | Z = z) &= P(T_1 > t_1, T_2 > t_2 | Z = z) \\ &= \prod_{j=1}^{2} S(t_j | Z = z) = \prod_{j=1}^{2} \exp\left[-\int_0^t z\lambda_j(x)dx\right] \\ &= \prod_{j=1}^{2} \exp\left[-z\int_0^t \lambda_j(x)dx\right] = \prod_{j=1}^{2} \exp\left[-z\Lambda_j(t)\right] \\ &= \exp\left[-z\sum_{j=1}^{2} \Lambda_j(t)\right] = \exp(-z\Lambda.) \end{aligned}$$

where $\Lambda. = \sum_{j=1}^{n} \Lambda_j(t)$ ■

**Definition 1** *[18, Hougaard (2000)] defines the positive stable distribution as:*

*Let $X_i$, $i = 1, ...n$, be independent, identically distributed random variables with positive supports. If there exists a scale factor function $c(n)$ having the form $n^{1/\alpha}$,*

*$\alpha \leq 1$, such that*

$$c\,(n)\,X =_D X_1 + ... + X_n,$$

*where $=_D$ denotes having the same distribution as. The the distribution of $X$ is called a positive stable distribution with two parameters $\alpha$ and $\delta$ given by the Laplace transform*

$$L\,(s) = \exp\left(-\delta s^{\alpha}/\alpha\right).$$

**Theorem 2.5** *Let $T_j$, $j = 1, 2$, be conditionally independent lifetimes given a specific quantity $Z$, which in term has a positive stable distribution with parameter $\alpha$ given by the Laplace transform*

$$E\left\{\exp\left(-\Lambda Z\right)\right\} = \exp\left(-\Lambda^{\alpha}\right), \quad \alpha \in (0, 1],$$

*The marginal hazard of $T_j$ is $Z\lambda_j$, and $\Lambda_j(t) = \int_0^t \lambda_j(x)dx$ is the cumulative hazard. Then the unconditional bivariate survivor function $P\left(T_1 \geq t_1, T_2 \geq t_2\right) = \exp\left(-\Lambda.^{\alpha}\right)$*

**Proof.** Assuming some regularity conditions and by lemma 1, we have the following result

$$\begin{aligned}
P\left(T_1 \geq t_1, T_2 \geq t_2\right) &= \int_{t_1}^{\infty}\int_{t_2}^{\infty} f\left(x_1, x_2\right) dx_1 dx_2 = \int_{t_1}^{\infty}\int_{t_2}^{\infty} \left[\int_{-\infty}^{\infty} f\left(x_1, x_2, z\right) dz\right] dx_1 dx_2 \\
&= \int_{t_1}^{\infty}\int_{t_2}^{\infty} \left[\int_{-\infty}^{\infty} f\left(x_1, x_2 | z\right) f\left(z\right) dz\right] dx_1 dx_2 \\
&= \int_{-\infty}^{\infty} \left[\int_{t_1}^{\infty}\int_{t_2}^{\infty} f\left(x_1, x_2 | z\right) dx_1 dx_2\right] f\left(z\right) dz \\
&= \int_{-\infty}^{\infty} P\left(T_1 > t_1, T_2 > t_2 | Z = z\right) f\left(z\right) dz \\
&= E\left[P\left(T_1 > t_1, T_2 > t_2 | Z = z\right)\right] \\
&= E\left[\exp\left(-z\Lambda.\right)\right] = \exp\left(-\Lambda.^{\alpha}\right)
\end{aligned}$$

■

### 2.2.1 The Bivariate Weibull Model (BVW) of Random Hazards

[17, Hougaard (2000)] derives a bivariate Weibull distribution with common shape parameter $\gamma$ such that the arbitrary hazard rate $\lambda_j(x) = \epsilon_j \gamma t^{\gamma-1}$. The marginal distributions are also Weibull with common shape $\alpha\gamma$, Using more conventional parameterizations, [31, Lu and Bhatacharyya (1990)] derive the same model without the assumption of conditional independence and common shape parameter, and render the survivor function as

$$\begin{aligned} S(t_1, t_2) &= P(T_1 \geq t_1, T_2 \geq t_2) \\ &= \exp\left\{-\left[\left(\frac{t_1}{\theta_1}\right)^{\beta_1/\alpha} + \left(\frac{t_2}{\theta_2}\right)^{\beta_2/\alpha}\right]^{\alpha}\right\}, \qquad 0 < \alpha \leq 1 \end{aligned} \tag{2.30}$$

as well as the moments

$$\begin{aligned} E(T_j) &= \theta_j \Gamma\left(1/\beta_j + 1\right) \\ Var(T_j) &= \theta_j^2 \left\{\Gamma\left(2/\beta_j + 1\right) - \Gamma^2\left(1/\beta_j + 1\right)\right\} \qquad j = 1,\ 2 \\ Cov(T_1, T_2) &= \theta_1\theta_2 \left[\Gamma(\alpha/\beta_1 + 1)\Gamma(\alpha/\beta_2 + 1)\Gamma(1/\beta_1 + 1/\beta_2 + 1)\right. \\ &\quad \left.\Gamma(1/\beta_1 + 1)\Gamma(1/\beta_2 + 1)\Gamma(\alpha/\beta_1 + \alpha/\beta_2 + 1)\right] \\ &\quad \div\Gamma(\alpha/\beta_1 + \alpha/\beta_2 + 1) \end{aligned}$$

Lu and Bhatacharyya also studied the statistical properties of the BVW and showed that for the bivariate exponential case where $\beta_1 = \beta_2 = 1$, the correlation

$$\begin{aligned} Corr(T_1, T_2) &= \frac{Cov(T_1, T_2)}{\sqrt{Var(T_1)\,Var(T_2)}} \\ &= 2\Gamma^2(\alpha + 1)/\Gamma(2\alpha + 1) - 1 \end{aligned}$$

is free of the marginal parameters and could be conveniently utilized in simulation.

### 2.2.2 The Farlie-Gumbel-Morgenstern (FGM) Family

A general class of bivariate distributions is the Farlie-Gumbel-Morgenstern (FGM) Family as proposed by [36], [15], and [12]. In general, the FGM family of bivariate distributions is defined as:

Let $F_{X_i}(x_i)$, $i = 1, 2$, be a cdf of the continuous random variable $X_i$, then the joint cdf of $X_1$ and $X_2$ with an FGM bivariate distribution is given as

$$F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1) F_{X_2}(x_2) \left[1 + \alpha \left(1 - F_{X_1}(x_1)\right)\left(1 - F_{X_2}(x_2)\right)\right] \tag{2.31}$$

where $\alpha$ is a dependence parameter such that $|\alpha| < 1$.

It can be shown that the properties of a bivariate cdf all hold for the cdf defined in the above (2.31):

1. $F_{X_1,X_2}(x_1, \infty) = F_{X_1}(x_1)$ and $F_{X_1,X_2}(\infty, x_2) = F_{X_2}(x_2)$,
2. $F_{X_1,X_2}(x_1, -\infty) = F_{X_1,X_2}(-\infty, x_2) = F_{X_1,X_2}(-\infty, -\infty) = 0$,
3. $F_{X_1,X_2}(\infty, \infty) = 1$.

Let $X_1, X_2$ be continuous random variables having marginal pdfs $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$ and follow the FGM bivariate distribution,then the joint pdf, if exists, is given as

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) \left[1 + \alpha \left(1 - 2F_{X_1}(x_1)\right)\left(1 - 2F_{X_2}(x_2)\right)\right]. \tag{2.32}$$

Assuming the joint cdf $F_{X_1,X_2}(x_1,x_2)$ is differenciable to the second order, then

$$\begin{aligned}
f_{X_1,X_2}(x_1,x_2) &= \frac{\partial^2}{\partial x_1 \partial x_2} F_{X_1,X_2}(x_1,x_2) \\
&= \frac{\partial}{\partial x_1}\left[\frac{\partial}{\partial x_2} F_{X_1,X_2}(x_1,x_2)\right] \\
&= \frac{\partial}{\partial x_1}\left\{\frac{\partial}{\partial x_2} F_{X_1}(x_1) F_{X_2}(x_2)\right. \\
&\quad \left. +\alpha \left(F_{X_1}(x_1) - F_{X_1}^2(x_1)\right)\left(F_{X_2}(x_2) - F_{X_2}^2(x_2)\right)\right\} \\
&= \frac{\partial}{\partial x_1}\left[f_{X_2}(x_2) F_{X_1}(x_1)\right. \\
&\quad \left. +\alpha f_{X_2}(x_2)\left(1 - 2F_{X_2}(x_2)\right)\left(F_{X_1}(x_1) - F_{X_1}^2(x_1)\right)\right] \\
&= f_{X_1}(x_1) f_{X_2}(x_2) + \alpha f_{X_1}(x_1) f_{X_2}(x_2)\left(1 - 2F_{X_1}(x_1)\right)\left(1 - 2F_{X_2}(x_2)\right) \\
&= f_{X_1}(x_1) f_{X_2}(x_2)\left[1 + \alpha\left(1 - 2F_{X_1}(x_1)\right)\left(1 - 2F_{X_2}(x_2)\right)\right]
\end{aligned}$$

The survival function of $(X_1, X_2)$ is given as

$$\begin{aligned}
S_{X_1,X_2}(x_1,x_2) &= \int_{x_1}^{\infty}\int_{x_2}^{\infty} f_{X_1,X_2}(t_1,t_2)\,dt_1dt_2 \qquad (2.33)\\
&= \int_{x_1}^{\infty}\int_{x_2}^{\infty} f_{X_1}(t_1)\,f_{X_2}(t_2)\left[1+\alpha\left(1-2F_{X_1}(t_1)\right)\left(1-2F_{X_2}(t_2)\right)\right]dt_1dt_2\\
&= \int_{x_1}^{\infty}\int_{x_2}^{\infty} f_{X_1}(t_1)\,f_{X_2}(t_2)\,dt_1dt_2\\
&\quad +\alpha\int_{x_1}^{\infty}\int_{x_2}^{\infty} f_{X_1}(t_1)\,f_{X_2}(t_2)\left(1-2F_{X_1}(t_1)\right)\left(1-2F_{X_2}(t_2)\right)dt_1dt_2\\
&= \int_{x_1}^{\infty} f_{X_1}(t_1)\left[\int_{x_2}^{\infty} f_{X_2}(t_2)\,dt_2\right]dt_1\\
&\quad +\alpha\int_{x_1}^{\infty} f_{X_1}(t_1)\left(1-2F_{X_1}(t_1)\right)\left[\int_{x_2}^{\infty} f_{X_2}(t_2)\left(1-2F_{X_2}(t_2)\right)dt_2\right]dt_1\\
&= S_{X_1}(x_1)\,S_{X_2}(x_2)\\
&\quad +\alpha\left[S_{X_1}(x_1)-F_{X_1}^2(t_1)\right]\left[S_{X_2}(x_2)-F_{X_2}^2(t_2)\right]
\end{aligned}$$

### 2.2.3 The Farlie-Gumbel-Morgenstern Family of BVWs

Let $X_i$, $i = 1, 2$, be a continuous random variable distributed as the two parameter Weibull defined in (1.7), i.e., $X_i \sim Weibull(\theta_i, \beta_i)$, $i = 1, 2$. Then the Farlie-Gumbel-Morgenstern Family of BVWs is defined by the joint cdf as

$$\begin{aligned}
F_{X_1,X_2}(x_1,x_2) &= F_{X_1}(x_1)\,F_{X_2}(x_2)\left[1+\alpha\left(1-F_{X_1}(x_1)\right)\left(1-F_{X_2}(x_2)\right)\right] \qquad (2.34)\\
&= \left[1-\exp\left(-\left(\frac{x_1}{\theta_1}\right)^{\beta_1}\right)\right]\left[1-\exp\left(-\left(\frac{x_2}{\theta_2}\right)^{\beta_2}\right)\right]\\
&\quad\times\left[1+\alpha\exp\left(-\left(\frac{x_1}{\theta_1}\right)^{\beta_1}-\left(\frac{x_2}{\theta_2}\right)^{\beta_2}\right)\right];
\end{aligned}$$

and the joint pdf is given as

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) &= f_{X_1}(x_1)\, f_{X_2}(x_2)\left[1+\alpha\left(1-2F_{X_1}(x_1)\right)\left(1-2F_{X_2}(x_2)\right)\right] \qquad (2.35)\\ &= \frac{\beta_1\beta_2}{\theta_1\theta_2}\left(\frac{x_1}{\theta_1}\right)^{\beta_1-1}\left(\frac{x_2}{\theta_2}\right)^{\beta_2-1}\exp\left[-\left(\frac{x_1}{\theta_1}\right)^{\beta_1}-\left(\frac{x_2}{\theta_2}\right)^{\beta_2}\right]\\ &\quad\times\left\{1+\alpha\left[2\exp\left(-\left(\frac{x_1}{\theta_1}\right)^{\beta_1}\right)-1\right]\left[2\exp\left(-\left(\frac{x_2}{\theta_2}\right)^{\beta_2}\right)-1\right]\right\}; \end{aligned}$$

and the joint survival function is given by

$$S_{X_1,X_2}(x_1,x_2) = S_{X_1}(x_1)\, S_{X_2}(x_2) + \alpha\left[S_{X_1}(x_1) - F_{X_1}^2(t_1)\right]\left[S_{X_2}(x_2) - F_{X_2}^2(t_2)\right] \qquad (2.36)$$

where the marginal survivor functions and cdfs are defined as in (1.9) and (1.7).

# Chapter 3

## Bivariate Location-Scale Weibull Lifetime Distributions

In this chapter, we firstly use the BVW defined in (2.30) as the standard pdf to generate a location-scale family of bivariate distributions through location-scale transformation. Then, by using the logarithm transform, we generated a location-scale Farlie-Gumbel-Morgenstern (FGM) Family of BVWs. Maximum likelihood estimates of the parameters and properties of the FGM BVW are also studied. Lastly, Bivariate Location-Scale Lifetime Distribution Regression Models are introduced and charted as future research extension.

The location-scale family of distributions have cdfs of the form $\Phi\left(\frac{x-\mu}{\sigma}\right)$, $-\infty < \mu < \infty$ and $\sigma > 0$. Many of the widely used statistical distributions belong to such a family of distributions. Examples of distributions that belong to the location-scale family are normal distribution, exponential distribution, double exponential distribution, Cauchy distribution, logistic distribution, and uniform distribution, etc. [34, Meeker and Escobar (1998)] emphasizes the importance of the widely used location-scale family with respect to its adaptivity and simplicity.

**Definition 2** *A group family of distributions is a family obtained by applying a suitable family of transformations to a random variable with a fixed distribution [29, Lehmann and Casella (1998)].*

**Definition 3** *Let $f(x)$ be a pdf. Then the family of pdfs $\frac{1}{\sigma}f(\frac{x-\mu}{\sigma})$, $-\infty < \mu < \infty$, $\sigma > 0$, is called the location-scale family with standard pdf $f(x)$; $\mu$ is called the location parameter and $\sigma$ is called the scale parameter. [2, Casella and Berger (2002)]*

The following three transformations result in three families of distributions, i.e., location families, scale families, and location-scale families. Examples of distributions that belong to the location-scale family are normal distribution, exponential distribution, double exponential distribution, Cauchy distribution, logistic distribution, and uniform distribution, etc.

Let $U$ be a random variable with a fixed distribution $F_U(u)$ with pdf $f_U(u)$ and let $\mu$ , the location parameter, and $\sigma > 0$ ,the scale parameter, be any given constants. Then the random variables $X = \mu + U$, $X = \sigma U$, and $X = \mu + \sigma U$ have distributions $F_X(x-\mu)$, $F_X(x/\sigma)$, and $F_X(\frac{x-\mu}{\sigma})$ with $f_X$'s equal to $f_U(x-\mu)$, $\frac{1}{\sigma}f_U(x/\sigma)$ and $\frac{1}{\sigma}f_U(\frac{x-\mu}{\sigma})$, which constitute a location family, a scale family, and a location-scale family, respectively.

[29, Lehmann and Casella (1998)] states that the families of transformations for the above location-scale family distributions are closed under composition and inversion.

Let the continuous random vector $\mathbf{U} = (U_1, ..., U_n)'$ have a fixed joint distribution $F_{\mathbf{U}'}(\mathbf{u}')$. The random vector $\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{\sigma}\mathbf{U}$, where $\boldsymbol{\mu}$ is an $n \times 1$ constant vector and $\boldsymbol{\sigma}$ is an $n \times n$ diagonal matrix of constants with the diagonal entry being $\sigma_{i,\ i=1,\ ...,\ n}$, and off-diagonal entries being zero. The marginal pdf of $X_i$ is $f_{X_i}(x) = \frac{1}{\sigma_i}f_{U_i}(\frac{x_i-\mu_i}{\sigma_i})$.

The joint pdf of the random vector $\mathbf{X'}=(X_1, ..., X_n)$ is given by

$$f_{\mathbf{X'}}\left(\mathbf{x'}\right) = \frac{1}{\prod_{i=1}^{n}\left(\sigma_i\right)} f_{\mathbf{U'}}\left(\frac{x_1-\mu_1}{\sigma_1}, ..., \frac{x_n-\mu_n}{\sigma_n}\right).$$

which defines a multivariate location-scale family of distributions. More specifically, when n=2, we have the bivariate case. The bivariate joint pdf of $\mathbf{X'}=(X_1, X_2)$ is given by

$$f_{X_1,X_2}\left(x_1, x_2\right) = \frac{1}{\sigma_1\sigma_2} f_{U_1,U_2}\left(\frac{x_1-\mu_1}{\sigma_1}, \frac{x_2-\mu_2}{\sigma_2}\right).$$

Since the bivariate location-scale transformations are one-to-one, the proof of the above proposition can be readily obtained using Jacobian of the transformation. Also it can be shown that the three properties of the bivariate distribution function reiterated in [23, Joe (1997)] are satisfied.

1. $\lim_{u_i\to\infty} S_{U_1,U_2}\left(u_1, u_2\right) = 0, \quad i = 1, 2;$

2. $\lim_{u_i\to\infty\forall i} F_{U_1,U_2}\left(u_1, u_2\right) = 1$; and

3. If $F_{U_1,U_2}$ has second-order derivatives, $\partial^2 F_{U_1,U_2}/\partial u_1 \partial u_2 \geq 0$ (the rectangle inequality).

A proof of the proposition for the specific case of BVW is given in the next section.

## 3.1 Bivariate Location-Scale Family Based on BVW with Random Hazards

Recall the BVW distribution defined by (2.30), its joint pdf is given by

$$\begin{aligned} f_{T_1,T_2}(t_1,t_2) &= \frac{(-1)^2 \partial^2 S_{T_1,T_2}(t_1,t_2)}{\partial t_1 \partial t_2} \\ &= \frac{\partial^2}{\partial t_1 \partial t_2} \exp\left\{-\left[\left(\frac{t_1}{\theta_1}\right)^{\beta_1/\alpha} + \left(\frac{t_2}{\theta_2}\right)^{\beta_2/\alpha}\right]^{\alpha}\right\} \\ &= \frac{\beta_1}{\theta_1}\frac{\beta_2}{\theta_2}\left(\frac{t_1}{\theta_1}\right)^{\frac{\beta_1}{\alpha}-1}\left(\frac{t_2}{\theta_2}\right)^{\frac{\beta_2}{\alpha}-1}\left[\left(\frac{t_1}{\theta_1}\right)^{\frac{\beta_1}{\alpha}} + \left(\frac{t_2}{\theta_2}\right)^{\frac{\beta_2}{\alpha}}\right]^{\alpha-2} \\ &\quad \times \left\{\left[\left(\frac{t_1}{\theta_1}\right)^{\beta_1/\alpha} + \left(\frac{t_2}{\theta_2}\right)^{\beta_2/\alpha}\right]^{\alpha} + \frac{1-\alpha}{\alpha}\right\} \\ &\quad \times \exp\left\{-\left[\left(\frac{t_1}{\theta_1}\right)^{\beta_1/\alpha} + \left(\frac{t_2}{\theta_2}\right)^{\beta_2/\alpha}\right]^{\alpha}\right\} \qquad 0 < \alpha \leq 1,\ 0 \leq t_1, t_2 < \infty. \end{aligned}$$

**Theorem 3.1** *Let the continuous random vector* $\mathbf{U} = (U_1, U_2)'$ *have a joint Bivariate Weibull distribution* $F_{U_1,U_2}(u_1, u_2)$ *as defined above, and define the transformation* $\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{\sigma}\mathbf{U}$*,where* $\boldsymbol{\mu}$ *is an* $2 \times 1$ *constant vector and* $\boldsymbol{\sigma}$ *is an* $2 \times 2$ *diagonal matrix of constants with the diagonal entry being* $\sigma_{i,\ i=1,\ 2}$*, and off-diagonal entries being zero. Then*

1. *The marginal pdf of* $X_i$ *is* $f_{X_i}(x) = \frac{1}{\sigma_i} f_{U_i}(\frac{x_i - \mu_i}{\sigma_i})$*;*

*2. The bivariate joint pdf of* $\mathbf{X} = (X_1, X_2)'$ *is given by*

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) &= \frac{1}{\sigma_1\sigma_2} f_{U_1,U_2}\left(\frac{x_1-\mu_1}{\sigma_1}, \frac{x_2-\mu_2}{\sigma_2}\right) \qquad (3.1) \\ &= \frac{\beta_1\beta_2}{\theta_1^*\theta_2^*}\left(\frac{x_1-\mu_1}{\theta_1^*}\right)^{\frac{\beta_1}{\alpha}-1}\left(\frac{x_2-\mu_2}{\theta_2^*}\right)^{\frac{\beta_2}{\alpha}-1} \\ &\quad \times\left[\left(\frac{x_1-\mu_1}{\theta_1^*}\right)^{\frac{\beta_1}{\alpha}}+\left(\frac{x_2-\mu_2}{\theta_2^*}\right)^{\frac{\beta_2}{\alpha}}\right]^{\alpha-2} \\ &\quad \times\left\{\left[\left(\frac{x_1-\mu_1}{\theta_1^*}\right)^{\frac{\beta_1}{\alpha}}+\left(\frac{x_2-\mu_2}{\theta_2^*}\right)^{\frac{\beta_2}{\alpha}}\right]^{\alpha}+\frac{1-\alpha}{\alpha}\right\} \\ &\quad \times\exp\left\{-\left[\left(\frac{x_1-\mu_1}{\theta_1^*}\right)^{\frac{\beta_1}{\alpha}}+\left(\frac{x_2-\mu_2}{\theta_2^*}\right)^{\frac{\beta_2}{\alpha}}\right]^{\alpha}\right\} \\ \text{where } \theta_1^* &= \theta_1\sigma_1,\ \theta_2^* = \theta_2\sigma_2,\ \text{and } 0<\alpha\le 1,\ 0\le x_1,x_2<\infty. \end{aligned}$$

*The location parameter vector is* $(\mu_1, \mu_2)$*, and the scale parameter vector is* $(\theta_1^* = \theta_1\sigma_1,\ \theta_2^* = \theta_2\sigma_2)$*.*

**Proof.** Assuming some regularity conditions as listed in [2, Casella and Berger (2002)], we have the new bivariate random vector $(X_1, X_2)$ defined by bivariate transformation

$$\mathbf{g} = \begin{pmatrix} X_1 = g_1(U_1,U_2) = \mu_1 + \sigma_1 U_1 \\ X_2 = g_2(U_1,U_2) = \mu_2 + \sigma_2 U_2 \end{pmatrix}',$$

Denote the inverse bivariate transformation by

$$\mathbf{h} = \begin{pmatrix} U_1 = h_1\left(X_1, X_2\right) = \frac{X_1 - \mu_1}{\sigma_1} \\ U_2 = h_2\left(X_1, X_2\right) = \frac{X_2 - \mu_2}{\sigma_2} \end{pmatrix}'.$$

Obviously the transformations are one-to-one, then by 4.3.2 of Casella and Berger (2002), we have

$$\begin{aligned} f_{X_1,X_2}\left(x_1, x_2\right) &= f_{U_1,U_2}\left[h_1\left(x_1, x_2\right), h_2\left(x_1, x_2\right)\right] \begin{vmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x 2} \end{vmatrix} \\ &= f_{U_1,U_2}\left(\frac{x_1 - \mu_1}{\sigma_1}, \frac{x_2 - \mu_2}{\sigma_2}\right) \begin{vmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{vmatrix} \\ &= \frac{1}{\sigma_1 \sigma_2} f_{U_1,U_2}\left(\frac{x_1 - \mu_1}{\sigma_1}, \frac{x_2 - \mu_2}{\sigma_2}\right) \end{aligned}$$

■

The survival function of $(X_1, X_2)$ is given as

$$\begin{aligned}
S_{X_1,X_2}(x_1,x_2) &= \int_{x_1}^{\infty}\int_{x_2}^{\infty} f_{X_1,X_2}(t_1,t_2)\,dt_1dt_2 \qquad (3.2)\\
&= \int_{x_1}^{\infty}\int_{x_2}^{\infty} \frac{1}{\sigma_1\sigma_2} f_{U_1,U_2}\left(\frac{t_1-\mu_1}{\sigma_1},\frac{t_2-\mu_2}{\sigma_2}\right) dt_1dt_2\\
&= S_{U_1,U_2}\left(\frac{x_1-\mu_1}{\sigma_1},\frac{x_2-\mu_2}{\sigma_2}\right)\\
&= \exp\left\{-\left[\left(\frac{x_1-\mu_1}{\theta_1\sigma_1}\right)^{\beta_1/\alpha}+\left(\frac{x_2-\mu_2}{\theta_2^*}\right)^{\beta_2/\alpha}\right]^{\alpha}\right\}\\
&= \exp\left\{-\left[\left(\frac{x_1-\mu_1}{\theta_1^*}\right)^{\beta_1/\alpha}+\left(\frac{x_2-\mu_2}{\theta_2^*}\right)^{\beta_2/\alpha}\right]^{\alpha}\right\} \qquad \theta_i^*=\theta_i\sigma_i
\end{aligned}$$

## 3.2 Maximum Likelihood Estimates of the Bivariate Location-Scale Family Based on BVW with Random Hazards

### 3.2.1 Likelihood Functions of Uncensored Lifetime Data

Let the bivariate lifetimes, $(X_{1i}, X_{2i})$, of a random sample of size $n$ without censoring have continuous joint survivor function (3.2). The bivariate likelihood function is given by

$$L = \prod_{i=1}^{n} f_{X_1,X_2}(x_{1i},x_{2i}), \qquad (3.3)$$

and the bivariate loglikelihood is

$$
\begin{aligned}
l \;=\;& n \log\left(\frac{\beta_1}{\theta_1^*}\right) + n \log\left(\frac{\beta_2}{\theta_2^*}\right) \qquad (3.4)\\
&+ \left(\frac{\beta_1}{\alpha} - 1\right) \sum_{i=1}^{n} \log\left(\frac{x_{1i} - \mu_1}{\theta_1^*}\right) + \left(\frac{\beta_2}{\alpha} - 1\right) \sum_{i=1}^{n} \log\left(\frac{x_{2i} - \mu_2}{\theta_2^*}\right)^{\frac{\beta_2}{\alpha} - 1}\\
&+ (\alpha - 2) \sum_{i=1}^{n} \log\left[\left(\frac{x_{1i} - \mu_1}{\theta_1^*}\right)^{\frac{\beta_1}{\alpha}} + \left(\frac{x_{2i} - \mu_2}{\theta_2^*}\right)^{\frac{\beta_2}{\alpha}}\right]\\
&+ \sum_{i=1}^{n} \log\left\{\left[\left(\frac{x_{1i} - \mu_1}{\theta_1^*}\right)^{\beta_1/\alpha} + \left(\frac{x_{2i} - \mu_2}{\theta_2^*}\right)^{\beta_2/\alpha}\right]^{\alpha} + \frac{1 - \alpha}{\alpha}\right\}\\
&- \left[\left(\frac{x_{1i} - \mu_1}{\theta_1^*}\right)^{\beta_1/\alpha} + \left(\frac{x_{2i} - \mu_2}{\theta_2^*}\right)^{\beta_2/\alpha}\right]^{\alpha}.
\end{aligned}
$$

### 3.2.2 Likelihood Functions for Right Censored Lifetime Data

Let the bivariate lifetimes, $(X_{1i}, X_{2i})$, of a random sample of size $n$ with right censoring have continuous joint survivor function 3.2. Assume that censoring times $(C_{1i}, C_{2i})$are independent of $(X_{1i}, X_{2i})$, and let the censoring indicator $\delta_{ji} = I\left[X_{ji} = \min\left(X_{ji},\; C_{ji}\right)\right]$, $j = 1,\; 2$. Then the bivariate likelihood function takes the form as given in [26, Lawless (2002)]

$$
\begin{aligned}
L \;=\;& \prod_{i=1}^{n} f_{X_1,X_2}\left(x_{1i}, x_{2i}\right)^{\delta_{1i}\delta_{2i}} \left[\frac{-\partial S_{X_1,X_2}\left(x_{1i}, x_{2i}\right)}{\partial x_{1i}}\right]^{\delta_{1i}(1-\delta_{2i})} \qquad (3.5)\\
&\times \left[\frac{-\partial S_{X_1,X_2}\left(x_{1i}, x_{2i}\right)}{\partial x_{2i}}\right]^{\delta_{2i}(1-\delta_{1i})} S_{X_1,X_2}\left(x_{1i}, x_{2i}\right)^{(1-\delta_{1i})(1-\delta_{2i})}\\
=\;& \prod_{i=1}^{n} f_{X_1,X_2}\left(x_{1i}, x_{2i}\right)^{\delta_{1i}\delta_{2i}} f_{X_1}\left(x_{1i}\right)^{\delta_{2i}(1-\delta_{1i})} f_{X_2}\left(x_{2i}\right)^{\delta_{1i}(1-\delta_{2i})} S_{X_1,X_2}\left(x_{1i}, x_{2i}\right)^{(1-\delta_{1i})(1-\delta_{2i})}
\end{aligned}
$$

where the marginal pdfs are given as

$$
\begin{aligned}
f_{X_j}(x_{ji}) &= \frac{-\partial S_{X_1,X_2}(x_{1i},x_{2i})}{\partial x_{ji}} \qquad (3.6)\\
&= \exp\left\{-\left[\left(\frac{x_{1i}-\mu_1}{\theta_1^*}\right)^{\beta_1/\alpha}+\left(\frac{x_{2i}-\mu_2}{\theta_2^*}\right)^{\beta_2/\alpha}\right]^{\alpha}\right\}\\
&\times\left[\left(\frac{x_{1i}-\mu_1}{\theta_1^*}\right)^{\beta_1/\alpha}+\left(\frac{x_{2i}-\mu_2}{\theta_2^*}\right)^{\beta_2/\alpha}\right]^{\alpha-1}\\
&\times\frac{\beta_j}{\theta_j^*}\left(\frac{x_{ji}-\mu_j}{\theta_j^*}\right)^{\frac{\beta_j}{\alpha}-1}
\end{aligned}
$$

The bivariate loglikelihood is given as

$$
\begin{aligned}
l &= \sum_{i=1}^{n}[\delta_{1i}\delta_{2i}\log f_{X_1,X_2}(x_{1i},x_{2i})+\delta_{2i}(1-\delta_{1i})\log f_{X_1}(x_{1i}) \qquad (3.7)\\
&\quad +\delta_{1i}(1-\delta_{2i})\log f_{X_2}(x_{2i})+(1-\delta_{1i})(1-\delta_{2i})\log S_{X_1,X_2}(x_{1i},x_{2i})]
\end{aligned}
$$

where

$$
\log S_{X_1,X_2}(x_{1i},x_{2i}) = -\left[\left(\frac{x_{1i}-\mu_1}{\theta_1^*}\right)^{\beta_1/\alpha}+\left(\frac{x_{2i}-\mu_2}{\theta_2^*}\right)^{\beta_2/\alpha}\right]^{\alpha},
$$

$$\log f_{X_1,X_2}(x_{1i}, x_{2i}) = \log\left(\frac{\beta_1\beta_2}{\theta_1^*\theta_2^*}\right) + \left(\frac{\beta_1}{\alpha} - 1\right)\log\left(\frac{x_{1i}-\mu_1}{\theta_1^*}\right) + \left(\frac{\beta_2}{\alpha} - 1\right)\log\left(\frac{x_{2i}-\mu_2}{\theta_2^*}\right)$$
$$+ (\alpha - 2)\log\left[\left(\frac{x_1-\mu_1}{\theta_1^*}\right)^{\frac{\beta_1}{\alpha}} + \left(\frac{x_2-\mu_2}{\theta_2^*}\right)^{\frac{\beta_2}{\alpha}}\right]$$
$$+ \log\left\{\left[\left(\frac{x_1-\mu_1}{\theta_1^*}\right)^{\frac{\beta_1}{\alpha}} + \left(\frac{x_2-\mu_2}{\theta_2^*}\right)^{\frac{\beta_2}{\alpha}}\right]^{\alpha} + \frac{1-\alpha}{\alpha}\right\}$$
$$- \left[\left(\frac{x_1-\mu_1}{\theta_1^*}\right)^{\frac{\beta_1}{\alpha}} + \left(\frac{x_2-\mu_2}{\theta_2^*}\right)^{\frac{\beta_2}{\alpha}}\right]^{\alpha},$$

and

$$\log f_{X_j}(x_{ji}) = -\left[\left(\frac{x_{1i}-\mu_1}{\theta_1^*}\right)^{\beta_1/\alpha} + \left(\frac{x_{2i}-\mu_2}{\theta_2^*}\right)^{\beta_2/\alpha}\right]^{\alpha}$$
$$+ (\alpha - 1)\log\left[\left(\frac{x_{1i}-\mu_1}{\theta_1^*}\right)^{\beta_1/\alpha} + \left(\frac{x_{2i}-\mu_2}{\theta_2^*}\right)^{\beta_2/\alpha}\right]$$
$$+ \log\left(\frac{\beta_j}{\theta_j^*}\right) + \left(\frac{\beta_j}{\alpha} - 1\right)\log\left(\frac{x_{ji}-\mu_j}{\theta_j^*}\right).$$

## 3.3 Location-Scale Family of BVWs Based on the Farlie-Gumbel-Morgenstern Family

The FGM BVWs as defined in section 2.2.3 are not of location-scale form, however, by exploring the relationship between a two parameter Weibull distribution and a smallest extreme value (SEV) distribution, the Weibull can be transformed into a location-scale form as discussed in chapter four of [34, Meeker and Escobar (1998)].

If random variable $X \sim Weibull\left(\theta, \beta\right)$, then $Y = Log\left(X\right) \sim SEV\left(\mu, \sigma\right)$,where $\mu = \log\left(\theta\right)$ is the location parameter and $\sigma = 1/\beta$ is the scale parameter of the SEV.

Then the Weibull cdf, pdf, survival and hazard function can be written as

$$\begin{aligned} F_X\left(x\right) &= F_Y\left(\frac{\log\left(x\right) - \mu}{\sigma}\right) \\ &= 1 - \exp\left[-\exp\left(\frac{\log\left(x\right) - \mu}{\sigma}\right)\right]; \end{aligned} \tag{3.8}$$

$$\begin{aligned} f_X\left(x\right) &= \frac{1}{\sigma x} f_Y\left(\frac{\log\left(x\right) - \mu}{\sigma}\right) \\ &= \frac{1}{\sigma x}\exp\left[\frac{\log\left(x\right) - \mu}{\sigma} - \exp\left(\frac{\log\left(x\right) - \mu}{\sigma}\right)\right]; \end{aligned} \tag{3.9}$$

$$\begin{aligned} S_X\left(x\right) &= S_Y\left(\frac{\log\left(x\right) - \mu}{\sigma}\right) \\ &= \exp\left[-\exp\left(\frac{\log\left(x\right) - \mu}{\sigma}\right)\right]; \end{aligned} \tag{3.10}$$

$$h_X\left(x\right) = \frac{1}{\sigma}\exp\left(\frac{\log\left(x\right) - \mu}{\sigma}\right). \tag{3.11}$$

Thus the FGM BVW cdf , pdf and survivor function in location-scale form are given by

$$\begin{aligned} F_{X_1,X_2}(x_1,x_2) &= F_{Y_1}\left(\frac{\log(x_1)-\mu_1}{\sigma_1}\right)F_{Y_2}\left(\frac{\log(x_2)-\mu_2}{\sigma_2}\right) \\ &\quad\times\left\{1+\alpha\left[1-F_{Y_1}\left(\frac{\log(x_1)-\mu_1}{\sigma_1}\right)\right]\left[1-F_{Y_2}\left(\frac{\log(x_2)-\mu_2}{\sigma_2}\right)\right]\right\} \\ &= \left\{1-\exp\left[-\exp\left(\frac{\log(x_1)-\mu}{\sigma}\right)\right]\right\}\left\{1-\exp\left[-\exp\left(\frac{\log(x_2)-\mu}{\sigma}\right)\right]\right\} \\ &\quad\times\left\{1+\alpha\exp\left[-\exp\left(\frac{\log(x_1)-\mu}{\sigma}\right)-\exp\left(\frac{\log(x_2)-\mu}{\sigma}\right)\right]\right\}; \end{aligned} \tag{3.12}$$

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) &= \frac{1}{\sigma_1\sigma_2x_1x_2}f_{Y_1}\left(\frac{\log(x_1)-\mu_1}{\sigma_1}\right)f_{Y_2}\left(\frac{\log(x_2)-\mu_2}{\sigma_2}\right) \\ &\quad\times\left[1+\alpha\left(1-2F_{Y_1}\left(\frac{\log(x_1)-\mu_1}{\sigma_1}\right)\right)\left(1-2F_{Y_2}\left(\frac{\log(x_2)-\mu_2}{\sigma_2}\right)\right)\right] \\ &= \frac{1}{\sigma_1\sigma_2x_1x_2}\times\exp\left[\frac{\log(x_1)-\mu_1}{\sigma_1}+\frac{\log(x_2)-\mu_2}{\sigma_2}\right. \\ &\quad\left.-\exp\left(\frac{\log(x_1)-\mu_1}{\sigma_1}\right)-\exp\left(\frac{\log(x_2)-\mu}{\sigma}\right)\right] \\ &\quad\times\left[1+\alpha\left(1-2F_{Y_1}\left(\frac{\log(x_1)-\mu_1}{\sigma_1}\right)\right)\left(1-2F_{Y_2}\left(\frac{\log(x_2)-\mu_2}{\sigma_2}\right)\right)\right] \end{aligned} \tag{3.13}$$

where $F_{Y_i}\left(\frac{\log(x_i)-\mu_i}{\sigma_i}\right)=1-\exp\left[-\exp\left(\frac{\log(x_i)-\mu_i}{\sigma_i}\right)\right]$, $i=1,\ 2.$

$$
\begin{aligned}
S_{X_1,X_2}(x_1,x_2) &= S_{X_1}(x_1)\,S_{X_2}(x_2) \\
&\quad +\alpha\left[S_{X_1}(x_1)-F_{X_1}^2(t_1)\right]\left[S_{X_2}(x_2)-F_{X_2}^2(t_2)\right] \\
&= \exp\left[-\exp\left(\frac{\log(x_1)-\mu_1}{\sigma_1}\right)-\exp\left(\frac{\log(x_2)-\mu_1}{\sigma_1}\right)\right] \\
&\quad +\alpha\left[e^{-\exp\frac{\log(x_1)-\mu_1}{\sigma_1}}-\left(1-e^{-\exp\frac{\log(x_1)-\mu_1}{\sigma_1}}\right)^2\right] \\
&\quad \times\left[e^{-\exp\frac{\log(x_2)-\mu_2}{\sigma_2}}-\left(1-e^{-\exp\frac{\log(x_2)-\mu_2}{\sigma_2}}\right)^2\right]
\end{aligned} \tag{3.14}
$$

## 3.4 Maximum Likelihood Estimates of the FGM BVWs

In this section, maximum likelihood method is applied to the FGM BVW distribution with two-parameter Weibull marginal. The two-parameter marginal are chosen not out of necessity, but of convenience since the location-scale parametrization of the FGM BVW has two-parameter Weibll marginal. Nevertheless all the following results apply to the FGM BVW with three-parameter Weibull marginal readily.

### 3.4.1 Likelihood Functions of Uncensored Lifetime Data

Let the bivariate lifetimes, $(X_{1i}, X_{2i})$, of a random sample of size $n$ without censoring have FGM BVW distribution as defined in (2.34). The bivariate likelihood function is defined in (3.3), and the bivariate loglikelihood is

$$
\begin{aligned}
l &= n\log\left(\frac{\beta_1\beta_2}{\theta_1\theta_2}\right) \qquad (3.15)\\
&+\sum_{i=1}^{n}\left[(\beta_1-1)\log\left(\frac{x_{1i}}{\theta_1}\right)+(\beta_2-1)\log\left(\frac{x_{2i}}{\theta_2}\right)-\left(\frac{x_{1i}}{\theta_1}\right)^{\beta_1}-\left(\frac{x_{2i}}{\theta_2}\right)^{\beta_2}\right]\\
&+\sum_{i=1}^{n}\left\{\log\left[1+\alpha\left(2\exp\left(-\left(\frac{x_{1i}}{\theta_1}\right)^{\beta_1}\right)-1\right)\left(2\exp\left(-\left(\frac{x_{2i}}{\theta_2}\right)^{\beta_2}\right)-1\right)\right]\right\}.
\end{aligned}
$$

For location-scale FGM BVW as defined in (3.12), the bivariate loglikelihood is

$$
\begin{aligned}
l &= -n\log(\sigma_1\sigma_2)-\sum_{i=1}^{n}\log(x_1x_2)\\
&+\frac{\log(x_1)-\mu_1}{\sigma_1}+\frac{\log(x_2)-\mu_2}{\sigma_2}-\exp\left(\frac{\log(x_1)-\mu_1}{\sigma_1}\right)-\exp\left(\frac{\log(x_2)-\mu}{\sigma}\right)\\
&+\log\left[1+\alpha\left(1-2F_{Y_1}\left(\frac{\log(x_1)-\mu_1}{\sigma_1}\right)\right)\left(1-2F_{Y_2}\left(\frac{\log(x_2)-\mu_2}{\sigma_2}\right)\right)\right],
\end{aligned}
$$

where $F_{Y_i}\left(\frac{\log(x_i)-\mu_i}{\sigma_i}\right)=1-\exp\left[-\exp\left(\frac{\log(x_i)-\mu_i}{\sigma_i}\right)\right]$, $i=1,\ 2$.

### 3.4.2 Likelihood Functions for Right Censored Lifetime Data

Assume that censoring times $(C_{1i}, C_{2i})$are independent of $(X_{1i}, X_{2i})$, and let the censoring indicator $\delta_{ji}=I\left[X_{ji}=\min\left(X_{ji},\ C_{ji}\right)\right]$, $j=1,\ 2$. The bivariate likelihood function is given by (3.5) where the marginal pdfs are given by (1.6), and the bivariate loglikelihood is given by (3.7), where the joint survival function and the joint pdf are defined by (2.36) and (2.35) for FGM BVW, and by (3.14) and (3.13) for the location-scale parametrization of the FGM BVW.

### 3.4.3 Optimization Procedures for MLEs of the FGM BVW

Let the vector of parameters $\boldsymbol{\xi} = (\alpha, \beta_1, \beta_2, \theta_1, \theta_2)'$ be in a parameter space $\Omega$. The maximum likelihood method is used to maximize the log-likelihood function $l(\boldsymbol{\xi})$, and the corresponding vector of parameters, $\hat{\boldsymbol{\xi}} = \left(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\theta}_1, \hat{\theta}_2\right)'$, is call the vector of mles. If the likelihood function has a unique maximum in the parameter space $\Omega$, then the mle vector $\hat{\boldsymbol{\xi}}$ can be found by solving $\mathbf{U}(\boldsymbol{\xi}) = \partial l(\boldsymbol{\xi}) / \partial \boldsymbol{\xi} = \mathbf{0}$, which are called score functions. For Weibull distribution, however, there is no closed form solutions for the score functions. [26, Lawless (2002)] summarizes numerical methods of solving the optimization problem, which are distinguished by their use of the first and second derivatives of the logarithm of the likelihood function. The methods include:

1. Search algorithm or heuristics without utilizing any derivatives.

2. Methods that utilize only the first derivatives.

3. Methods that utilize both the first and second derivatives. Moreover, the second derivative matrix (or the Hessian Matrix) is defined as $H(\boldsymbol{\xi}) = \partial^2 l(\boldsymbol{\xi}) / \partial \boldsymbol{\xi} \partial \boldsymbol{\xi}'$.

The SAS procedure NLP provides all three types of methods for the optimization. The Newton-Raphson Method with Line Search (NEWRAP), which is of type three method above, is the method of choice in the optimization procedures in the simulation study for this thesis.

The NEWRAP technique uses the first derivative vector $\mathbf{U}\left(\boldsymbol{\xi}^k\right)$ and the Hessian matrix $H\left(\boldsymbol{\xi}^k\right)$ in its iterations and requires that the logarithm of the likelihood function have continuous first- and second-order derivatives inside the parameter space

$\Omega$. The NEWRAP method is a relatively efficient algorithm for medium to large problems since it does not need many function, gradient, and Hessian calls. Nevertheless the SAS derivative compiler is not efficient in the computation of second-order derivatives, and a complicated Hessian does affect the algorithm's efficiency. The algorithm also requires a positive definite Hessian. When the Hessian is not positive definite, a multiple of the identity matrix is added to the Hessian matrix to make it positive definite [11, Eskow and Schnabel 1991]. The default line-search method uses quadratic interpolation and cubic extrapolation in each iteration to compute an approximate optimum of the objective function.

Similar to that described for the FGM BVW's, the optimization procedures for MLEs of the bivariate location-scale family consist of three major type of numerical methods. Again, the Newton-Raphson Method is used in the simulation study. Details of the simulation study can be found in the appendix.

### 3.5 Bivariate Location-Scale Lifetime Distribution Regression Models

The location-scale family includes many important distributions. Bivariate regression models can improve estimation of marginal covariate effects when two or more response variables are correlated(e.g. [41, Zellner (1962)]).

[19, He and Lawless (2005)] considers bivariate location-scale regression models: Let the true bivariate distribution of response variables $Y_1$ and $Y_2$ be given by

$$F\left(y_1, y_2 | \mathbf{x}_1, \mathbf{x}_2\right) = H_\psi\left(\omega_1, \omega_2\right) = H_\psi\left(\frac{y_1 - \mu_{10} - \mathbf{x}_1'\boldsymbol{\mu}_1}{\tau_1}, \frac{y_2 - \mu_{20} - \mathbf{x}_2'\boldsymbol{\mu}_2}{\tau_2}\right), \quad (3.16)$$

where $\mathbf{x}_1,\ \mathbf{x}_2$ are $p \times 1$ and $q \times 1$ covariate vectors, $H_\psi(\omega_1, \omega_2)$ is a bivariate cdf specified with an association parameter $\psi$.

The location-scale form of the true marginal distributions of $Y_1$ and $Y_2$ can be written as

$$Y_j = \mu_{j0} + \mathbf{x}_j'\boldsymbol{\mu}_j + \tau_j\omega_j, \qquad j = 1, 2 \tag{3.17}$$

where the distribution of the error $\omega_j$ is independent of that of the covariate's. $\mu_{j0} + \mathbf{x}_j'\boldsymbol{\mu}_j$ is the location parameter and $\tau_j$ is the scale parameter.

[19, He and Lawless (2005)] proposes a location-scale regression model to investigate distribution misspecification effects on the estimation of regression coefficients. The regression model is given as

$$\begin{aligned} Y_1 &= \beta_{10} + \mathbf{x}_1'\boldsymbol{\beta}_1 + \sigma_1\varepsilon_1 \\ Y_2 &= \beta_{20} + \mathbf{x}_2'\boldsymbol{\beta}_2 + \sigma_2\varepsilon_2 \end{aligned} \tag{3.18}$$

where $\beta_{j0} + \mathbf{x}_j'\boldsymbol{\beta}_j$ is the location parameter and $\sigma_j > 0$ is the scale parameter, $\boldsymbol{\beta}_j$ is the corresponding regression coefficient vector, and the errors $\varepsilon_1$, $\varepsilon_2$ have a joint distribution specified by a copula function $C_\phi$ with an association parameter $\phi$.

[19, He and Lawless (2005)] shows that estimators of the regression coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ are consistent estimators and robust to misspecification of the marginal distributions of the errors. Also, they examine the relative efficiency for using the bivariate model to estimate $\mu_1$ and $\mu_2$ compared with using the marginal distributions.

In their simulation study, a specific bivariate location-scale regression model as proposed in [5, Clayton (1978)] is studied. The model is defined by the joint survivor distribution

$$H_\phi\left(\varepsilon_1, \varepsilon_2\right) = \left[S_1\left(\varepsilon_1\right)^{-1/\phi} + S_2\left(\varepsilon_2\right)^{-1/\phi} - 1\right]^{-\phi}, \qquad \phi > 0 \tag{3.19}$$

where the survivor functions $S_1\left(\varepsilon_1\right)$ and $S_2\left(\varepsilon_2\right)$ define the location-scale marginal distributions of $Y_j = \beta_{j0} + \mathbf{x}_j'\boldsymbol{\beta}_j + \sigma_j\varepsilon_j$.

Chapter 4

Simulation Study

The simulation study is focused on maximum likelihood estimation of several bivariate models. Bivariate data generation, maximum likelihood estimation and statistical properties are studied, and validation examples are provided.

## 4.1 Linearly Associated Bivariate Exponential and Weibull Models

### 4.1.1 Bivariate Data Generation

To generate the linearly associated bivariate data set, $(X_{i1}, X_{i2}), i = 1, ..., n$, of BVE$(\lambda_1, \lambda_2, a)$ as defined in section 2.1.1, we utilize the fact that $X_1$ and $Z$ are independent of each other and generate the two random variables first. $X_1 \sim \exp(\lambda_1)$ is generated with the SAS exponential random number generator. Since $Z$ is the product of of a Bernoulli random variable with $P(Z = 0) = a\lambda_2/\lambda_1$ and a continuous random variable having the same distribution as $X_2 \sim \exp(\lambda_2)$, it is generated using the SAS Bernoulli random number generator and the SAS exponential random number generator. By the linear association defined in 2.4, $X_2$ is then generated as $aX_1 + Z$.

The bivariate data set of BVW$(\lambda_1, \lambda_2, \beta, a)$ as defined in section 2.1.2 can be readily obtained by a power transform of the BVE data set, i.e., letting $Y_{ij} = X_{ij}^{1/\beta}$, $j = 1,\ 2$.

### 4.1.2 Simulation Results

Table 4.1 summarizes the results of a simulation study where 10,000 simulated samples of size 25 pairs from a BVE$(1, 1, a)$ were generated.

Several other simulations were conducted using various combinations of $\lambda_1$, $\lambda_2$ and $a$ and similar results were found as given in Table 1. Note that since $\lambda_1 = \lambda_2 = 1$, Corr$(X_1, X_2) = a\lambda_2/\lambda_1 = a$. The empirical MSE was computed for both $\hat{\lambda}_1$, the estimator based on the joint likelihood given in (25), and $\hat{\lambda}_1^*$, the usual maximum likelihood estimator based on the marginal distribution. Also computed, were the maximum likelihood estimators of the $\rho$, given as $\hat{\rho} = a\hat{\lambda}_2/\hat{\lambda}_1$ and $\hat{\rho}^* = a\hat{\lambda}_2^*/\hat{\lambda}_1^*$. Percent MSE improvement was computed as

$$(MSE(\hat{\theta}_1^*) - MSE(\hat{\theta}_1))/MSE(\hat{\theta}_1^*) \cdot 100\%.$$

The joint MLE estimator $\lambda$ gave MSE improvement over the marginal MLE for all values of $\rho$. Interestingly, percent improvement is a concave function of $\rho$, with maximum occurring at $\rho = 0.5$, giving over 25%. The joint MLE for the correlation coefficient gives monotonically increasing percent improvement over the estimator based on the marginal MLE's, with 44% improvement when $\rho = 0.99$.. Therefore, ignoring the multivariate relationship between $X_1$ and $X_2$ comes at a significant cost with respect to MSE.

Similarly, Table 4.2 summarizes the results of a simulation study where 10,000 simulated samples of size 25 pairs from a BVW(1,1,$\beta$,a), $\beta = 0.5, 1, 1.5, 2.0$ were generated. The empirical MSE was computed for both $\hat{\lambda}_1$, the estimator based on the joint likelihood given in (25), and $\hat{\lambda}_1^*$, the usual maximum likelihood estimator based on the marginal distribution. Similar patterns of MSE improvement emerge for the bivariate Weibull as in the bivariate exponential case, for all values of $\beta$.

## 4.2 Bivariate Location-Scale Models

### 4.2.1 Data Generation for BVW of Hougaard's Model

[27, Lee (1979)] and [31, Lu and Bhattacharyya (1990)] show that $(X_1, X_2)$ of BVW defined by 2.30 can be represented by two independent random variables $(U,\ V)$ as

$$X_1 = U^{\delta/\beta_1} V^{1/\beta_1} \theta_1, \quad X_2 = (1-U)^{\delta/\beta_2} V^{1/\beta_2} \theta_2,$$

where $U \sim Uniform\,(0,1)$, and $V$ is distributed as the mixture of a standard exponential and standard Gamma(2). The pdf of $V$ is given by

$$f\,(v) = \delta v \exp\left(-v\right) + \left(1-\delta\right) \exp\left(-v\right), \qquad v > 0.$$

So we start by generating $(U,\ V)$. $U$ is generated by SAS uniform random number generator. $V$ is obtained by generating other four standard uniform random variables $u_1, ...u_4$, and using the logarithm transform as on page 248 of [2, Casella and Berger

| | Mean Squared Error (MSE) | | | | | |
|---|---|---|---|---|---|---|
| $\rho$ | $\hat{\lambda}_1^*$ | $\hat{\lambda}_1$ | %-imp | $\hat{\rho}_1^*$ | $\hat{\rho}_1$ | %-imp |
| 0.01 | 0.04585 | 0.04621 | 0.782 | 0.00001 | 0.00001 | 0.529 |
| 0.05 | 0.04522 | 0.04723 | 4.251 | 0.00021 | 0.00022 | 2.990 |
| 0.10 | 0.04512 | 0.05005 | 9.849 | 0.00077 | 0.00082 | 6.410 |
| 0.20 | 0.04118 | 0.04877 | 15.570 | 0.00252 | 0.00284 | 11.243 |
| 0.30 | 0.03930 | 0.04937 | 20.406 | 0.00456 | 0.00546 | 16.563 |
| 0.40 | 0.03627 | 0.04749 | 23.615 | 0.00634 | 0.00815 | 22.176 |
| 0.50 | 0.03757 | 0.04834 | 22.293 | 0.00741 | 0.01011 | 26.710 |
| 0.60 | 0.03771 | 0.04982 | 24.307 | 0.00807 | 0.01169 | 31.014 |
| 0.70 | 0.03799 | 0.04713 | 19.384 | 0.00763 | 0.01131 | 32.517 |
| 0.80 | 0.04210 | 0.04913 | 14.311 | 0.00611 | 0.00966 | 36.764 |
| 0.90 | 0.04326 | 0.04675 | 7.465 | 0.00343 | 0.00574 | 40.287 |
| 0.99 | 0.04939 | 0.04978 | 0.796 | 0.00039 | 0.00067 | 42.560 |

Based on 10,000 simulated samples

Table 4.1: Simulation Study for BVE($\lambda_1 = \lambda_2 = 1$, $n = 25$)

| | Percent-Improvement in MSE | | | | |
|---|---|---|---|---|---|
| a | $\beta = 0.5$ | $\beta = 1$ | $\beta = 1.5$ | $\beta = 2$ | $\beta = 10$ |
| 0.01 | 1.864 | 0.152 | 0.141 | 0.233 | 0.490 |
| 0.05 | 4.904 | 4.836 | 4.568 | 3.986 | 1.973 |
| 0.1 | 5.987 | 5.378 | 2.763 | 4.352 | 5.882 |
| 0.2 | 4.285 | 5.021 | 7.807 | 4.352 | 10.089 |
| 0.3 | 9.974 | 9.035 | 12.566 | 5.988 | 6.809 |
| 0.4 | 9.408 | 15.852 | 20.258 | 10.532 | 11.684 |
| 0.5 | 15.240 | 12.699 | 15.718 | 18.472 | 18.161 |
| 0.6 | 16.083 | 10.914 | 16.048 | 13.707 | 14.489 |
| 0.7 | 7.228 | 12.788 | 12.214 | 13.676 | 12.571 |
| 0.8 | 9.946 | 10.300 | 9.967 | 10.299 | 9.189 |
| 0.9 | 4.922 | 6.716 | 4.595 | 5.940 | 6.174 |
| 0.99 | 0.367 | 0.478 | 0.590 | 1.118 | 1.499 |

Based on 10,000 simulated samples of size $n = 25$.

Table 4.2: Simulation Study BVW($\lambda_1 = \lambda_2 = 1$)

(2002)] such that

$$V = \begin{cases} -\ln u_1 - \ln u_2, & if\ u_4 \leq \delta \\ -\ln u_4, & if\ u_4 > \delta \end{cases}$$

where $(-\ln u_1 - \ln u_2) \sim gamma\,(2)$, and $-\ln u_4 \sim \exp(1)$

### 4.2.2 Data Generation for BVW of FGM Model (Sequential Monte Carlo Simulation)

To generate bivariate data set, $(X_{i1}, X_{i2}), i = 1, ..., n$, of the FGM BVW as defined in 2.34, we first generate $X_1 \sim Weibull(\theta_1, \beta_1)$ by setting its marginal CDF, $F_{X_1}$, equal to a random number of standard Uniform distribution, then

$$x_{i1} = \theta_1 \left[-\log(1-u_{i1})\right]^{1/\beta_1},$$

where $u_{i1} \sim Uniform\,(0,1)$. The censoring indicator $\delta_{i1}$ is then determined by comparing $x_{i1}$ with the censoring value $c_{i1}$.

$X_2$ is generated sequentially by setting its conditional CDF, $F_{X_2|X_1}$, qual to a random number of standard Uniform distribution. The conditional CDF is given by

$$\begin{aligned} F_{X_2|X_1} &= \frac{F_{X_1,X_2}(x_1,x_2)}{F_{X_1}(x_1)} \qquad (4.1) \\ &= \frac{F_{X_1}(x_1)\,F_{X_2}(x_2)\left[1+\alpha\left(1-F_{X_1}(x_1)\right)\left(1-F_{X_2}(x_2)\right)\right]}{F_{X_1}(x_1)} \\ &= F_{X_2}(x_2)\left[1+\alpha\left(1-F_{X_1}(x_1)\right)\left(1-F_{X_2}(x_2)\right)\right] \\ &= (1-V)(1+\alpha WV) \end{aligned}$$

where $V = \exp\left(-\left(\frac{x_2}{\theta_2}\right)^{\beta_2}\right), W = \exp\left(-\left(\frac{x_1}{\theta_1}\right)^{\beta_1}\right)$.

Setting (4.1) equal to a standard uniform random number, then we have a quadratic equation in terms of $V$

$$\alpha W V^2 + (1 - \alpha W) V + (u_{i2} - 1) = 0,$$

where $u_{i2} \sim Uniform(0, 1)$ is independent of $u_{i1}$.

By Quadratic Formula and $V \geq 0$,

$$\begin{aligned} V &= \frac{(\alpha W - 1) + \sqrt{(1 - \alpha W)^2 - 4\alpha W (u_{i2} - 1)}}{2\alpha W} \\ &= \frac{1}{2} + \frac{\sqrt{(1 + \alpha W)^2 - 4\alpha W u_{i2}} - 1}{2\alpha W}, \end{aligned} \tag{4.2}$$

and then $X_2$ is given by

$$x_{i2} = \theta_2 \left[-\log(V)\right]^{1/\beta_2}.$$

The censoring indicator $\delta_{i2}$ is then determined by comparing $y_{i2}$ with the censoring value $c_{i2}$.

### 4.2.3 Simulation Settings and Results for Bivariate Location-Scale Models

**Simulation Settings**

1. Data sets are generated by methods listed in section 4.2.1 and 4.2.2;

2. Sample size, n, is set at 25;

3. Simulation iteration is set to 500;

4. True parameter values are $\beta_1 = 0.5$, $\beta_2 = 2$, $\theta_1 = \theta_2 = 10$.(equal scale parameters);

5. Depedence parameters are set as $\delta = 0.5$, $\alpha = 0.5$ for the RE and FGM models, respectively.

Both the joint MLEs and working independence MLEs are obtained and compared against each other.

Percent improvements/losses in terms of absolute bias and empirical MSE (mean squared errors) are also calculated. We found mixed turnout of improvements and losses.

**The Random Effect(Hougaard) Model Simulation Results**

Table 4.3 summarizes maximum likelihood estimation results based on the joint and the working independence models. Tables 4.4 and 4.5 summarize the percentage improvement/losses obtained by comparing joint mle's against working mle's in terms of their biasses and Empirical Mean Squared Errors(EMSE). Table 4.5 shows overall improvement in MSE for mle's of the shape parameters $\beta_1$ and $\beta_1$. But for mle's of the scale parameters, no such pattern is found. It is expected as stated in [24, Johnson, Evans and Green (1999)] that neither the sample correlation nor the population correlation depends on the values of the underlying scale parameters of the marginal distributions.

### The FGM Model Simulation Results

Table 4.6 summarizes maximum likelihood estimation results based on the joint and the working independence models. Tables 4.7 and 4.8 summarize the improvements/losses in biasses and EMSE obtained by comparing the joint mle's against working mle's. Table 8 shows overall improvement in MSE only for mle's of $\beta_1$ and slightly loss in $\beta_2$.

## 4.3 Example: DMBA-Induced Tumors

Table 4.9 contains the first and second tumor times for 30 control and 30 treated animals, simulated as $BVW(0.000009, 0.000005, \beta = 3)$ and $BVW(0.000001, 0.0000005, \beta = 3)$, respectively, with $a = 1$ in both cases. We see that there were 18 and 15 simultaneous tumors for the control and treated animals, respectively. We first estimated the shape parameter based on the marginal Weibull likelihoods as $\hat{\beta} = 2.93$. Then we compute the scale parameter estimates for each population using the estimates given in (24). This yielded $\hat{\lambda} = (0.00001287, 0.00000776)$ and $(0.00000179, 0.00000082)$ for the controls and treated animals, respectively.

| $\delta$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\beta}_1^*$ | $\hat{\beta}_2^*$ | $\hat{\theta}_1^*$ | $\hat{\theta}_2^*$ |
|---|---|---|---|---|---|---|---|---|
| 0.1 | 0.52768 | 2.11475 | 10.5071 | 9.9578 | 0.52775 | 2.11411 | 10.4546 | 9.9468 |
| 0.2 | 0.52601 | 2.10351 | 10.3663 | 9.9479 | 0.52588 | 2.10134 | 10.3499 | 9.9376 |
| 0.3 | 0.52724 | 2.11299 | 10.5591 | 9.9889 | 0.52688 | 2.10869 | 10.5284 | 9.9826 |
| 0.4 | 0.53299 | 2.11051 | 10.7050 | 9.9997 | 0.53218 | 2.11411 | 10.6547 | 9.9946 |
| 0.5 | 0.53413 | 2.16718 | 10.6873 | 10.0069 | 0.53587 | 2.16407 | 10.6663 | 9.9937 |
| 0.6 | 0.52750 | 2.11608 | 10.3724 | 9.9060 | 0.52631 | 2.11830 | 10.3399 | 9.8988 |
| 0.7 | 0.52602 | 2.09350 | 10.4076 | 9.9371 | 0.52619 | 2.09460 | 10.3778 | 9.9301 |
| 0.8 | 0.53550 | 2.11476 | 10.8448 | 10.0144 | 0.53444 | 2.11509 | 10.8130 | 10.0098 |
| 0.9 | 0.53219 | 2.12814 | 10.6916 | 9.9758 | 0.53207 | 2.12961 | 10.6822 | 9.9707 |
| 1.0 | 0.52696 | 2.10093 | 10.5949 | 9.8745 | 0.52768 | 2.10355 | 10.6092 | 9.8756 |

Where $\hat{\beta}^*$ and $\hat{\theta}^*$ are working independence MLEs.

Table 4.3: Joint and Working MLEs with $\beta_1 = 0.5$, $\beta_2 = 2$, $\theta_1 = \theta_2 = 10$ and varying $\delta$

| $\delta$ | $\beta_1$ bias improv. | $\beta_2$ bias improv. | $\theta_1$ bias improv. | $\theta_2$ bias improv. |
|---|---|---|---|---|
| 0.1 | -2477.84 | 0.24710 | -0.56025 | -11.5567 |
| 0.2 | -1968.70 | -0.49651 | -2.13658 | -4.6773 |
| 0.3 | -9211.99 | -1.34267 | -3.95333 | -5.8183 |
| 0.4 | -328586.83 | -2.51654 | 3.15616 | -7.6807 |
| 0.5 | -17276.79 | 4.84207 | -1.89587 | -3.1514 |
| 0.6 | -1058.00 | -4.51075 | 1.87806 | -9.5496 |
| 0.7 | -1560.73 | 0.63741 | 1.16725 | -7.8866 |
| 0.8 | -7062.51 | -3.07633 | 0.28694 | -3.9194 |
| 0.9 | -4334.31 | -0.37612 | 1.13056 | -1.3817 |
| 1.0 | -853.27 | 2.60031 | 2.52758 | 2.3545 |

Table 4.4: Bias Improvements/Losses(%) Over the Working Estimates

| $\delta$ | $\beta_1$ mse improv. | $\beta_2$ mse improv. | $\theta_1$ mse improv. | $\theta_2$ mse improv. |
|---|---|---|---|---|
| 0.1 | 90.8621 | 1.61977 | 1.35134 | -1.40621 |
| 0.2 | 81.2788 | 1.28675 | 4.61830 | -0.01806 |
| 0.3 | 70.9359 | 4.39391 | -1.04580 | 0.26043 |
| 0.4 | 63.3886 | 5.83908 | 6.00907 | -0.99596 |
| 0.5 | 58.1423 | 7.17802 | 3.45869 | 1.05814 |
| 0.6 | 44.2761 | 4.93833 | 0.76780 | -0.44087 |
| 0.7 | 32.2627 | 4.65393 | 3.68219 | -0.62015 |
| 0.8 | 22.5092 | -0.46238 | -0.63960 | 0.20701 |
| 0.9 | 18.3216 | 0.68599 | -1.02801 | -0.20352 |
| 1.0 | 17.0853 | -0.25471 | 1.01207 | -0.14121 |

Table 4.5: Empirical MSE Improvements/Losses(%) Over the Working Estimates

| $\delta$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\beta}_1^*$ | $\hat{\beta}_2^*$ | $\hat{\theta}_1^*$ | $\hat{\theta}_2^*$ |
|---|---|---|---|---|---|---|---|---|
| 0.1 | 0.53263 | 2.08660 | 10.3397 | 9.76154 | 0.53228 | 2.08518 | 10.3515 | 9.76080 |
| 0.2 | 0.52865 | 2.09570 | 10.4262 | 9.67449 | 0.52839 | 2.09407 | 10.4271 | 9.68024 |
| 0.3 | 0.53248 | 2.05551 | 10.5498 | 9.60398 | 0.53197 | 2.05479 | 10.5549 | 9.61140 |
| 0.4 | 0.52701 | 2.02371 | 10.1626 | 9.39615 | 0.52585 | 2.02132 | 10.1284 | 9.39784 |
| 0.5 | 0.53353 | 2.06266 | 10.8290 | 9.24950 | 0.53263 | 2.05816 | 10.7823 | 9.24247 |
| 0.6 | 0.53342 | 2.02963 | 10.5895 | 9.10675 | 0.53249 | 2.02805 | 10.5501 | 9.10546 |
| 0.7 | 0.52007 | 2.01353 | 10.4359 | 8.95880 | 0.51912 | 2.01013 | 10.3927 | 8.95503 |
| 0.8 | 0.53457 | 2.01028 | 10.7978 | 8.87893 | 0.53256 | 2.00488 | 10.6808 | 8.86241 |
| 0.9 | 0.53131 | 1.99020 | 10.7238 | 8.60068 | 0.52961 | 1.98629 | 10.6340 | 8.60127 |
| 1.0 | 0.52763 | 2.00859 | 10.8060 | 8.50162 | 0.52544 | 2.00157 | 10.7181 | 8.48742 |

Where $\hat{\beta}^*$ and $\hat{\theta}^*$ are working independence MLEs.

Table 4.6: Joint and Working MLEs with $\beta_1 = 0.5$, $\beta_2 = 2$, $\theta_1 = \theta_2 = 10$ and varying $\delta$

| $\delta$ | $\beta_1$ bias improv. | $\beta_2$ bias improv. | $\theta_1$ bias improv. | $\theta_2$ bias improv. |
|---|---|---|---|---|
| 0.1 | -342.920 | -1.10145 | -1.660 | 3.3686 |
| 0.2 | -288.550 | -0.92829 | -1.731 | 0.1915 |
| 0.3 | -205.638 | -1.59816 | -1.311 | 0.9308 |
| 0.4 | -134.917 | -4.48150 | -11.252 | -26.6615 |
| 0.5 | -120.447 | -2.76162 | -7.725 | -5.9750 |
| 0.6 | -118.480 | -2.84038 | -5.632 | -7.1674 |
| 0.7 | -94.343 | -4.99189 | -33.548 | -10.9885 |
| 0.8 | -104.619 | -6.18119 | -110.420 | -17.1839 |
| 0.9 | -103.403 | -5.74152 | 28.534 | -14.1625 |
| 1.0 | -119.249 | -8.62320 | -447.628 | -12.2393 |

Table 4.7: Bias Improvements/Losses(%) Over the Working Estimates

| $\delta$ | $\beta_1$ mse improv. | $\beta_2$ mse improv. | $\theta_1$ mse improv. | $\theta_2$ mse improv. |
|---|---|---|---|---|
| 0.1 | 90.4678 | -1.16397 | 0.13321 | 0.37230 |
| 0.2 | 84.3979 | -0.35227 | -1.69771 | -0.51571 |
| 0.3 | 75.6162 | -0.94049 | -0.72841 | -0.61527 |
| 0.4 | 71.8620 | -2.32073 | 0.54219 | -2.55347 |
| 0.5 | 69.7788 | -1.72126 | -2.74445 | -2.78227 |
| 0.6 | 69.4358 | -1.16677 | -0.15066 | -3.16534 |
| 0.7 | 65.3872 | -1.84386 | 0.36968 | -1.85708 |
| 0.8 | 65.0311 | -1.49496 | 0.06637 | -2.62671 |
| 0.9 | 68.4773 | -1.58717 | -0.34932 | -4.46156 |
| 1.0 | 69.2833 | -2.11464 | -0.88495 | -1.56760 |

Table 4.8: Empirical MSE Improvements/Losses(%) Over the Working Estimates

| Control Animals ($n = 30, a = 1, k = 18$) |
|---|
| (28,80), (53,53), (30,47), (25,63), (75,75), (21,21), (55,67), (55,66), (42,43), (79,79), (56,56), (56,56), (42,42), (64,64), (44,56), (39,57), (63,63), (41,41), (28,28), (49,49), (34,34), (10,82), (53,55), (26,26), (43,52), (16,16), (57,86), (56,56), (29,29), (19,19) |
| Treated Animals ($n = 30, a = 1, k = 15$)) |
| (50,50), (69,69), (53,53), (66,95), (77,77), (102,102), (114,142), (83,83), (63,63), (58,116), (80,137), (122,122), (90,90), (42,65), (106,106), (90,90), (114,114), (117,172), (98,98), (82,82), (22,99), (102,138), (123,147), (61,120), (80,138), (75,142), (78,146), (51,166), (12,180), (147,147), |

Table 4.9: Bivariate Weibull Times to first and second tumor

# Bibliography


[1] Au, C. and Tam, J. (1999), "Transforming variables using the dirac generalized function," *The American Statistician*, 53, pp. 270-273.

[2] Casella, G and Berger, R. L. (2002), *Statistical Inference,* Duxbury Thomson Learning, Pacific Grove.

[3] Csorgo, S. and Welsh, A.H. (1989), "Testing for exponential and Marshall-Olkin distributions," *Journal of Statistical Planning and Inference*, 23, pp. 287-300.

[4] Dudewicz, E. J. and Mishra, S. (1988), *Modern Mathematical Statistics*, Wiley, New York.

[5] Clayton, D. G. (1978). "A model for association in bivariate lifetables and its application in epidemiological studies of familial tendency in chronic disease incidence, " *Biometrika,* 65, pp. 141-151.

[6] Conway, D. (1983). "Farlie–Gumbel–Morgenstern distributions," *Encyclopedia of Statistical Sciences*, 3, pp. 28–31.

[7] Cohen, A. C. (1965). "Maximum likelihood estimation in the weibull distribution based on complete and on censored samples, " *Technometrics,* 7, pp. 579-588.

[8] Cohen, A. C. (1975). "Multi-Censored Sampling in the Three Parameter Weibull Distribution, " *Technometrics,* 17, pp. 347-351.

[9] Devore, J. L. (2000). *Probability and Statistics for Engineering and the Sciences,* Duxbury Thomson Learning, Pacific Grove.

[10] Efron, B. (1985). "Bootstrap confidence intervals for a class of parametric problems, " *Biometrika,* 72, pp. 45-58.

[11] Eskow, E. and Schnabel, R. B. (1991). "Algorithm 695: software for a new modified Cholesky factorization," *ACM Transactions on Mathematical Software*, 17, pp. 306 -312.

[12] Farlie, D. J. G. (1960). "The performance of some correlation coefficients for a general bivariate distribution," *Biometrika,* 47, pp. 307-323.

[13] Freund, J. E. (1961). "A bivariate extension of the exponential distribution," *J. Am. Statist. Ass.,* 56, pp. 971-977.

[14] Gould, A. and Lawless, J. F. (1988). "Consistency and efficiency of regression coefficient estimates in location-scale models, " *Biometrika,* 75, pp. 535-540.

[15] Gumbel, E. J. (1960). "Bivariate exponential distributions," *J. Am. Statist. Ass.,* 55, pp. 698-707.

[16] Hanagal, D. D. (1996). "Multivariate Weibull distribution," *Economic Quality Control,* 11, pp. 193-200.

[17] Hougaard, P. (1986). "A class of multivariate failure time distributions." *Biometrika*, 73, 3, pp. 671-678.

[18] Hougaard, P. (2000). *Analysis of Multivariate Survival Data,* Springer-Verlag, New York.

[19] He, W. and Lawless, J. F. (2005). "Bivariate location-scale models for regression analysis, with applications to lifetime data, " *J. R. Statist. Soc.,* 67, pp. 63-78.

[20] Huang, J. S. and Kotz, S. (1984). "Correlation structure in iterated Farlie–Gumbel–Morgenstern distributions," *Biometrika*, 71, pp. 633–636.

[21] Iyer, S. K., Manjunath, D. and Manivasakan, R. (2002). "Bivariate Exponential Distributions Using Linear Structures," *Sankhya*, 64(A), pp. 156-166.

[22] Iyer, S. K. and Manjunath, D. (2004). "Correlated Bivariate Sequence for queueing and reliability applications," *Communications in Statistics*, 33, pp 331-350.

[23] Joe, H. (1997). *Multivariate Models and Dependence Concepts,* Chapman and Hall, New York.

[24] Johnson, R. A., Evans, J. W. and Green, D. W.(1999). *Some Bivariate Distributions for Modeling the Strength Properties of Lumber*, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison.

[25] Khuri, A. (2004). "Applications of Dirac's Delta Function in Statistics," *Int. J. Math. Educ. Sci. Technol*, 35, 2, pp.

[26] Lawless, J. F. (2002). *Statistical Models and Methods for Lifetime Data*, Wiley, New York.

[27] Lee, L. (1979), "Multivariate Distributions having Weibull Properties," *Journal of Multivariate Analysis*, 9, pp. 267-277.

[28] Lehmann, E. L. ( 1983). *Theory of Point Estimation*, Wiley, New York.

[29] Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation,* Springer-Verlag, New York.

[30] Liu, C. (1997), *A Comparison Between the Weibull and Lognormal Models Used to Analyse Reliability Data*, University of Nottingham Doctoral Thesis, UK.

[31] Lu, J. C. and Bhattacharyya, G. K. (1990). "Some New Constructions of Bivariate Weibull Models," *Ann. Inst. Statist. Math.*, 43(3), pp. 543-559.

[32] Marshall, A. W. and Olkin, I. (1967). "A Multivariate Exponential Distribution," *J. Amer. Stat. Assoc.*, 63, pp. 30-44.

[33] Marshall, A. W. and Olkin, I. (1988). "Families of multivariate distributions," *J. Amer. Stat. Assoc.*, 83, 834–841.

[34] Meeker, W. Q. and Escobar, L. A. (1998). *Statistical Methods for Reliability Data,* Wiley, New York.

[35] Miller, G. (1984). "Inference on a future reliability parameter with the Weibull process model, " *Naval Research Logistics Quarterly,* 31, pp. 91-96.

[36] Morgenstern, D. (1956). "Einfache beispiele zweidimensionaler verteilungen," *Mitt. Math. Statis.*, 8, pp. 234-235.

[37] Murthy, D. N. P., Xie, M., Jiang, R. (2004). *Weibull Models,* Wiley, New York.

[38] Rao. C. R. ( 1973). *Linear Statistical Inference and Its Applications*, Wiley, New York.

[39] Rausand, M. and Hoyland, A. (1994). *System Reliability Theory; Models, Statistical Methods and Applications*, Wiley, New York.

[40] Sarkar, S. K. (1987). "A continuous bivariate exponential distribution," *J. Amer. Stat. Assoc.*, 82, pp. 667–675.

[41] Zeller, A. (1962). "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias," *J. Am. Statist. Ass.,* 57, pp. 400-405.

# Appendices

# Appendix A

# SAS Simulation Codes

## A.1 Data Generation and MLE for Linearly Associated BVE and BVW

```
data _null_;
retain c 1;
do i=.10, .2,.3,.4,.5,.6;
call symput('loop'||left(c),i);
call symput('step', c);
c+1;
end;
run;
%MACRO sim(lambda1, lambda2,beta, n, iter,table);
%do i= 1 %to &step;
title "lambda1=&lambda1 lambda2=&lambda2 roh=&&loop&i samplesize=&n";
DATA data1;
lambda1=&lambda1;
lambda2=&lambda2;
    beta=&beta;
roh=&&loop&i;
a=roh*(lambda1/lambda2);
```

```
p=a*lambda2/lambda1;
DO iter=1 to &iter;
DO i = 1 to &n;
d=0;
z=RANBIN(0,1,1-p)*RANEXP(0)/lambda2;
X1=RANEXP(0)/lambda1;
X2=a*X1 + z;
    y1=x1**(1/beta); y2=x2**(1/beta);
if z=0 then d=1;
ind=1-d;
OUTPUT;
    end;
END;
run;
PROC NLP tech=newrap DATA=data1 OUTEST=init noprint;
MAX logf;
PARMS Lam1=1, Lam2=1,b=1;
logf=2*log(b) + log(Lam1)+ log(Lam2)
+(b-1)*log(y1) + (b-1)*log(y2)
-Lam1*y1**b - Lam2*y2*b;
BY iter;
RUN;
```

```
DATA init;
SET init;
IF _TYPE_='PARMS';
KEEP iter lam1 lam2 b;
RENAME b=bhat;
RUN;
DATA data2a;
MERGE data1 init; by iter;
y11=(y1**bhat); y22=(y2**bhat);
RUN;
proc means data=data2a noprint;
var y11 y22 a d;
by iter;
output out=data2 mean(y11 y22 a bhat lam1 lam2)=y1bar y2bar a bhat lam1
lam2 sum(d)=k;
run;
data mle;
set data2;
    lambda1=&lambda1;
lambda2=&lambda2;
roh=&&loop&i;
a=roh*(lambda1/lambda2);
```

```
lambhat11=a/y2bar + (&n-k)/(&n*y1bar);
lambhat12=1/y2bar;
rohhatstar=a*lambhat12/lambhat11;
lamb11bias=lambhat11-&lambda1;
lamb12bias=lambhat12-&lambda2;
rohhatstarbias=rohhatstar-&&loop&i;
lamb11mse=lamb11bias**2;
lamb12mse=lamb12bias**2;
rohhatstarmse=rohhatstarbias**2;
se_lamb11=sqrt((lambhat11*(lambhat11-a*lambhat12)+a**2*lambhat12**2)/&n);
LCL = lambhat11 - 2*se_lamb11;
UCL = lambhat11 + 2*se_lamb11;
conf=(lcl le lambda1 le ucl);
se_lam1=sqrt(lam1**2/&n);
LCL1 = lam1 - 2*se_lam1;
UCL1 = lam1 + 2*se_lam1;
conf1=(lcl1 le lambda1 le ucl1);
range=(ucl-lcl)/2;
range1=(ucl1-lcl1)/2;
rohhat=a*lam2/lam1;
rohhatbias=rohhat-&&loop&i;
lam1bias=lam1-&lambda1;
```

```
lam2bias=lam2-&lambda2;
lam1MSE=lam1bias**2;
lam2MSE=lam2bias**2;
      rohhatMSE=rohhatbias**2;
run;
PROC MEANS data=mle noprint;
var roh conf lcl ucl conf1 lcl1 ucl1 range range1 lambhat11 lam1 lamb11bias
lam1bias lamb11mse lam1mse
lambhat12 lam2 lamb12bias lam2bias lamb12mse lam2mse
rohhatstar rohhat rohhatstarbias rohhatbias rohhatstarmse rohhatmse;
      output out=stats
mean=rho conf lcl ucl conf1 lcl1 ucl1 range range1 lambhat11 lam1 lamb11bias
lam1bias lamb11mse lam1mse
lambhat12 lam2 lamb12bias lam2bias lamb12mse lam2mse
rohhatstar rohhat rohhatstarbias rohhatbias rohhatstarmse rohhatmse;
run;
PROC DATASETS nodetails nolist force;
APPEND BASE=work.table&table DATA=stats;
RUN;
%end;
DATA sim.table&table;
SET table&table;
```

```
lbiasimp=(abs(lam1bias)-abs(lamb11bias))/lam1bias*100;
lmseimp=(lam1mse-lamb11mse)/lam1mse*100;
rhobiasimp=(abs(rohhatbias)-abs(rohhatstarbias))/rohhatbias*100;
rhomseimp=(rohhatmse-rohhatstarmse)/rohhatmse*100;
RUN;
PROC EXPORT DATA= sim.TABLE&table
OUTFILE= "C:\Documents and Settings\BB\My Documents\PAPER\Spring05\code\tables.
DBMS=EXCEL REPLACE;
SHEET="sheet&table";
RUN;
%MEND sim;
%sim(10,1,2, 50, 1000,118);
/*%sim(100,1,50, 10000,6);
%sim(1,100,50, 10000,7);
%sim(100,100,50, 10000,8);*/
```

## A.2 Data Generation and MLE for BVW of Hougaard's Model

```
%LET directory=C:\;
LIBNAME sim "&directory";
%let seed=0;
%let beta1=0.5;
%let beta2=2;
```

```
%let theta1=10;
%let theta2=10;
%let n=25;
%macro hougsimu(iter);
%do dl=1 %to 10;
%let delta=&dl/10;
%do k=1 %to &iter;
DATA houg;
    DO i = 1 to &n;
U=uniform(0); U2=uniform(0); U3=uniform(0); U4=uniform(0); U5=uniform(0);
    V=(-log(U2)-log(u3))*(u5 le &delta) -log(u4) *(U5 > &delta);
    X1=U**(&delta/&beta1)*v**(1/&beta1)*&theta1;
    x2=(1-U)**(&delta/&beta2)*v**(1/&beta2)*&theta2;
        delta=&delta;
    OUTPUT;
END;
RUN;
PROC NLP tech=newrap outest=out1 DATA=houg noprint;
    MAX logf;
    PARMS bhat1 that1;
    logf=-log(that1)+log(bhat1)+(bhat1-1)*log(x1/that1)
    -(x1/that1)**bhat1;
```

```
    by delta;
RUN;
data out1; set out1;
    KEEP _type_ bhat1 that1 delta;
    if _type_='PARMS';
RUN;
PROC NLP tech=newrap outest=out2 DATA=houg noprint;
    MAX logf;
    PARMS bhat2 that2;
    logf=-log(that2)+log(bhat2)+(bhat2-1)*log(x2/that2)
    -(x2/that2)**bhat2;
    by delta;
RUN;
DATA out2; set out2;
    KEEP _type_ bhat2 that2 delta;
    IF _type_='PARMS';
RUN;
*working independent mle;
DATA seedwrk;
    merge out1 out2;
RUN;
```

```
PROC NLP tech=congra DATA=houg inest=seedwrk outest=hougwrkest noprint;
    MAX logf;
    PARMS bhat1 that1 bhat2 that2;
    logf=log(bhat1)-log(that1)+log(bhat2)-log(that2)
+(bhat1-1)*log(x1/that1)+(bhat2-1)*log(x2/that2)
-log((x1/that1)**bhat1 + (x2/that2)**bhat2)
+log((x1/that1)**bhat1 + (x2/that2)**bhat2)
-(x1/that1)**bhat1 - (x2/that2)**bhat2;
    by delta;
RUN;
data simwrk;
    set hougwrkest;
    if _type_='PARMS';
    bhat1wrk=bhat1;
    bhat2wrk=bhat2;
    that1wrk=that1;
    that2wrk=that2;
    keep bhat1wrk bhat2wrk that1wrk that2wrk;
run;
*joint mle;
DATA seed1;
```

```
set seedwrk;
      dhat=0.1;
RUN;
PROC NLP tech=congra DATA=houg inest=seed1 outest=seed2 noprint;
      MAX logf;
      PARMS dhat; bounds 0<dhat<=1;
      logf=log(bhat1/that1)+log(bhat2/that2)
+(bhat1/dhat-1)*log(x1/that1)+(bhat2/dhat-1)*log(x2/that2)
+(dhat-2)*log((x1/that1)**(bhat1/dhat) + (x2/that2)**(bhat2/dhat))
+log(((x1/that1)**(bhat1/dhat) + (x2/that2)**(bhat2/dhat))**dhat+1/dhat-
1)
-((x1/that1)**(bhat1/dhat) + (x2/that2)**(bhat2/dhat))**dhat;
      by delta;
RUN;
DATA seed2;
      SET seed2;
      KEEP _type_ bhat1 bhat2 that1 that2 dhat delta;
      if _type_='PARMS';
RUN;
PROC NLP tech=nmsimp DATA=houg inest=seed2 outest=hougest noprint;
      MAX logf;
      PARMS dhat bhat1 bhat2 that1 that2; bounds 0<dhat<=1;
```

```
logf=log(bhat1)-log(that1)+log(bhat2)-log(that2)
+(bhat1/dhat-1)*log(x1/that1)+(bhat2/dhat-1)*log(x2/that2)
+(dhat-2)*log((x1/that1)**(bhat1/dhat) + (x2/that2)**(bhat2/dhat))
+log(((x1/that1)**(bhat1/dhat) + (x2/that2)**(bhat2/dhat))**dhat+1/dhat-
1)
-((x1/that1)**(bhat1/dhat) + (x2/that2)**(bhat2/dhat))**dhat;
by delta;
RUN;
data sim;
set hougest;
if _type_='PARMS';
run;
data simrslt;
merge simwrk sim;
bhat1bias=bhat1-&beta1;
bhat2bias=bhat2-&beta2;
that1bias=that1-&theta1;
that2bias=that2-&theta2;
bhat1wrkbias=bhat1wrk-&beta1;
bhat2wrkbias=bhat2wrk-&beta2;
that1wrkbias=that1wrk-&theta1;
that2wrkbias=that2wrk-&theta2;
```

```
    bhat1mse=bhat1bias**2;
    bhat2mse=bhat2bias**2;
    that1mse=that1bias**2;
    that2mse=that2bias**2;
    bhat1wrkmse=bhat1wrkbias**2;
    bhat2wrkmse=bhat2wrkbias**2;
    that1wrkmse=that1wrkbias**2;
    that2wrkmse=that2wrkbias**2;
run;
PROC DATASETS nodetails nolist force;
APPEND BASE=sim.hougresult DATA=simrslt;
RUN;
%end;
%end;
%mend hougsimu;
%hougsimu(1)
title theta1=&theta1 beta1=&beta1 theta2=&theta2 beta2=&beta2 delta=&delta;
proc sort data=sim.hougresult;
    by delta;
run;
PROC MEANS data=sim.hougresult noprint;
```

```
        var bhat1 bhat2 that1 that2 bhat1wrk bhat2wrk that1wrk that2wrk dhat
delta
            bhat1mse bhat2mse that1mse that2mse bhat1wrkmse bhat2wrkmse that1wrkmse
that2wrkmse
            bhat1bias bhat2bias that1bias that2bias bhat1wrkbias bhat2wrkbias
that1wrkbias that2wrkbias;
        output out=stats
        mean=bhat1 bhat2 that1 that2 bhat1wrk bhat2wrk that1wrk that2wrk dhat
            bhat1mse bhat2mse that1mse that2mse bhat1wrkmse bhat2wrkmse that1wrkmse
that2wrkmse
            bhat1bias bhat2bias that1bias that2bias bhat1wrkbias bhat2wrkbias
that1wrkbias that2wrkbias;
        by delta;
    run;
    DATA hougtable;
        SET stats;
        b1biasimp=(abs(bhat1wrkbias)-abs(bhat1bias))/abs(bhat1wrkbias)*100;
        b2biasimp=(abs(bhat2wrkbias)-abs(bhat2bias))/abs(bhat2wrkbias)*100;
        t1biasimp=(abs(that1wrkbias)-abs(that1bias))/abs(that1wrkbias)*100;
        t2biasimp=(abs(that2wrkbias)-abs(that2bias))/abs(that2wrkbias)*100;
        b1mseimp=(bhat1wrkmse-bhat1mse)/bhat1wrkmse*100;
        b2mseimp=(bhat2wrkmse-bhat2mse)/bhat2wrkmse*100;
```

```
t1mseimp=(that1wrkmse-that1mse)/that1wrkmse*100;
t2mseimp=(that2wrkmse-that2mse)/that2wrkmse*100;
keep b1biasimp b2biasimp t1biasimp t2biasimp
     b1mseimp b2mseimp t1mseimp t2mseimp delta;
RUN;
proc print data=stats;
  var delta dhat bhat1 bhat2 that1 that2 bhat1wrk bhat2wrk that1wrk that2wrk;
run;
proc print data=hougtable;
run;
quit;
```

## A.3 Data Generation and MLE for BVW of FGM Model

```
*x 'del C:\fgmresult.sas7bdat';
goptions reset=all; options nodate;
%LET directory=C:\Documents and Settings\Zhigang\My Documents\Yi Han
thesis simulation;
LIBNAME sim "&directory";
%let seed=0;
%let beta1=0.5;
%let beta2=2;
%let theta1=10;
```

```
%let theta2=10;
%let n=25;
%macro fgmsimu(iter);
%do al=1 %to 10;
%let alpha=&al/10;
%do k=1 %to &iter;
DATA fgm;
    DO i = 1 to &n;
    U1=uniform(&seed); U2=uniform(&seed);
        x1=&theta1*(-log(1-U1))**(1/&beta1);
    W=exp(-(x1/&theta1)**&beta1);
        b=(((1+&alpha*W)**2-4*&alpha*W*U2)**0.5-1)/(2*&alpha*W);
        V1=0.5+(((1+&alpha*W)**2-4*&alpha*W*U2)**0.5-1)/(2*&alpha*W);
        V2=0.5-(((1+&alpha*W)**2-4*&alpha*W*U2)**0.5+1)/(2*&alpha*W);
        V=max(V1, V2);
    X2=&theta2*(-log(V))**(1/&beta2);
        alpha=&alpha;
    OUTPUT;
END;
RUN;
/*PROC CORR DATA=fgm out=corrout noprint;
VAR x1 x2;
```

```
RUN;*/
PROC NLP tech=newrap outest=out1 DATA=fgm noprint;
MAX logf;
PARMS bhat1 that1;
logf=log(bhat1)-bhat1*log(that1)+(bhat1-1)*log(x1)
-(x1/that1)**bhat1;
    by alpha;
RUN;
data out1; set out1;
KEEP _type_ bhat1 that1 alpha;
    if _type_='PARMS';
RUN;
PROC NLP tech=newrap outest=out2 DATA=fgm noprint;
MAX logf;
PARMS bhat2 that2;
logf=log(bhat2)-bhat2*log(that2)+(bhat2-1)*log(x2)
-(x2/that2)**bhat2;
    by alpha;
RUN;
DATA out2; set out2;
KEEP _type_ bhat2 that2 alpha;
    IF _type_='PARMS';
```

```
RUN;
*working independent mle;
data seedwrk;
    merge out1 out2;
run;
PROC NLP tech=newrap DATA=fgm inest=seedwrk outest=fgmwrkest noprint;
MAX logf;
PARMS bhat1 bhat2 that1 that2; * bounds bhat1=1, bhat2=2;
logf=log(bhat1)-log(that1)+log(bhat2)-log(that2)
+(bhat1-1)*(log(x1)-log(that1))+(bhat2-1)*(log(x2)-log(that2))
    -(x1/that1)**bhat1-(x2/that2)**bhat2;
    by alpha;
RUN;
data simwrk;
    set fgmwrkest;
    if _type_='PARMS';
    bhat1wrk=bhat1;
    bhat2wrk=bhat2;
    that1wrk=that1;
    that2wrk=that2;
    keep bhat1wrk bhat2wrk that1wrk that2wrk;
run;
```

```
*joint mle;
DATA seed1;
set seedwrk;
    alphat=0.1;
RUN;
PROC NLP tech=newrap DATA=fgm inest=seed1 outest=seed2 noprint;
MAX logf;
PARMS alphat; bounds -1<=alphat<=1;
logf=log(bhat1)-log(that1)+log(bhat2)-log(that2)
+(bhat1-1)*(log(x1)-log(that1))+(bhat2-1)*(log(x2)-log(that2))
    -(x1/that1)**bhat1-(x2/that2)**bhat2
+log(1+alphat*(2*exp(-(x1/that1)**bhat1)-1)*(2*exp(-(x2/that2)**bhat2)-1));
    by alpha;
RUN;
DATA seed2;
SET seed2;
KEEP _type_ bhat1 bhat2 that1 that2 alphat;
if _type_='PARMS';
RUN;
PROC NLP tech=newrap DATA=fgm inest=seed2 outest=fgmest noprint;
MAX logf;
PARMS bhat1 bhat2 that1 that2 alphat; bounds -1<=alphat<=1;
```

```
logf=log(bhat1)-log(that1)+log(bhat2)-log(that2)
+(bhat1-1)*(log(x1)-log(that1))+(bhat2-1)*(log(x2)-log(that2))
    -(x1/that1)**bhat1-(x2/that2)**bhat2
+log(1+alphat*(2*exp(-(x1/that1)**bhat1)-1)*(2*exp(-(x2/that2)**bhat2)-1));
    by alpha;
RUN;
data sim;
    set fgmest;
    if _type_='PARMS';
run;
data simrslt;
    merge simwrk sim;
    bhat1bias=bhat1-&beta1;
    bhat2bias=bhat2-&beta2;
    that1bias=that1-&theta1;
    that2bias=that2-&theta2;
    bhat1wrkbias=bhat1wrk-&beta1;
    bhat2wrkbias=bhat2wrk-&beta2;
    that1wrkbias=that1wrk-&theta1;
    that2wrkbias=that2wrk-&theta2;
    bhat1mse=bhat1bias**2;
    bhat2mse=bhat2bias**2;
```

```
    that1mse=that1bias**2;
    that2mse=that2bias**2;
    bhat1wrkmse=bhat1wrkbias**2;
    bhat2wrkmse=bhat2wrkbias**2;
    that1wrkmse=that1wrkbias**2;
    that2wrkmse=that2wrkbias**2;
run;
PROC DATASETS nodetails nolist force;
APPEND BASE=sim.fgmresult DATA=simrslt;
RUN;
%end;
%end;
%mend fgmsimu;
%fgmsimu(500)
title theta1=&theta1 beta1=&beta1 theta2=&theta2 beta2=&beta2;
proc sort data=sim.fgmresult;
    by alpha;
run;
PROC MEANS data=sim.fgmresult noprint;
    var bhat1 bhat2 that1 that2 bhat1wrk bhat2wrk that1wrk that2wrk alphat
alpha
```

```
        bhat1mse bhat2mse that1mse that2mse bhat1wrkmse bhat2wrkmse that1wrkmse
that2wrkmse
        bhat1bias bhat2bias that1bias that2bias bhat1wrkbias bhat2wrkbias
that1wrkbias that2wrkbias;
    output out=stats
    mean=bhat1 bhat2 that1 that2 bhat1wrk bhat2wrk that1wrk that2wrk al-
phat
        bhat1mse bhat2mse that1mse that2mse bhat1wrkmse bhat2wrkmse that1wrkmse
that2wrkmse
        bhat1bias bhat2bias that1bias that2bias bhat1wrkbias bhat2wrkbias
that1wrkbias that2wrkbias;
    by alpha;
  run;
  DATA fgmtable;
    SET stats;
    b1biasimp=(abs(bhat1wrkbias)-abs(bhat1bias))/abs(bhat1wrkbias)*100;
    b2biasimp=(abs(bhat2wrkbias)-abs(bhat2bias))/abs(bhat2wrkbias)*100;
    t1biasimp=(abs(that1wrkbias)-abs(that1bias))/abs(that1wrkbias)*100;
    t2biasimp=(abs(that2wrkbias)-abs(that2bias))/abs(that2wrkbias)*100;
    b1mseimp=(bhat1wrkmse-bhat1mse)/bhat1wrkmse*100;
    b2mseimp=(bhat2wrkmse-bhat2mse)/bhat2wrkmse*100;
    t1mseimp=(that1wrkmse-that1mse)/that1wrkmse*100;
```

```
t2mseimp=(that2wrkmse-that2mse)/that2wrkmse*100;
keep b1biasimp b2biasimp t1biasimp t2biasimp
     b1mseimp b2mseimp t1mseimp t2mseimp alpha;
RUN;
proc print data=stats;
     var alpha alphat bhat1 bhat2 that1 that2 bhat1wrk bhat2wrk that1wrk
that2wrk ;
run;
proc print data=fgmtable;
run;
quit;
```

### A.4 Example: DMBA-Induced Tumors

```
DATA data1;
INPUT Treatment$ x1 x2 censor1 censor2;
z=X2-X1; k=(z=0); n=1;
DATALINES;
Cont    78    99    0 0
Cont    41    54    0 0
⋮
EGCG    33    87    0 0
EGCG    73    115    0 0
```

```
⋮
    Res     83     118     0 0
⋮
    Res     78     92     0 0
    ;
    PROC SORT DATA=data1;
    BY treatment;
    RUN;
    PROC MEANS data=data1 noprint;
    VAR x1 x2 n k censor1 censor2; BY Treatment;
    OUTPUT out=stats sum=x1 x2 n k censor1 censor2;
    RUN;
    DATA stats; set stats;
    DROP _freq_ _type_;
    lambda11=(n-censor2)/x2 + (n-k)/(x1);
    lambda12=(n-censor1)/x1;
    lambda2=(n-censor2)/x2;
    se11=sqrt((lambda11*(lambda11-lambda2) + lambda2**2)/n);
    se12=sqrt(lambda12**2/n);
    se2=sqrt(lambda2**2/(n-censor2));
    cov=lambda2**2/n;
    corr=cov/(se11*se2);
```

```
rhohat1=lambda2/lambda11;
rhohat2=lambda2/lambda12;
lb1=lambda11-2*se11; ub1=lambda11+2*se11;
lb2=lambda12-2*se12; ub2=lambda12+2*se12;
lb3=lambda2-2*se2; ub3=lambda2+2*se2;
RUN;
PROC PRINT; RUN;
/*PROC LIFETEST data=data1 plots=(s,ls, lls);
TIME tum1;
strata treatment;
OUTSURV OUT=tum1;
RUN;
PROC LIFETEST data=data1 plots=(s,ls, lls);
TIME tum2*delta(1);
strata treatment;
OUTSURV OUT=tum2;
RUN;
PROC SORT DATA=data1;
BY treatment;
RUN;
PROC NLP tech=newrap DATA=data1 OUTEST=init ;
MAX logf;
```

```
PARMS Lam1=1, b=1;
logf=log(b) + log(Lam1)+
+(b-1)*log(tum1)
-Lam1*tum1**b ;
BY treatment;
RUN;
PROC NLP tech=newrap DATA=data1 OUTEST=init ;
MAX logf;
PARMS lam1=1, Lam2=1, b=1;
logf=log(b) + log(Lam1)+
+(b-1)*log(tum1)
-Lam1*tum1**b + (1-delta)*log(b) + (1-delta)*log(Lam2)+
+(b-1)*(1-delta)*log(tum2)
-(1-delta)*Lam2*tum2**b -delta*lam2*tum2**b;
BY treatment;
RUN;
DATA data2; set data1;
IF treatment ='Cont' THEN DO; y1=tum1**3; y2=tum2**3; END;
IF treatment ='Res' THEN DO; y1=tum1**3; y2=tum2**4; END;
IF treatment ='EGCG' THEN DO; y1=tum1**3; y2=tum2**4; END;
RUN;
PROC LIFETEST data=data2 plots=(s,ls, lls);
```

```
TIME y1;
strata treatment;
RUN;
PROC LIFETEST data=data2 plots=(s,ls, lls);
TIME y2*delta(1);
strata treatment;
RUN;
```