NOTE TO USERS

This reproduction is the best copy available.
RICE UNIVERSITY

Nonparametric Estimation of Bivariate Mean Residual Life Function

by

Musie S. Ghebremichael

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Approved, Thesis Committee:

Javier Rojo, Chair
Professor of Statistics

David W. Scott
Noah Harding, Professor of Statistics

Gary L. Rosner
Professor of Biostatistics

Richard A. Tapia
Noah Harding Professor of Computational
and Applied Mathematics

HOUSTON, TEXAS

MAY, 2004
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI Microform 3168086
Copyright 2005 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346
Abstract

Nonparametric Estimation of Bivariate Mean Residual Life Function

by

Musie S. Ghebremichael

In survival analysis the additional lifetime that an object survives past a time $t$ is called the residual life function of the object. Mathematically speaking if the lifetime of the object is described by a random variable $T$ then the random variable $R(t) = [T - t | T > t]$ is called the residual life random variable. The quantity $c(t) = E(R(t)) = E[T - t | T > t]$ is called the mean residual lifetime (mrl) function or the life expectancy at age $t$.

There are numerous situations where the bivariate mrl function is important. Times to death or times to initial contraction of a disease may be of interest for litter mate pairs of rats or for twin studies in humans. The time to a deterioration level or the time to reaction of a treatment may be of interest in pairs of lungs, kidneys, breasts, eyes or ears of humans. In reliability, the distribution of the lifelengths
of a particular pair of components in a system may be of interest. Because of the
dependence among the event times, we can not get reliable results by using the
univariate mrl function on each event times in order to study the aging process. The
bivariate mrl function is useful in analyzing the joint distribution of two event times
where these times are dependent.

In recent years, though considerable attention has been paid to the univariate mrl
function, relatively little research has been devoted to the analysis of the bivariate
mrl function. The specific contribution of this dissertation consists in proposing,
and examining the properties of, nonparametric estimators of the bivariate mean
residual life function when a certain order among such functions exists. That is, we
consider the problem of nonparametric estimation of a bivariate mrl function when it
is bounded from above by another known or unknown mrl function. The estimators
under such an order constraint are shown to perform better than the empirical mrl
function in terms of mean squared error. Moreover, they are shown to be projections,
onto an appropriate space, of the empirical mean residual life function. Under suitable
technical conditions, the asymptotic theory of these estimators is derived. Finally,
the procedures are applied to a data set on bivariate survival. More specifically, we
have used the Diabetic Retinopathy Study (DRS) data to illustrate our estimators.
In this data set, the survival times of both left and right eyes are given for two groups
of patients: juvenile and adult diabetics. Thus, it seems natural to assume that the
mrl for the juvenile diabetics be longer than the mrl of the adult diabetics. Under
this assumption, we calculated the estimators of the mrl function for each group.

We have also calculated the empirical mrl functions of the two groups and compared them with the estimators of the mrl function obtained under the above assumption.
Acknowledgements

First, I would like to acknowledge God for His grace and mercy in blessing me with family, friends, and mentors whose help made it possible for me to complete my graduate studies.

I wish to express my appreciation to Professor Javier Rojo both for his guidance and help as well as his enlightening comments in the preparation of this research. The time he devoted to discussions relative to the work has proved invaluable. His encouragement was inspirational and his enthusiasm contagious. During the course of this research, he carried me on his shoulders of patience and expertise. His friendship, advice and humble personality will be long remembered.

I would also like to express my gratitude to Professor Richard Tapia for the several illuminating discussions that we had on research, academic career, diversity and other issues related to graduate study. Had it not been for the financial assistance provided to me through his help, this research would not have been possible. Sincere thanks to Professor David Scott and Professor Gary Rosner for their encouragement and willingness to be in my thesis committee.

Finally, I would like to thank my family for all their love, support, sacrifice and struggle that has allowed me to achieve this level in my education. They started me on the path of higher education and this dissertation is a testament to the constant support they provided me. I will always cherish their love and affection. It is with
great fondness that I remember the times I spent with my friend Clayton Earl Simien and all members of the Houston Eritrean Emmanuel church during my stay at Rice University.
Contents

Abstract ii

Acknowledgements v

List of Figures xi

List of Tables xiv

1 Introduction 1

1.1 Distribution Function 3

1.2 Introduction to Survival Concepts 5

1.2.1 Survival and hazard functions 6

1.2.2 Survival analysis and censoring 7

1.2.3 Kaplan-Meier and Nelson-Aalen estimators 11

2 The Mean Residual Life Function 14

2.1 Definition, Properties and Characterizations 16

2.2 MRL and Hazard Functions 20
2.3 The Empirical MRL Function ............................................ 24
   2.3.1 Definition ......................................................... 24
   2.3.2 Asymptotic properties of the empirical mrl estimator ....... 26
2.4 The Proportional MRL Model ............................................. 28
   2.4.1 Specification and existence criterion of the model ........... 28
   2.4.2 Relation between PH and PMRL models ...................... 29
2.5 Estimation of the MRL Function under Censoring ................. 32
2.6 The MRL Function of Discrete Random Variables .................. 34

3 Bivariate Mean Residual Function ...................................... 35
3.1 Definition, Properties and Characterization ....................... 36
   3.1.1 Definition of the bivariate mrl function .................... 38
   3.1.2 Relationship between the bivariate mrl and survival functions 41
   3.1.3 Necessary and sufficient conditions for a bivariate function to
          be a bivariate mrl function .................................... 45
3.2 The Empirical Bivariate MRL Function ............................. 46
   3.2.1 Definition ....................................................... 46
   3.2.2 Asymptotic properties ......................................... 48
3.3 Tests and Confidence Sets for MRL Functions ..................... 55
3.4 Estimating the Bivariate MRL Function under Censoring ........ 58
3.5 Simulation Results .................................................... 60
4 Estimation of Two Bivariate Ordered MRL Functions

4.1 Introduction .................................. 64

4.2 The One Sample Problem ..................... 66
  4.2.1 The estimator under mrl ordering ............. 66
  4.2.2 Asymptotic properties of the estimator .......... 69

4.3 The Two Sample Problem .................... 76
  4.3.1 The estimator under mrl ordering ............. 76
  4.3.2 Asymptotic properties of the estimators ........ 81

4.4 The Mirror Image of the One Sample Problem .... 87
  4.4.1 The estimator under mrl ordering ............. 87
  4.4.2 Asymptotic properties of the estimator .......... 89

4.5 The Censored Data Case ..................... 93

4.6 Tests and Confidence Sets for MRL Functions .... 96

4.7 Simulation Results .......................... 97

5 Application to a Real Data Set .......................... 100

5.1 AML Maintenance Study ..................... 101

5.2 Bladder Cancer Data Set ..................... 103

5.3 The Diabetic Retinopathy Data ................ 110

6 Conclusions .................................. 115

6.1 Summary .................................. 115
6.2 Extension to the Multivariate Case .................. 116

A Properties of The Kaplan-Meier Estimator .......... 119

B Definition of Weak Convergence and Gaussian Processes 123
   B.1 Weak Convergence and Tightness .................. 123
   B.2 Gaussian Processes ................................ 126

C Simulation Results for Chapter Three ............... 128

D Simulation Results for Chapter Four ................ 132

Bibliography ........................................... 139
List of Figures

2.1 MRL function and hazard function of the uniform distribution with probability mass \([a, b]\). .............................................. 23

3.1 (a) Gumbel density with \(\theta = 0.5\); (b) Pareto density with \(a = 3, \theta_1 = 0.8, \theta_2 = 0.4\); (c) Morgenstern density with \(\alpha = 1\) and (d) Sarmanov density with \(\alpha = -1\). .............................................. 62

5.1 Kaplan-Meier curves for the AML study. ............................. 102

5.2 Mean residual functions for the AML study. .......................... 102

5.3 Surface plots for the empirical mrl function for patients on the placebo (top), pyridoxine (center) and thiotepa (bottom) treatments. ....... 105

5.4 Empirical survival curves for DRS data. .............................. 112

5.5 Empirical mrl function for the DRS data: juvenile (top) and adult (bottom). ..................................................... 113

5.6 Estimators of the mrl function under mrl ordering for the DRS data: juvenile (top) and adult (bottom). ............................ 114
C.1 Mean squared error for the empirical mrl function in the case of bivariate Gumbel distribution ($\theta = 1$). 129

C.2 Mean squared error for the empirical mrl function in the case of Morgenstern distribution ($\alpha = 0.5$). 130

C.3 Mean squared error for the empirical mrl function in the case of Sarmanov distribution ($\alpha = 0.6$). 131

D.1 Ratios of mean squared errors of the restricted estimator to the empirical estimator for two Gumbel distributions ($\theta_1 = 1, \theta_2 = 0.5$): (Left)-one sample problem, (Right)-mirror image problem. 133

D.2 Ratios of mean squared errors of the restricted estimator to the empirical estimator for two Morgenstern distributions ($\alpha_1 = 0.5, \alpha_2 = 0.6$): (Left)-one sample problem, (Right)-mirror image problem. 134

D.3 Ratios of mean squared errors of the restricted estimator to the empirical estimator for two Sarmanov distributions ($\alpha_1 = 0.5, \alpha_2 = 0.6$): (Left)-one sample problem, (Right)-mirror image problem. 135

D.4 Two sample problem: Ratios of mean squared errors of the restricted estimator to the empirical estimator for two Gumbel distributions ($\theta_1 = 1, \theta_2 = 0.5$): (Left)-minimum estimator, (Right)-maximum estimator. 136
D.5 Two Sample Problem: Ratios of mean squared errors of the restricted estimator to the empirical estimator for two Morgenstern distributions 

\( (a_1 = 10, a_2 = 5) \): (Left)-minimum estimator, (Right)-maximum estimator. .................................................. 137

D.6 Two Sample Problem: Ratios of mean squared errors of the restricted estimator to the empirical estimator for two Sarmanov distributions

\( (a_1 = 10, a_2 = 5) \): (Left)-minimum estimator, (Right)-maximum estimator. .................................................. 138
List of Tables

5.1 Data for AML maintenance study ........................................ 101
5.2 Lifetimes in months for bladder tumor patients undergoing placebo pills treatment ............................................... 104
5.3 Lifetimes in months for bladder tumor patients undergoing pyridoxine treatment ............................................... 106
5.4 Lifetimes in months for bladder tumor patients undergoing thiotepa treatment ............................................... 106
5.5 The empirical mrl function for patients on the thiotepa treatment .................................................. 107
5.6 The empirical mrl function for patients on the pyridoxine treatment .................................................. 108
5.7 The empirical mrl function for patients on the placebo treatment .................................................. 109
5.8 Survival times (months) for adult diabetics. ........................................ 111
5.9 Survival times (months) for juvenile diabetics. ........................................ 112
Chapter 1

Introduction

The mean residual life (mrl) function or remaining life expectancy at age $x$ is defined as the expected remaining life given survival to age $x$. It is a concept of obvious interest and, indeed, one of the most important notions in actuarial, reliability and survivorship studies. Our intent in this research is to provide some estimators of the bivariate mrl function under a certain order among mean residual life functions and study the asymptotic properties of the estimators. Chapter 1 provides a brief review of some basic mathematical and statistical concepts required for a good understanding of the remaining chapters. Chapter 2 discusses some general results on the univariate mrl function and Chapter 3 contains extensions of the results to cover the bivariate case. That is, Chapter 3 presents a review of the bivariate mrl function and the asymptotic properties of the bivariate empirical mrl function. Chapter 4 explores the problem of estimating the bivariate mrl function when a certain order among
mean residual life functions obtains. Chapter 5 illustrates the estimators presented in Chapters 3 and 4 by applying them to data from the biomedical sciences. Finally, Chapter 6 presents remarks which summarize the work and a multivariate extension of the problem considered in Chapter 4. In an effort to make this work as self-contained as possible, there are four appendices. The first appendix discusses some properties of the Kaplan-Meier estimator. The second appendix presents the main results on weak convergence which will be used in this thesis. The last two appendices contain simulation results for the estimators presented in Chapter 3 and Chapter 4.

The rest of this section presents a summary of the principal abbreviations and notation followed in the text.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.s.</td>
<td>almost surely</td>
</tr>
<tr>
<td>d.f.</td>
<td>distribution function</td>
</tr>
<tr>
<td>iff</td>
<td>if and only if</td>
</tr>
<tr>
<td>i.i.d.</td>
<td>independent and identically distributed</td>
</tr>
<tr>
<td>p.d.f.</td>
<td>probability density function</td>
</tr>
<tr>
<td>p.s.d.</td>
<td>positive semi-definite</td>
</tr>
<tr>
<td>r.v.</td>
<td>random variables (vectors)</td>
</tr>
<tr>
<td>w.p.1</td>
<td>with probability one</td>
</tr>
</tbody>
</table>

Let $r$ and $s$ be real numbers. Then $r \wedge s$ and $r \lor s$ denote respectively $\min(r, s)$ and $\max(r, s)$. In general, boldface lowercase (uppercase) characters denote vectors
(matrices) and $x'$ denotes the transpose of $x$. Also, $tr(A)$ and $|A|$ denote the trace and the determinant of a matrix $A$. The symbols $\Phi$ and $\varphi$ are reserved for the d.f. and the p.d.f. of the standard normal distribution; we also let $X \sim \mathcal{N}(a, b)$ indicate that a random variable $X$ follows the normal distribution with mean $a$ and variance $b$. Convergence in distribution is denoted either by $\Rightarrow$ or $\Rightarrow_D$. The symbol $C[0, \infty)$ denotes the collection of real-valued continuous functions defined on the domain $[0, \infty)$ and $D[0, \infty)$ denotes the collection of cad-lag functions (functions which are right-continuous and have left-hand limits) on $[0, \infty)$. For vectors $a = (a_1, a_2)$ and $b = (b_1, b_2)$, $a \leq b$ will denote the order in $\mathbb{R}^2$ defined by $a_1 \leq b_1$ and $a_2 \leq b_2$. Other notation will be introduced as the need arises.

### 1.1 Distribution Function

A function $F: \mathbb{R} \to [0, 1]$ is called a d.f. if it is right-continuous and monotone non-decreasing with $F(-\infty) = 0$, $F(+\infty) = 1$. For a random sample $X_1, \ldots, X_n$ from a distribution $F$, the empirical distribution function is denoted by $F_n(x)$ and is defined as the proportion of sample values that do not exceed the number $x$ for all real numbers $x$. Thus $F_n(x)$ is a step function which increases by the amount $\frac{1}{n}$ at its jump points, which are the order statistics of the sample. That is,

$$F_n(x) = n^{-1} \sum_{j=1}^{n} I_{\{X_j \leq x\}}, \ x \in \mathbb{R}.$$  \hspace{1cm} (1.1.1)
For each fixed sample, $F_n$ is a distribution function when considered as a function of $x$. For every fixed $x \in \mathbb{R}$, when considered as a function of $X_1, X_2, \ldots, X_n$, $F_n(x)$ is a random variable; in this context, since the $I_{\{X_i \leq x\}}$, $i = 1, \ldots, n$, are i.i.d. zero-one valued random variables, the random variable $nF_n(x)$ is the sum of $n$ independent Bernoulli random variables, which follows the Binomial distribution with parameter $F(x)$. Therefore, for the discrete random variable $F_n(x)$ which is the empirical distribution function of a random sample $X_1, \ldots, X_n$ from a d.f. $F(x)$ we have

$$P \left[ F_n(x) = \frac{j}{n} \right] = \binom{n}{j} [F_X(x)]^j [1 - F_X(x)]^{n-j}$$  \hspace{1cm} (1.1.2)

for $j = 0, \ldots, n$ and with $E[F_n(x)] = F(x)$ and $\text{Var}[F_n(x)] = \frac{F(x)(1-F(x))}{n}$. It is well-known that $F_n(x)$ is a consistent estimator of $F(x)$, or, in other words, $F_n(x)$ converges with probability one to $F(x)$. This strong pointwise convergence follows from the law of large numbers, and as it turns out the strong convergence is uniform on $(-\infty, \infty)$ (Glivenko-Cantelli Theorem) as denoted below:

$$\sup_{-\infty < x < \infty} |F_n(x) - F(x)| \xrightarrow{a.s.} 0.$$  \hspace{1cm} (1.1.3)

The proof of (1.1.3) follows easily from the Borel-Cantelli lemma and the results of Kiefer et al (1956) who proved that there exists a universal constant $C$ such that, for all $n > 0$ and $r > 0$,

$$P \left\{ \sup_{-\infty < x < \infty} |\sqrt{n}[F_n(x) - F(x)]| \geq r \right\} \leq C \exp(-2r^2).$$  \hspace{1cm} (1.1.4)

After several efforts by various authors, Massart (1990) showed that the best possible $C$ in (1.1.4) is 2. Kiefer and Wolfowitz (1958) extended the above result to $\mathbb{R}^m$ for
$m > 1$. They showed that for each $m$ there exist positive constants $c_0$ and $c$ such that, for all $n$, all $F$, and all positive $r$, \( P\{\sup_{x \in \mathbb{R}^m} |\sqrt{n}[F_n(x) - F(x)]| \geq r\} < c_0 e^{-cr^2} \). The constants $c_0$ and $c$ in general depend upon $m$.

Another useful property of the empirical distribution is its asymptotic normality. Observing that $\text{Var}F_n(x) \to 0$ as $n \to \infty$ and using the Central Limit Theorem we may conclude that $\sqrt{n}[F_n(x) - F(x)] \overset{D}{\to} N(0, F(x)[1 - F(x)])$ for each fixed $x$. Moreover, $\{\sqrt{n}(F_n(x) - F(x)), -\infty < x < \infty\}$ converges weakly to a Gaussian process. For brief discussions on weak convergence and Gaussian processes, the reader is referred to Appendix B.

### 1.2 Introduction to Survival Concepts

Survival analysis encompasses a variety of statistical techniques for analyzing positive-valued random variables. Typically the value of the random variable is the time to death of a biological unit (patient, animal, cell, etc.) or the time to failure of a physical component (mechanical or electrical). However, the time of interest could be the time to learning a skill, or it may not even be a time at all. For example, it could be the number of dollars that a health insurance company pays in a particular case.

While the origins of survival analysis might be attributed to the early work on mortality tables centuries ago, the more modern era started about a half century ago with applications to engineering. World War II stimulated interest in the reliability
of military equipment, and this interest in reliability carried over into the postwar era for military and commercial products. Most of the statistical research for engineering applications was concentrated on parametric models. The second half of the 20th century experienced an increase in the number of clinical trials in medical research, and a corresponding increase in nonparametric approaches to analyzing survival data. Our research will focus on nonparametric methods in survival analysis.

1.2.1 Survival and hazard functions

Let $T$ be a nonnegative random variable with density function $f(t)$ and distribution function $F(t)$. The survival function $S(t)$ is defined by

$$S(t) = 1 - F(t) = P \{T > t\} \quad (1.2.1)$$

and the hazard rate or hazard function $h(t)$ is defined by

$$h(t) = \frac{f(t)}{S(t)}. \quad (1.2.2)$$

The function $h(t)$ has been called the instantaneous force of mortality in epidemiology because of the following expression

$$h(t) = \lim_{\Delta t \to 0} \frac{P\{\text{dying in } (t, t + \Delta t) \mid \text{alive at time } t\}}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{P\{t \leq T < t + \Delta t \mid T > t\}}{\Delta t}$$

$$= \frac{f(t)}{S(t)}.$$

Thus, $h(t)$ expresses the instantaneous force of mortality at time $t$. Integrating $h(t)$,

$$\int_0^t h(u)du = \int_0^t \frac{f(u)}{S(u)}du = -\ln S(t)$$

which leads to the important expression $S(t) =$
\exp \left(- \int_0^t h(u)du \right).

Let \( \{T_n, n \geq 1\} \) be independent random variables (representing the lifetimes of individuals) with common distribution \( F \). Suppose that \( F \) is unknown and on the basis of the sample \( T_1, T_2, \ldots, T_n \) we seek to estimate the survival function. In the absence of any additional information about \( F \), a reasonable estimator is the empirical survival function defined by

\[
S_n(t) = \frac{1}{n} \sum_{i=1}^{n} I_{\{T_i > t\}} \quad \forall t \geq 0. \tag{1.2.3}
\]

Clearly, the finite-sample and asymptotic distribution theories for the empirical survival function are inherited from the corresponding theories for the empirical distribution function. Thus, \( S_n(t) \to S(t) \) a.s. as \( n \to \infty \). In fact the convergence is uniform in \( t \), that is, \( \sup_t |S_n(t) - S(t)| \to 0 \) a.s. as \( n \to \infty \). By the Central Limit Theorem we get that for each fixed \( t \), \( \sqrt{n} [S_n(t) - S(t)] \overset{D}{\to} N(0, S(t)[1 - S(t)]) \).

We have also the following stronger result: \( \{\sqrt{n} (S_n(t) - S(t)) - \infty < t < \infty\} \) converges weakly to a zero mean Gaussian process with covariance function given by \( \min \{S(x), S(y)\} - S(x)S(y) \), for \( 0 \leq x < y < \infty \).

### 1.2.2 Survival analysis and censoring

What distinguishes survival analysis from other areas in statistics is censoring. Informally, a censored observation contains partial information about the random variable of interest. We consider three types of censoring. Let \( T_1, \ldots, T_n \) be i.i.d. with d.f. \( F \).
**Type I Censoring:** Let $t_c$ be some (preassigned) fixed number which we call the fixed censoring time. Instead of observing $T_1, \ldots, T_n$ (the random variables of interest) we only observe $Y_1, \ldots, Y_n$ where

$$Y_i = \begin{cases} T_i & \text{if } T_i \leq t_c \\ t_c & \text{if } t_c < T_i. \end{cases}$$

Note that the distribution function of $Y$ has positive mass $P\{T > t_c\} > 0$ at $y = t_c$.

**Type II Censoring:** Let $r < n$ be fixed, and let $T_{(1)}, \ldots, T_{(n)}$ be the order statistics of $T_1, \ldots, T_n$. Observation ceases after the $r^{th}$ failure so we only observe $T_{(1)}, \ldots, T_{(r)}$. The full ordered sample is

$$Y_{(i)} = \begin{cases} T_{(i)} & \text{if } i = 1, 2, \ldots, r \\ T_{(r)} & \text{if } i = r + 1, r + 2, \ldots, n. \end{cases}$$

**Random Censoring:** Let $C_1, C_2, \ldots, C_n$ be i.i.d. with distribution function $G$. Let $C_i$ be the censoring time associated with $T_i$. It is assumed that $C_i$ is independent of the $T_j$’s and independent of $C_j$’s. The data obtained consists of $(Y_1, \delta_1), \ldots, (Y_n, \delta_n)$ where

$$Y_i = T_i \wedge C_i$$

$$\delta_i = \begin{cases} 1 & \text{if } T_i \leq C_i \text{ (Uncensored)} \\ 0 & \text{if } T_i > C_i \text{ (Censored)}. \end{cases}$$

Note that $Y_1, \ldots, Y_n$ are i.i.d. with some survival function $\bar{H} = (1 - G)(1 - F)$.

Both Type I and Type II censoring arise in engineering applications. In such situations there is, for example, a batch of transistors or tubes; we put them all on test at $t = 0$, and record their times to failure. Some transistors may take a long
time to burn out, and we will not want to wait that long to stop the experiment. Therefore, we might stop the experiment at a specified time $t_e$, in which case we have Type I censoring, or we might decide to wait until a prespecified fraction $F_n$ of the transistors has burned out, in which case we have Type II censoring.

Random censoring arises in medical applications. In a clinical trial, patients may enter the study at different times; then each patient is treated with one of several possible therapies. We want to observe their lifetimes, but censoring occurs in one of the following forms:

- *Lost to follow-up.* The patients may decide to move elsewhere; we never see them again.

- *Drop out.* The therapy may have such bad side effects that it is necessary to discontinue the treatment. Or the patient may still be in contact (he/she has not moved), but refuses to continue the treatment.

- *Termination of the study.*

Henceforth, it will be assumed that $T_i$ and $C_i$ are independent. The assumption seems justified when entries to the study are random in time, or when losses to follow-up occur randomly. However, if the reason for dropping out is related to the course of the therapy, there may be dependence between $T_i$ and $C_i$. If $T_i$ and $C_i$ are dependent, it will be assumed that $P\{t \leq T < t + dt | T \geq t\} = P\{t \leq T < t + dt | T \geq t, C \geq t\}$. Note that this assumption is weaker than the independent censoring assumption and
is necessary for the distribution of $T$ to be identifiable in inference from observations on $\{\min(T_i, C_i), I(T_i \leq C_i)\}$.

The previous types of censoring fall under the heading of right censoring: if the random variable of interest is too large, we do not get to observe it. There is also left censoring. For example, in random left censoring, we can only observe $(Y_1, \varepsilon_1), \ldots, (Y_n, \varepsilon_n)$ where $Y_i = \max(T_i, C_i)$ and $\varepsilon_i = I(C_i \leq T_i)$. Both left and right censoring may occur in the same data set. Suppose a biologist wants to know the time from injection of a carcinogen until the onset of tumors in certain animals. The data is obtained by killing the animal and observing if tumors are present at the time of death. If tumors are present at the time of death, the onset time is smaller than the time of death; the observation is left-censored. But if tumors are not present at the time of death, the onset time is larger than the time actually observed, the observation is right-censored.

Both right and left censoring are special cases of interval censoring. Interval censoring mechanisms arise when the event of interest cannot be directly observed and it is only known to have occurred during a possibly random interval of time. For example, two examinations at different times to see if a certain event has occurred yield a censored observation of that particular event. Whether the observation is left-censored, right-censored or interval-censored depends on whether the event occurred before the first examination, after the second examination and between the two examinations. The literature dealing with interval censoring is vast. In the context of
AIDS survival data, the reader is referred to the following works and the references therein: Gruttola and Lakagos (1989) and Gentleman and Geyer (1994).

In contrast to interval censoring there is truncation in which the random variable of interest falls outside some interval and even its existence is unobserved. For example, suppose we want to get the distribution and expected size of a certain organelle in the cell. Because of limitations on the measuring equipment, if an organelle is below a certain size it can not be detected and hence its size is taken to be zero. Qin and Wang (2001) discussed the analysis of truncated data.

Henceforth, attention will be focused on the case of independent random censoring from the right.

1.2.3 Kaplan-Meier and Nelson-Aalen estimators

Let $X_1, \ldots, X_n$ be i.i.d. failure times with d.f. $F$ and let $C_1, \ldots, C_n$ be i.i.d. censoring times with d.f. $G$. Suppose that $(C_1, \ldots, C_n)$ is independent of $(X_1, \ldots, X_n)$. The data which we observe consists of the pairs $(T_i, \delta_i)$, $i = 1, \ldots, n$ where $T_i = \min(X_i, C_i)$ and $\delta_i = I(X_i \leq C_i)$ is the indicator function which takes the values 1 and 0 depending on whether $X_i \leq C_i$ or $X_i > C_i$. Suppose that $t_1 < t_2 < \ldots < t_k$ ($k \leq n$) are the distinct times at which the deaths occur. The possibility of there being more than one death at $t_j$ is allowed, and we let $d_j$ represent the number of deaths at $t_j$. Let us denote the number of individuals who are at risk just prior to $t_j$ by $n_j$.

The Kaplan-Meier (1958) estimator (hereafter referred to as K-M) estimates the
survival function in the presence of censoring and is given by

\[ S_{KM}(t) = \prod_{j: t_j \leq t} \left(1 - \frac{d_j}{n_j}\right) \]

\[ = \prod_{k=1}^{j} \left(1 - \frac{d_k}{n_k}\right), \text{ for } t_j \leq t < t_{j+1}. \]  

A detailed explanation of the Kaplan-Meier estimator together with its large-sample properties is given in Appendix A.

Estimating the hazard function \( h(t) \) is essentially equivalent to the problem of estimating a density. An easier problem is estimating the cumulative hazard function which is defined as \( H(t) = \int_0^t h(u)du \). Thus, a natural estimate of \( H(t) \) is

\[ \hat{H}(t) = -\log S_{KM}(t). \]  

An alternate estimate of \( H(t) \) is the Nelson-Aalen estimator which is sometimes called the empirical (cumulative) hazard function. The idea is simple: if at time \( t_j \), \( d_j \) individuals out of a risk set of \( n_j \) die, that means the hazard at time \( t_j \) is \( \frac{d_j}{n_j} \). Cumulatively, sum these from time 0 to time \( t \) to get the cumulative hazard rate. The formula is:

\[ \hat{H}(t) = \sum_{j: t_j \leq t} \frac{d_j}{n_j} \]

\[ = \sum_{k=1}^{j} \frac{d_k}{n_k}, \text{ for } t_j \leq t < t_{j+1}. \]

The plots of \( \hat{H}(t) \) and \( \hat{H}(t) \) are often very useful. For example, it is much easier to see from such plots whether a life distribution might have a constant, decreasing, or increasing hazard function than from plots of \( \hat{S}(t) \).
For continuous models, $\hat{H}(t)$ and $\bar{H}(t)$ are asymptotically equivalent and do not differ greatly in most situations, except for large values of $t$. Since $\hat{H}(t) = -\log S_{KM}(t)$, we have

\[
\hat{H}(t) = -\log \prod_{j: t_j \leq t} \left(1 - \frac{d_j}{n_j}\right)
\]

\[
= -\sum_{j: t_j \leq t} \log \left(1 - \frac{d_j}{n_j}\right)
\]

\[
= \sum_{j: t_j \leq t} \left(\frac{d_j}{n_j} + \frac{d_j^2}{2n_j^2} + \cdots\right)
\]

and thus $\hat{H}(t)$ is a first-order approximation to $\bar{H}(t)$. An estimate for the asymptotic variance of $\hat{H}(t)$ is obtained by applying the $\delta$-method to $\hat{H}(t) = -\log S_{KM}(t)$ to obtain

\[
\text{Var}[\hat{H}(t)] = \text{Var}[S_{KM}(t)] \cdot \frac{S_{KM}(t)}{S_{KM}^2(t)}.
\]

This is an estimate for the asymptotic variance of $\hat{H}(t)$ as well, since $\bar{H}(t)$ and $\hat{H}(t)$ are asymptotically equivalent. Because of the simple relationship of $H(t)$ to $S(t)$, a lengthy discussion of $H(t)$ and its estimates is not required. The asymptotic properties of $\hat{H}(t)$, and of course of $\bar{H}(t)$, are discussed by Breslow and Crowley (1974).
Chapter 2

The Mean Residual Life Function

The concept of mean residual life (mrl) has been of much interest in the actuarial sciences, survival studies and reliability theory. A literature survey by Guess and Proschan (1985) traces its history to the third century A.D. It has been used extensively to model the 'burn-in' problem in reliability theory. Watson and Wells (1961) explore improving the mrl of a batch of articles (e.g. light bulbs, transistors) by running them, for some time, under realistic conditions. Defective articles fail early on and are eliminated, leaving the rest of the batch with a longer mrl. They found conditions on the life distribution of the original articles to ensure that the eliminated items have shorter mean remaining lifetimes. The concepts of increasing mean residual life, decreasing mean residual life, new better than used etc., have been proven to be useful in the social sciences to analyze duration of wars and strikes. For example, the papers by Guess and Proschan (1988) and Morrison (1978) present an application

14
of the mrl function in the social sciences. The paper by Swartz (1973) provides the necessary and sufficient conditions for a function to be an mrl function. Yang (1978) proposed a nonparametric estimator of the mrl function and studied its convergence properties on a closed interval. Hall and Wellner (1979) have extended her results to $\mathbb{R}_+$. Oakes and Dasu (1990) as well as Arnold and Zahedi (1991) proposed the proportional mean residual life model. They stated that two survivor functions $S_0(x)$ and $S_1(x)$ are said to have proportional mean residual life if $e_1(x) = \theta e_0(x)$, for $\theta > 0$, where $e_0(x)$ and $e_1(x)$ are the mrl functions which correspond to $S_0(x)$ and $S_1(x)$ respectively. The proportional mrl model was extended to a regression model with explanatory variables by Maguluri and Zhang (1994). Yuen, Zhu and Tang (2001) developed goodness of fit tests for the proportional mrl regression model.

In this chapter our aim is to give an extensive review of mean residual life functions for univariate real-valued random variables. Since the concept of the mrl function was often introduced in connection with aging processes (particularly in its application in reliability and survival analysis), most of the literature on the subject is centered around nonnegative continuous random variables. This chapter contains five sections. Section 2.1 is devoted to the definition, properties and characterization of mrl functions (i.e., the characterization of the survival function in terms of mrl function). One of the most important properties of the mrl function is the fact that it uniquely determines the corresponding probability distribution function. In Section 2.2 we will investigate the relationship between the hazard and the mrl function. Section 2.3
deals with the empirical mrl function and its asymptotic properties. Section 2.4
discusses the proportional mrl model and Section 2.5 considers the estimation of the
mrl function in the presence of censoring. Section 2.6 will discuss the mrl function
of a discrete random variable.

2.1 Definition, Properties and Characterizations

Let $T$, for example, represent the time to failure of a component, or the repair time
of a component, or the time required to perform a service, or the waiting time in a
queue. Thus, in typical contexts, the dimension of $T$ is time. Given that a unit is of
age $t$, the remaining life after time $t$ is random. The expected value of this random
residual life is called the mean residual life (mrl) at time $t$. Since the mrl is defined
for each time $t$, we also speak of the mrl function. Henceforth, it will be assumed
that $e(0) = E(T) = \mu < \infty$, and that $T$ has a density $f(t)$. The expected residual
value, or mean residual value of $T$ given that $t$ units of time has elapsed, is denoted
by $e_T(t) = e(t)$ and is defined as

$$e(t) = E[T - t | T > t], \ t \geq 0$$

$$= \frac{\int_t^\infty S(u)du}{S(t)}$$

$$= \frac{\int_t^\infty uf(u)du}{S(t)} - t.$$ 

Note that when $T$ represents a lifetime, or time to failure of a component, then $e(0)$
is the expected life of a new component and $e(t)$ is the expected life remaining in a
component of age \( t \). Here age \( t \) means that the component has survived up to time \( t \) and is still operational at time \( t \). When \( T \) represents the repair time of a component, then \( e(t) \) is the expected time required to complete the repair of a component on which \( t \) units of repair time have already been expended. Hall and Wellner (1981) derived the variance of the residual lifetime:

\[
Var[T - t | T > t] = \frac{\int_t^\infty e^2(u)f(u)du}{S(t)}.
\]

In the theory of renewal processes, the mrl function arises naturally. For an equilibrium renewal process with survival function \( S \), the survival function of the forward recurrence time (residual life time) is given by \( S_R(t) = \frac{\int_0^\infty S(x)dx}{\mu} \). Thus, the failure rate of the residual life times is inversely related to the mrl of the inter-event distribution, i.e \( h_R(t) = 1/e(t) \). In general however, although it is the case that \( h(t)e(t) \rightarrow 1 \) as \( t \rightarrow \infty \), it is not true that \( h(t) = 1/e(t) \) for all \( t \).

Several classes of life distributions have been defined using the mrl function: (1) A life distribution \( F \) has decreasing mean residual life (DMRL) if its mrl function is decreasing, (2) A life distribution \( F \) is new better than used in expectation (NBUE) if \( e(0) \geq e(t) \) for all \( t \geq 0 \) and (3) A life distribution \( F \) has increasing then decreasing mean residual life (IDMRL) if there exists \( \tau \geq 0 \) such that \( e \) is increasing on \([0, \tau)\) and decreasing on \([\tau, \infty)\). The length of time employees stay with certain companies is a typical example of IDMRL class. An employee who has been with a company for ten years has more time and career invested in the company than an employee who has been with the company for only two months. The mrl of the ten-year employee
is likely to be longer than the mrl of the two-month employee. After this initial period of increasing MRL, the processes of aging and retirement take over to yield a DMRL period. Note that each of the classes defined above has an obvious dual class associated with it. That is, increasing mean residual life (IMRL), new worse than used in expectation (NWUE), and decreasing then increasing mean residual life (DIMRL), respectively. Guess and Proschan (1988) provides an excellent survey on the application of these classes.

The distribution function of $T$ may be expressed uniquely in terms of its survival function $S(t)$ and its hazard function $h(t)$. It is also possible to establish relationships between $e(t)$ and these functions. It is known that $e(t)$ provides a unique representation of the probability law of $T$. Write,

$$e(t) = \frac{\int_t^\infty S(u)du}{S(t)}$$

$$= \frac{\int_0^\infty S(u)du - \int_0^t S(u)du}{S(t)}$$

$$= \frac{1}{S(t)} \left[ \mu - \int_0^t S(u)du \right].$$

Multiplying (2.1.2) by $S(t)$ and differentiating with respect to $t$ yields $e'(t)S(t) + e(t)S'(t) = -S(t)$, and after arranging terms we get

$$\frac{1 + e'(t)}{e(t)} = h(t)$$

(2.1.3)

where $h(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)}$ is the hazard function. Since both $e(t)$ and $h(t)$ must be non-negative for all $t$ it follows from (2.1.3) that $e'(t) \geq -1$. The survival function is
expressed in terms of $h(t)$ as

$$S(t) = \exp \left[ - \int_0^t h(x) \, dx \right]$$

(2.1.4)

and substituting (2.1.3) in (2.1.4) we obtain

$$S(t) = \exp \left[ - \int_0^t \frac{1 + e'(u)}{e(u)} \, du \right].$$

Further, since $\int_0^t \frac{1 + e'(u)}{e(u)} \, du = \ln e(t) - \ln e(0) + \int_0^t \frac{du}{e(u)}$ we find

$$S(t) = \frac{e(0)}{e(t)} \exp \left[ - \int_0^t \frac{du}{e(u)} \right].$$

(2.1.5)

Equation (2.1.5) shows that $S(t)$ can be computed from $e(t)$, giving a unique representation of the probability law of $T$.

Swartz (1973) provided necessary and sufficient conditions for a non-negative function to be an mrl function. The conditions are given as follows:

(i) $0 \leq e(t) < \infty$.

(ii) $e'(t) \geq -1$.

(iii) $\int_0^t \frac{du}{e(u)} \to \infty$ as $t \to \infty$.

The first condition ensures that the mrl function is non-negative. Condition two ensures that the p.d.f. defined through $e, \frac{1 + e'(u)}{e(t)} e(0) \exp[ - \int_0^t 1/e(x) \, dx ]$, is non-negative.

The third condition is required so that the survival function corresponding to $e$ is decreasing.

Since $e(t)$ characterizes the probability law of $T$, it is, in principle, possible to determine any moment of $T$ from $e(t)$, provided that the moment exists. The first
moment is simply \( e(0) \). All other moments require the entire function \( e(t) \) for their determination. Starting with the definition of the \( n \)th moment

\[
E[T^n] = \int_0^\infty u^n f(u) du = n \int_0^\infty u^{n-1} S(u) du. \tag{2.1.6}
\]

But from equation (2.1.2) we know that \( \frac{d}{dx} [e(x)S(x)] = -S(x) \). Substituting this in (2.1.6) we get

\[
E[T^n] = -n \int_0^\infty u^{n-1} d[e(u)S(u)]
\]

and integrating by parts, we obtain

\[
E[T^n] = n(n-1) \int_0^\infty u^{n-2} e(u)S(u) du \quad \text{for } n \geq 2. \tag{2.1.7}
\]

In arriving at the above expression we have used the fact that \( \lim_{x \to \infty} x^{n-1} e(x)S(x) = 0 \) if the \( n \)th moment exists.

### 2.2 MRL and Hazard Functions

In this section we address the question of how well the process of "aging" or "wearout" in the context of time to failure is described by either the hazard function or the mrl function.

The mrl function gives a different picture of survival or aging than is seen through the more commonly studied survival or hazard rate functions. To illustrate a different perspective, consider a cancer patient undergoing chemotherapy begun 3 months ago
(t = 0 denote the beginning of the treatment). The expected remaining lifetime of
a patient who has survived the treatment for 3 months is e(3), whereas S(3) and
h(3) are the proportion of cancer patients who survive the first 3 months and the
instantaneous probability of dying tomorrow, respectively. An essential difference
between the hazard function and the mrl function is that the former accounts only
for the immediate future in assessing the event of component failure, whereas the
latter considers the entire future in providing the mean life after t. This is readily
seen if we multiply both h(t) and e(t) by S(t). Mathematically, one can represent
this as follows:

\[ h(t)S(t) = -\frac{d}{dt}S(t) \quad (2.2.1) \]
\[ e(t)S(t) = \int_{t}^{\infty} S(u)du. \quad (2.2.2) \]

The right side of (2.2.1) depends on the probability law at the point t only, whereas
the right side of (2.2.2) depends on the probability law at all points in \([t, \infty)\). This
explains why a component can experience aging at time t, in the sense that its mrl
is decreasing, and yet have zero hazard, because the failure event cannot occur in
the immediate future. Such a situation is exemplified by the uniform distribution in
\([a, b]\), for which the hazard function and the mrl function are shown in Figure 2.1.
This example is taken from Muth's (1977) paper. The mrl function decreases linearly
in \([0, a]\) with derivative -1, indicating that the expected life decreases at the same
rate at which time elapses. In a reliability context, the component is clearly wearing
out. The hazard function on the other hand, being zero, gives no indication of wear
out. In the range \([a, b]\) the mrl function decreases linearly with derivative \(-0.5\). The hazard function makes a jump at \(t = a\) and then grows as \(\frac{1}{b-t}\).

The mrl function provides a more descriptive measure of the aging process than the hazard function. In reliability theory, the advantages of the mrl function over the hazard function are clearly described by Muth (1977): "The advantages of using the mrl function as a decision making criterion for replacement policies, should be obvious. In these cases it is the expected residual life left in the component which gives us an indication of whether to replace, or to schedule maintenance. For example, we might establish a policy to replace a component when 80% of its expected useful life has been used up. In that case, the replacement time \(t_i\), for a component of type \(i\), with decreasing mrl function \(e(t)\), is determined by the equation \(\frac{e(t_i)}{e(0)} = 0.2\). It is hard to conceive that an equally meaningful and consistent criterion could be established with the hazard function as the yardstick".

A strictly decreasing mrl function represents positive aging at all times. An increasing hazard function implies a decreasing mrl function, and therefore represents positive aging also. Note that IHF (increasing hazard function) is stronger than the condition DMRL (decreasing mrl function), in the sense that IHF \(\Rightarrow\) DMRL, but DMRL \(\nRightarrow\) IHF.
Figure 2.1: MRL function and hazard function of the uniform distribution with probability mass \([a, b]\).
2.3 The Empirical MRL Function

2.3.1 Definition

In characterizing probability distributions from observed data, the sample mrl function is a useful tool. For example, the behavior of the empirical mrl at great age may suggest a corresponding tail behavior of the underlying survival function.

Let \( T_1, T_2, \ldots, T_n \) be i.i.d. random variables representing the lifetimes of individuals. Then the empirical mrl function at age \( t \), denoted by \( e_n(t) \), is the sample average of the residual lifetimes of those individuals whose lifetimes are larger than \( t \). That is

\[
e_n(t) = \frac{\sum_{i=1}^{n} (T_i - t) I(T_i > t)}{\sum_{i=1}^{n} I(T_i > t)} = \frac{\int_t^\infty S_n(u) du}{S_n(t)} I(t < T_{(n)}),
\]

(2.3.1)

where \( T_{(n)} = \max_{1 \leq i \leq n} T_i \) and \( S_n(t) \) is the empirical survival function which is given by \( S_n(t) = \frac{1}{n} \sum_{i=1}^{n} I(T_i > t) \). The empirical mrl function jumps at each \( t = T_i \) and decreases with a slope of 45° elsewhere.

Yang (1978) derived the expected value of the empirical mrl function. The univariate case can be handled easily by utilizing Zahedi’s (1985) results for the multivariate empirical mrl function:

\[
E[e_n(t)] = \sum_{j=1}^{n} E[e_n(t) | nS_n(t) = j] P(nS_n(t) = j)
\]

(2.3.2)

\[
= \sum_{j=1}^{n} E \left[ \int_t^\infty \frac{S_n(u) du}{S_n(t)} | nS_n(t) = j \right] P(nS_n(t) = j)
\]

\[
= \sum_{j=1}^{n} \frac{P(nS_n(t) = j)}{j} E \left[ \int_t^\infty nS_n(u) du | nS_n(t) = j \right]
\]
\[
= \sum_{j=1}^{n} \frac{P(nS_n(t) = j)}{n} \int_{t}^{\infty} E[nS_n(u) | nS_n(t) = j] du.
\]

From the definition of \( S_n(t) \) it follows that \( nS_n(t) \sim Bin (n, S(t)) \). Conditioned on \( \{nS_n(t) = j\} \), \( nS_n(u) \) for \( u > t \) is simply the number of individuals, out of \( j \), who survive past time \( u \). Thus, \( nS_n(u) | (nS_n(t) = j) \sim Bin \left( j, \frac{S(u)}{S(t)} \right) \), and therefore,

\[
E[nS_n(u) | nS_n(t) = j] = \frac{jS(u)}{S(t)}.
\]  \hspace{1cm} (2.3.3)

Substituting (2.3.3) in (2.3.2) we get

\[
E[e_n(t)] = \sum_{j=1}^{n} \binom{n}{j} S^j(t) \left[ 1 - S(t) \right]^{n-j} \int_{t}^{\infty} \frac{jS(u)}{S(t)} du
\]

\[
= \sum_{j=1}^{n} \binom{n}{j} S^j(t) \left[ 1 - S(t) \right]^{n-j} e(t)
\]

\[
= \left[ 1 - P(nS_n(t) = 0) \right] e(t)
\]

\[
= \left[ 1 - (1 - S(t))^n \right] e(t).
\]

It follows that the bias of the empirical estimator is \( -e(t)[1 - S(t)]^n \) and hence \( e_n(t) \) is asymptotically unbiased, with bias decaying exponentially to zero as \( n \to 0 \). When \( E[T^2] < \infty \), one can also apply Zahedi’s (1985) procedures to derive the variance of the empirical mrl function which is given by:

\[
Var \left( e_n(t) \right) = [1 - S(t)]^n \left[ 1 - (1 - S(t))^n \right] e_n^2(t) \hspace{1cm} (2.3.4)
\]

\[
+ Var \left[ T - t \mid T \geq t \right] \sum_{j=1}^{n} \frac{P_j}{j}
\]

where \( P_j = \binom{n}{j} S(t)^j (1 - S(t))^{n-j} \) for all \( t < \infty \). Therefore, when \( E[T^2] < \infty \),

\[
Var[e_n(t)] \to 0 \hspace{0.5cm} \text{as} \hspace{0.5cm} n \to \infty.
\]

To prove this, we only need to show that the expression

\[
\sum_{j=1}^{n} \frac{P_j}{j}, \hspace{0.5cm} \text{in the second term of} \hspace{0.5cm} (2.3.4), \hspace{0.5cm} \text{goes to zero as} \hspace{0.5cm} n \to \infty
\]
\[ \sum_{j=1}^{n} \frac{P_j}{j} = \sum_{j=1}^{n} \left\{ \frac{1}{j(j+1)} + \frac{1}{j+1} \right\} P_j \]
\[ \leq \sum_{j=1}^{n} \frac{2}{j+1} P_j \]
\[ = 2 \sum_{j=1}^{n} \frac{(n)}{j+1} S(t)^{j}(1 - S(t))^{n-j}. \]

By letting \( k = j + 1 \), the above expression reduces to:

\[ \sum_{j=1}^{n} \frac{P_j}{j} = \frac{2}{(n+1)S(t)} \sum_{k=1}^{n+1} \binom{n+1}{k} S(t)^{k}(1 - S(t))^{n+1-k} \]
\[ = \frac{2\{1 - [1 - S(t)]^{n+1}\}}{(n+1)(S(t))}. \]

Thus, since \( 0 < S(t) < 1 \), it follows that

\[ \lim_{n \to \infty} \sum_{j=1}^{n} \frac{P_j}{j} \leq \lim_{n \to \infty} \frac{2\{1 - [1 - S(t)]^{n+1}\}}{(n+1)(S(t))} = 0. \]

### 2.3.2 Asymptotic properties of the empirical mrl estimator

Yang (1978) first proposed the empirical mrl estimator \( e_n(t) \) and considered some of its asymptotic properties. For example, it was shown by Yang, that \( e_n(t) \) is asymptotically unbiased, and strongly uniformly consistent estimator of \( e(t) \) on \([0, b]\) for any fixed \( b \). Yang also showed that, when properly centered and normalized, the empirical mrl converges weakly to a certain Gaussian process on \([0, b]\). That is, she proved the following three statements:

- \( E[e_n(t)] \to e(t) \) as \( n \to \infty \).

- \( P \left[ \sup_{0 \leq t \leq b} |e_n(t) - e(t)| \to 0, \text{ as } n \to \infty \right] = 1. \)
\bullet \sqrt{n} [e_n(t) - e(t), 0 \leq t \leq b] \Rightarrow \{U(t), 0 \leq t \leq b\},

where $U(t)$ is a Gaussian process with mean zero and covariance function given by:

\[
\text{Cov}(s, t) = \frac{(1-s)(1-t)\sigma^2(t, 1) - t(1-s)\theta^2(t, 1)}{(1-s)^2(1-t)^2}, 0 \leq s \leq t \leq b,
\]

where $\theta(t, u) = \mathbb{E}[TI_{t<F(T)\leq u}], \sigma(t, u) = \text{Var}[TI_{t<F(T)\leq u}]$ for $0 \leq t < u \leq 1$ and $F(t) = 1 - S(t)$ is the d.f. of $T$. Hall and Wellner (1979) extended her results (strong uniform consistency and weak convergence) to the interval $[0, b_n]$, where $\{b_n\}_{n \geq 1}$ is a sequence of positive numbers with $S(b_n) \to 0$ and $nS(b_n) \to \infty$ as $n \to \infty$. Besides, they obtained nonparametric confidence bands for the mean residual life function. They re-stated Yang’s Theorem of weak convergence as follows:

\[
Z_n(t) = \sqrt{n} [e_n(t) - e(t)] \Rightarrow Z(t) \text{ on } [0, T_F)
\]

where $Z(t) = \frac{\sigma_0}{S(t)}B(G(t)), G(t) = \frac{S(t)\sigma^2(t)}{\sigma_0^2}, \sigma^2(t) = \text{Var}(T - t | T > t)$ is the variance of the residual lifetime distribution, $\sigma_0 = \sqrt{\sigma^2(0)} = \text{Var}(T), T_F = \inf \{t : F(t) = 1\}$ and $B$ represents the standard Brownian motion. Thus, $Z$ is a zero-mean Gaussian process on the support $[0, T_F)$ with covariance function

\[
\text{Cov}(Z(x), Z(y)) = \frac{\sigma^2(y)}{S(x)} \text{ for } 0 \leq x \leq y < T_F.
\]

Thus, replacing $\sigma^2$ by the sample variance $\hat{\sigma}^2$, we get

\[
\sqrt{n} [e_n(t) - e(t)] \sqrt{\frac{S(t)}{\hat{\sigma}^2(t)}} \overset{D}{\to} N(0, 1).
\]

(2.3.5)

Hence, from the above results we can get an asymptotic nonparametric confidence interval for the mrl function $e$ at $t$. Hall and Wellner (1979) constructed nonparametric confidence bands for the mrl function $e$. They showed that if $E[T^r] < \infty$ for
some \( r > 2 \) and \( a \) is a number such that \( P[Y \leq a] = P[\|B\|_0 \leq a] = 1 - \alpha \), where
\[
P[Y \leq a] = \sum_{k=-\infty}^{\infty} (-1)^k \{ \Phi((2k+1)a) - \Phi((2k-1)a) \} \approx 1 - 4(1 - \Phi(a)),
\]

then
\[
\lim_{n \to \infty} P \left\{ e_n(t) - \frac{a\hat{\sigma}}{\sqrt{n}S_n(t)} \leq e(t) \leq e_n(t) + \frac{a\hat{\sigma}}{\sqrt{n}S_n(t)} \text{ for all } t \leq t < \infty \right\} \geq 1 - \alpha.
\]

### 2.4 The Proportional MRL Model

#### 2.4.1 Specification and existence criterion of the model

Motivated by the proportional hazards model of Cox (1972), Oakes and Dasu (1990) as well as Arnold and Zahedi (1991) proposed a new semiparametric proportional mrl model. Consider two survival functions \( S_0 \) and \( S_1 \). Then \( S_0 \) and \( S_1 \) are said to have proportional mrl functions if

\[
ed_1(t) = \theta e_0(t), \quad \forall t \tag{2.4.1}
\]

\[
\iff \quad \frac{\int_t^\infty S_1(u)du}{S_1(t)} = \theta \frac{\int_t^\infty S_0(u)du}{S_0(t)}
\]

for some \( \theta > 0 \). Using the inversion formula, we get

\[
S_1(x, \theta) = \frac{e_1(0)}{e_1(x)} \exp \left( - \int_0^x \frac{du}{e_1(u)} \right)
\]

\[
= \frac{e_0(0)}{e_0(x)} \exp \left( - \int_0^x \frac{du}{\theta e_0(u)} \right)
\]

\[
= \frac{e_0(0)}{e_0(x)} \left( \frac{S_0(x)e_0(x)}{e_0(0)} \right)^{\frac{1}{\theta}}
\]

\[
= S_0(x) \left( \frac{\int_x^\infty S_0(u)du}{e_0(0)} \right)^{\frac{1}{\theta} - 1}.
\]
Families of distributions need to satisfy certain conditions before they can be considered suitable for the PMRL model. The sufficient conditions can be determined by checking that \( S_1(x, \theta) \) is a valid survivor function. If \( S_0(x) \) is a proper survivor function and \( \theta < 1 \), then \( S_1(x, \theta) \) is always a valid survivor function. But when \( \theta > 1 \), \( \left( \frac{\int_x^\infty S_0(u)du}{\mu_0} \right)^{\frac{1}{\theta} - 1} \) is not a decreasing function which implies that \( S_1(x, \theta) \) is not a valid survival function. Hence we have to impose conditions on \( c_0(x) \) to make the model meaningful for all positive values of \( \theta \). Note that \( \frac{\partial}{\partial x} S_1(x, \theta) = \frac{c_0(0)}{c_0(x)} \exp \left( - \int_x^\infty \frac{du}{\theta c_0(u)} \right) \left[ \frac{\partial}{\partial x} c_0(x) + \frac{1}{\theta} \right] (-1) \). Therefore, \( S_1(x, \theta) \) is a decreasing function if \( \frac{\partial}{\partial x} c_0(x) + \frac{1}{\theta} \) is non-negative, i.e. \( \frac{\partial}{\partial x} c_0(x) \geq -\frac{1}{\theta} \). In particular, this inequality has to hold in the limit as \( \theta \to \infty \), which implies that \( \frac{\partial}{\partial x} c_0(x) \geq 0, \forall x \). Hence, \( S_1(x, \theta) \) represents a valid survival function for all \( \theta > 0 \) if and only if \( c_0(x) \) in nondecreasing in \( x \).

2.4.2 Relation between PH and PMRL models

The hazard rate \( h(t) \) and the mrl function \( e(t) \) are mathematically equivalent in the sense that knowing one of them, the other can be determined. However, for modeling purposes, the hazard rate and the mrl function play somewhat different roles. The hazard rate \( h(t) \) attempts to capture the risk of instantaneous death at age \( t \) whereas the mrl function \( e(t) \) summarizes the entire distribution of the remaining lifetime of an individual of age \( t \). If the main interest is the risk of immediate death (failure), an extremely useful model for the analysis of survival data is the celebrated proportional
hazards (PH) model of Cox (1972). The PH model has proved to be of tremendous value in survival analysis. However, it may not be the most appropriate model when one is concerned, as it is the case when dealing with replacement and repair models, with the remaining lifetime of an individual (equipment) of age $t$. For example, suppose we want to estimate the life expectancy of a particular kind of batteries and the manufacturer has data available only on batteries which are run under laboratory conditions. If it can be assumed that the life expectancy under actual operating conditions is proportional to the life expectancy under the laboratory conditions with a known proportionality constant, one can easily estimate the life expectancy under the actual operating conditions even though the data on such batteries are not available. Hence, it appears natural to see how far a mean residual life analogue of the PH model can be useful.

The PH, Cox (1972), model states that

$$h_1(t) = \beta h_0(t) \iff S_1(t) = \{S_0(t)\}^\beta$$

(2.4.2)

where $\beta > 0$ is independent of $t$. The PMRL model, which is defined earlier, states that

$$\epsilon_1(t) = \theta \epsilon_0(t) \iff S_1(t) = S_0(t) \left\{ \int_t^\infty \frac{S_0(u)\,du}{\epsilon_0(0)} \right\}^{\frac{1}{\beta}-1}$$

(2.4.3)

where $\theta > 0$ is independent of $t$.

An interesting connection between the mrls in the PMRL occurs in the case of renewal processes. Consider two stationary renewal processes with underlying survival
functions $S_0$ and $S_1$. Let $R_0$ and $R_1$ be the corresponding residual lifetimes at a fixed time. Then the PMRL model requires

$$e_1(t) = \theta e_0(t)$$

$$\iff$$

$$\frac{1}{h_{R_1}(t)} = \theta \frac{1}{h_{R_0}(t)}$$

$$\iff$$

$$h_{R_1}(t) = \frac{1}{\theta} h_{R_0}(t)$$

which shows that $R_0$ and $R_1$ follow a proportional hazards model with a proportionality parameter $\theta^{-1}$. Therefore, if the inter-event distributions of two renewal processes are PMRL, then the residual lifetimes of the corresponding stationary (equilibrium) renewal processes are PH (note that this is true only in the case of renewal processes). Maguluri and Zhang (1994) took advantage of this relationship to employ techniques for the proportional hazards model to analyze data from the PMRL model. Moreover, they extended the above simple PMRL model to a more general model with explanatory variables $Z$:

$$e(t | z) = \exp (\theta^t z) e_0(t)$$

(2.4.4)

where $e(t | z)$ is the conditional mrl function given $Z = z$. Goodness-of-fit-tests for the proportional mean residual life regression model, (2.4.4), were derived by K. Yuen et al (2001).
2.5 Estimation of the MRL Function under Censoring

Let \( X_1, X_2, \ldots, X_n \) denote a random sample of failure times from a random variable \( X \) with distribution \( F \). Suppose that the failure times \( X_1, X_2, \ldots, X_n \) are subject to right censoring. Let \( C_1, C_2, \ldots, C_n \) denote a random sample of censoring times from a random variable \( C \) with distribution function \( G \). For each \( i \), only one of \( X_i \) and \( C_i \) can be observed; when \( X_i \leq C_i \), we observe \( X_i \) (an exact observation), otherwise we observe \( C_i \) (a censored observation). Let \( \delta_i = 1 \) if \( X_i \leq C_i \) and 0 otherwise. Define \( T_i = \min(X_i, C_i) \), \( i = 1, 2, \ldots, n \). Further, let \( T_{(1)} \leq T_{(2)} \leq \cdots \leq T_{(n)} \) be the order statistics corresponding to \( T_1, T_2, \ldots, T_n \) and let \( \delta_i \) be the value of \( \delta \) associated with \( T_i \), i.e., \( \delta_i = \delta_j \) if \( T_i = T_j \). Define \( \tau_T = \max\{T_i : \delta_i = 1, 1 \leq i \leq n\} \). Then the most common estimator of \( S = 1 - F \) is the K-M estimator, \( S_{km} \), which was discussed in Chapter 1. Thus, a reasonable estimator of \( e(t) \) under censoring is given by

\[
\hat{e}(t) = \frac{\int_t^\infty S_{KM}(x)dx}{S_{KM}(t)}
\]

(2.5.1)

for \( 0 \leq t < \tau_T \). The following results can be found in Li (1997).

Assume that \( \tau_F = \sup\{x|F(x) < 1\} \leq \sup\{x|G(x) < 1\} \). Let \( W(t) \), for \( 0 \leq t < \tau_F \), be Brownian motion defined in (A.0.5). Since the Brownian motion process has independent increments, i.e., \( W(t) \) and \( |W(t + \epsilon) - W(t)| \), for all \( \epsilon > 0 \) are independent, \( W(t) \) and \( \int_t^\infty f(x)dW(x) \) are also independent, where \( f \) is a Riemann
integrable function. Thus, as \( n \to \infty \),

\[
\sqrt{n}[\hat{e}(t) - e(t)] = \frac{\sqrt{n}S(t)}{S_{KM}(t)S(t)} \int_t^\infty [S_{KM}(u) - S(u)]du - \frac{\sqrt{n}[S_{KM}(u) - S(u)]}{S_{KM}(t)S(t)} \int_t^\infty S(u)du \\
\overset{D}{\rightarrow} \frac{1}{S(t)}Z(t) - \frac{W(t)}{S^2(t)} \int_t^\infty S(u)du \\
= Y(t),
\]

(2.5.2)

where \( W(t) \) and \( Z(t) \) are defined as in (A.0.5) and (A.0.7). By the independence of \( Z(t) \) and \( W(t) \), \( Y(t) \) is a normal variate with mean zero and variance given by

\[
\sigma_Y^2 = \frac{Var[Z(t)]}{S(t)^2} + \frac{Var[W(t)][\int_t^\infty S(u)du]^2}{S(t)^4}
\]

(2.5.3)

\[
= S(t)^2 \int_t^\infty \frac{[\int_x^\infty S(v)dv]^2dF(x)}{S(x)H(x)} + \frac{U(t)[\int_t^\infty S(u)du]^2}{S(t)^4}
\]

where \( U(t) \), defined in (A.0.6), is the covariance function of \( W(t) \). \( U \) can be estimated by the Greenwood’s formula which is defined in (A.0.3). Replacing \( S \) and \( U \) in (2.5.4) by \( S_{KM} \) and the variance of \( S_{KM} \) (i.e., Greenwood’s formula) respectively, we obtain a consistent estimator \( \hat{\sigma}_Y \) of \( \sigma_Y \). Thus, a \((1 - \alpha)100\%\) confidence interval for \( e(t) \) is obtained as follows:

\[
\hat{e}(t) - z_{\alpha/2} \frac{\hat{\sigma}_Y}{\sqrt{n}} \leq e(t) \leq \hat{e}(t) + z_{\alpha/2} \frac{\hat{\sigma}_Y}{\sqrt{n}}.
\]

(2.5.4)

For further discussion on the estimation of the mrl function under censoring, refer to Li (1997).
2.6 The MRL Function of Discrete Random Variables

The mrl function was introduced in the previous sections as a way of representing the probability law of a continuous positive random variable. However, the definition of the mrl function given in Section 2.1 is also valid in the case of discrete or mixed positive random variables. Although it is unlikely that discrete models will be used in practice to describe phenomena such as time to failure, or repair time, this case is included for the sake of completeness.

Let \( T \) be a positive discrete random variable with probability mass function \( p(t) \), and with nonzero probability mass at the points \( t_i, \ i = 1, 2, \ldots \). Then from the definition of the mrl function it follows that

\[
e(t) = E[T - t | T > t] = \frac{\sum_{t_j \geq t} t_j p(t_j)}{\sum_{t_j > t} p(t_j)} - t.
\]

(2.6.1)

It is seen from equation (2.6.1) that \( e(t) \) makes upward jumps when \( t = t_i \), and decreases with a slope of 45° elsewhere. The sequence values \( e(t_i), i = 1, 2, \ldots \), determines uniquely the probability distribution of \( T \). From equation (2.6.1) we obtain the following recursive formula for the probability mass function

\[
p(t_{i+1}) = \frac{S(t_i)}{e(t_{i+1})} [t_{i+1} - t_i + e(t_{i+1}) - e(t_i)]
\]

(2.6.2)

where \( S(t_0) = 1, \ p(t_0) = 0 \), so that \( p(t_1) = \frac{1}{e(t_1)} [t_1 - t_0 + c(t_1) - \mu] \), and \( S(t_{i+1}) = S(t_i) - p(t_{i+1}) \).
Chapter 3

Bivariate Mean Residual Function

In recent years considerable attention has been focused on the study of the univariate mrl function. In contrast, relatively little attention has been devoted to the analysis of bivariate mrl functions. In this chapter attention is focused on the problem of nonparametric estimation of the bivariate mean residual function. Kulkarni and Rattihalli (2002) have generalized Yang (1978) results to the bivariate case. We will discuss their asymptotic results. We will start with an overview of the bivariate mean residual life function. This is done in Section 3.1. Section 3.2 deals with the large sample properties of the bivariate empirical mean residual life function. Section 3.3 presents a brief description of the tests and confidence sets which we have developed for the bivariate mrl function. Section 3.4 considers the problem of estimating the bivariate mrl function under censoring. Finally, in Section 3.5 some simulation results are carried out to study the bias and mean squared error of the empirical mrl
function. Although we have closed form expressions for the mean and variance of the empirical estimator, it is rather complicated to calculate the variance. Hence, we used simulation results to examine the behavior of the empirical mrl function.

### 3.1 Definition, Properties and Characterization

Let $Z = (X, Y)$ be a random vector representing the lifetimes of two individuals. That is, $Z = (X, Y)$ is a random vector in the first quadrant $Q = \{(x_1, x_2) : x_i \geq 0, i = 1, 2\}$ of $\mathbb{R}^2$, and let $z = (x, y)$ be a vector of non-negative real numbers. A bivariate distribution function is a function $F(z) = F(x, y)$ on $\mathbb{R}^2$ with the following properties:

1. $F$ is everywhere right continuous (i.e., continuous from above);

2. $0 \leq F(x, y) \leq 1$ for all $(x, y)$, $F$ is nondecreasing in each variable;

3. $F(x, y) \rightarrow 0$ as $x \rightarrow 0$ or $y \rightarrow 0$, and $F(x, y) \rightarrow 1$ as $x \rightarrow \infty$ and $y \rightarrow \infty$.

For $F(z) = P[X \leq x, Y \leq y]$, the joint distribution function of $(X, Y)$, $f(z) = f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$ is called the joint density function of $X$ and $Y$. The bivariate survival function $S(z)$ is a function on $\mathbb{R}_+^2$ and it is given by

$$S(z) = S(x, y) = P[X > x, Y > y]. \quad (3.1.1)$$

The cumulative bivariate hazard function at time $(x, y)$ is given by

$$H(x, y) = -\log S(x, y), \quad (3.1.2)$$
and, assuming $H(x, y)$ is absolutely continuous with partial derivatives that exist almost everywhere, the bivariate hazard rate or hazard gradient of $H(x, y)$ is defined as

$$h(x, y) = \nabla H(x, y)$$

$$= \left( \frac{\partial}{\partial x} H(x, y), \frac{\partial}{\partial y} H(x, y) \right)$$

$$= \left( \frac{\partial}{\partial x} - \log S(x, y), \frac{\partial}{\partial y} - \log S(x, y) \right)$$

$$= (h_X(x, y), h_Y(x, y))$$

where $h_X(x, y)$ and $h_Y(x, y)$ are the components of the bivariate hazard rate. Then $H(x, y)$ can be represented as the path-independent integral of $h(x, y)$ from $(0, 0)$ to $(x, y)$. In particular, for the linear path $(0, 0)$ to $(x, 0)$ to $(x, y)$,

$$H(x, y) = \int_0^x h_X(u, 0) \, du + \int_0^y h_Y(x, v) \, dv.$$

If both components of $h(x, y)$ are increasing (decreasing) functions of the corresponding variable, i.e., $h_X(x, y)$ is an increasing (decreasing) function of $x$ and $h_Y(x, y)$ is an increasing (decreasing) function of $y$, then the distribution is called bivariate IHR (DHR). One can easily show that if $X$ and $Y$ are independent then $h_X(x, y) = h_X(x)$ and $h_Y(x, y) = h_Y(y)$ where $h_X(x)$ and $h_Y(y)$ are the univariate hazard rates that correspond to $X$ and $Y$ respectively. So, if $X$ and $Y$ are independent, their joint distribution is IHR (DHR) if and only if the distribution of both $X$ and $Y$ are IHR (DHR).
3.1.1 Definition of the bivariate mrl function

The random vector \( R(z) = (Z - z | Z > z) \) is the bivariate residual life function at age \( z = (x, y) \) and its average value \( e(z) = E(R(z)) \) is called the bivariate mean residual life function and is given by

\[
e(z) = E[Z - z | Z > z] = (e_X(z), e_Y(z)) = (e_X(x, y), e_Y(x, y))
\]  

(3.1.5)

where

\[
e_X(x, y) = E[X - x | Z > z],
\]  

(3.1.6)

and

\[
e_Y(x, y) = E[Y - y | Z > z].
\]  

(3.1.7)

In reliability analysis \( e_X(x, y) \) measures the expected remaining life of the first component conditioned on the fact that the first component has survived \( x \) units of time and the second unit has survived \( y \) units of time. The component \( e_Y(x, y) \) is interpreted similarly. In actuarial statistics, \( e_X(x, y) \) can be interpreted as follows: Let \( X \) and \( Y \) be the ages at death of a woman and her husband who bought life insurance at possibly different ages. Then \( e_X(65, 75) \) is the remaining lifetime of the woman given that her current age is 65 and her husband age 75.
Equations (3.1.6) and (3.1.7) can be further simplified and expressed in terms of the bivariate survival function. Consider, for example, equation (3.1.6).

\[
e_X(x, y) = E[X - x | Z > z] = \frac{\int_x^\infty \int_y^\infty (u - x) f(u, v) dvdu}{P[Z > z]} = \left\{ \frac{1}{S(x, y)} \right\} \left\{ \int_x^\infty \int_y^\infty (u - x) \frac{\partial^2}{\partial x \partial y} S(u, v) dvdu \right\} = \left\{ \frac{1}{S(x, y)} \right\} \left\{ \int_x^\infty (u - x) \frac{\partial}{\partial x} \left[ \int_y^\infty \frac{\partial}{\partial y} S(u, v) dv \right] du \right\} = \left\{ \frac{1}{S(x, y)} \right\} \left\{ \int_x^\infty (u - x) \left( -\frac{\partial}{\partial x} S(u, y) \right) du \right\} = \left\{ \frac{1}{S(x, y)} \right\} \left[ -(u - x) S(u, y) \right]_{x}^{\infty} + \int_x^\infty S(u, y) du = \frac{1}{S(x, y)} \int_x^\infty S(u, y) du.
\]

In a similar fashion, equation (3.1.7) can be simplified into the following expression

\[
e_Y(z) = e_Y(x, y) = E[Y - y | Z > z] = \frac{1}{S(x, y)} \int_y^\infty S(x, v) dv.
\]

The univariate mrl functions can be obtained from the components of the bivariate mrl function. That is, \( e_X(x) = e_X(x, 0) \) and \( e_Y(y) = e_Y(0, y) \). Moreover, if \( X \) and \( Y \) are independent then \( e_X(x, y) = e_X(x) \) and \( e_Y(x, y) = e_Y(y) \) where \( e_X(x) \) and \( e_Y(y) \) are the univariate mrl functions that correspond to the random variables \( X \) and \( Y \) respectively. These results can be easily proved using equations (3.1.6) and (3.1.7).
Recently attention has been directed toward extending the concepts of univariate NBUE (NWUE) to the bivariate case. Such extensions are useful in modeling and analyzing bivariate failure data, when there is lack of independence among the components’ lifetimes. Bivariate NBUE and its dual bivariate NWUE have been defined and their properties have been developed by several authors. For a bibliography of available results see Ebrahimi and Zahedi (1989) and references therein. Some bivariate notions of IFR and DMRL are discussed in the paper of Bassan et al (2002).

Let $X$ and $Y$ denote the survival times of two individuals having joint cumulative d.f. $F(x, y)$ and joint survival function $S(x, y)$. For simplicity we assume that $F$ is absolutely continuous with $F(x, y) < 1$ for all $(x, y) \in \mathbb{R}_+^2$. We also assume that $S(0, 0) = 1$, and that $E(X)$ and $E(Y)$ are finite. The following definition of Bivariate New Better than Used in Expectation (BNBUE) and Bivariate New Worse than Used in Expectation (BNWUE) appeared in Ebrahimi and Zahedi (1989):

**Definition 3.1.1.** A bivariate survival function $S$ is said to be BNBUE (BNWUE) if $e_X(x, y) \leq (\geq) e_X(0, y)$ and $e_Y(x, y) \leq (\geq) e_Y(x, 0)$ for all $x, y \geq 0$, where $e_X$ ($e_Y$) denotes the first (second) component of the bivariate mrl function corresponding to the random vector $(X, Y)$.

If $X$ and $Y$ are independent, then the joint distribution is BNBUE (BNWUE) if the distributions of both $X$ and $Y$ are BNBUE (BNWUE). Ebrahimi and Zahedi (1989) showed that a bivariate survival function is both NBUE and NWUE if and only if it is a bivariate Gumbel distribution.
3.1.2 Relationship between the bivariate mrl and survival functions

As pointed out in Chapter 2 there exits a one-to-one relationship between the mrl and survival function as well as between the mrl and hazard function of a univariate random variable. The extensions of these one-to-one relationships to the bivariate case is discussed by Nair and Nair (1989).

**Theorem 3.1.2.** *(Nair and Nair (1989)).* The components of the hazard gradient and the bivariate mean residual function are connected by the relation:

\[
    h_X(x, y) = \frac{\frac{\partial}{\partial x} e_X(x, y) + 1}{e_X(x, y)} \tag{3.1.8}
\]

and

\[
    h_Y(x, y) = \frac{\frac{\partial}{\partial y} e_Y(x, y) + 1}{e_Y(x, y)}.
\]

**Proof.** For brevity, we only show the proof for (3.1.8). The proof of (3.1.9) follows from the arguments used to prove (3.1.8). We begin with \( e_X(x, y) = \frac{\int_{x}^{\infty} S(u,y) du}{S(x,y)} \); then, differentiating both sides of the expression, we find that

\[
    \frac{\partial}{\partial x} e_X(x, y) = \frac{-\frac{\partial}{\partial x} S(x, y)}{S^2(x, y)} \int_{x}^{\infty} S(u, y) du + \frac{1}{S(x, y)} [S(\infty, y) - S(x, y)]
\]

\[
    = \left\{ -\frac{\frac{\partial}{\partial x} S(x, y)}{S(x, y)} \right\} \left\{ \int_{x}^{\infty} S(u, y) du \right\} \frac{1}{S(x, y)} - 1
\]

\[
    = h_X(x, y) e_X(x, y) - 1.
\]

Simplifying the above expression for \( h_X(x, y) \) gives \( h_X(x, y) = \frac{\frac{\partial}{\partial x} e_X(x, y) + 1}{e_X(x, y)} \). \( \Box \)
Roy (1989) proved that if there exists some constant $c$ such that,

$$e_x(x, y)h_x(x, y) = e_y(x, y)h_y(x, y) = c$$

then

1. $c = 1$ if and only if $Z = (X, Y)$ follows the bivariate Gumbel exponential distribution with

$$S(x, y) = \exp\{-(\theta_1 x + \theta_2 y + \theta_{12} xy)\}$$

where $\theta_1$, $\theta_2$ and $\theta_{12}$ are positive constants.

2. $c > 1$ if and only if $Z = (X, Y)$ follows the bivariate Lomax distribution, denoted by $Z \sim ML_2(a, \theta_1, \theta_2)$, with

$$S(x, y) = (1 + \theta_1 x + \theta_2 y)^{-a}$$

where $a$, $\theta_1$ and $\theta_2$ are positive constants.

3. $0 < c < 1$ if and only if $Z$ follows the bivariate rescaled Dirichlet distribution, denoted by $Z \sim MRD_a(a, \theta_1, \theta_2)$, with

$$S(x, y) = \begin{cases} 
(1 - [\theta_1 x + \theta_2 y])^a, & 0 \leq \theta_1 x + \theta_2 y \leq 1 \\
0, & \text{otherwise}
\end{cases}$$

where $a$, $\theta_1$ and $\theta_2$ are positive constants.

The following theorem indicates that the bivariate survival function may be expressed in terms of the bivariate mean residual lifetime function:
Theorem 3.1.3. (Nair and Nair (1989)): The bivariate mrl function uniquely determines the joint survival function. That is,

\[
S(x, y) = \frac{e_y(x, 0) S(x, 0)}{e_y(x, y)} \exp \left( - \int_0^y \frac{du}{e_y(x, u)} \right)
= \frac{e_x(0, y) S(0, y)}{e_x(x, y)} \exp \left( - \int_0^x \frac{du}{e_x(u, y)} \right).
\] (3.1.10)

Proof. From equation (3.1.3) we know that \( h_X(x, y) = \frac{\partial}{\partial x} H(x, y) \). By integrating each side of this expression between 0 and \( x \) we obtain

\[
\int_0^x h_X(u, y) du = \int_0^x \frac{\partial}{\partial u} H(u, y) du.
\]

This implies that

\[
H(x, y) = \int_0^x h_X(u, y) du + H(0, y).
\] (3.1.11)

Once again, from equation (3.1.2) we know that \( S(x, y) = \exp(-H(x, y)) \) which is equivalent to \( S(x, y) = \exp \left( - \int_0^x h_X(u, y) du - H(0, y) \right) \). The last equality follows from equation (3.1.11). Therefore,

\[
S(x, y) = \exp(-H(x, y))
= \exp \left( - \int_0^x h_X(u, y) du - H(0, y) \right)
= \exp \left( - \int_0^x \left[ \frac{\partial}{\partial u} e_X(u, y) + \frac{1}{e_X(u, y)} \right] du \right) \exp(-H(0, y))
= \exp \left( - \int_0^x \left[ \frac{\partial}{\partial u} \log(e_X(u, y)) \right] du \right) \exp \left( - \int_0^x \left[ \frac{1}{e_X(u, y)} \right] du \right) S(0, y)
= \exp \left( \log \left[ \frac{e_X(0, y)}{e_X(x, y)} \right] \right) \exp \left( - \int_0^x \frac{du}{e_X(u, y)} \right) S(0, y)
= \exp \left( \log \left[ \frac{e_X(0, y)}{e_X(x, y)} \right] \right) \exp \left( - \int_0^x \frac{du}{e_X(u, y)} \right) S(0, y)
\]
\[
\frac{e_X(0,y)S(0,y)}{e_X(x,y)} \exp \left( - \int_0^x \frac{du}{e_X(u,y)} \right) \\
= \frac{e_X(0,y)S_Y(y)}{e_X(x,y)} \exp \left( - \int_0^x \frac{du}{e_X(u,y)} \right).
\]

Starting with \( h_X(x,y) = \frac{\partial}{\partial x} H(x,y) \) and using the same argument, one can also show that
\[
S(x,y) = \frac{e_Y(x,0)S(x,0)}{e_Y(x,y)} \exp \left( - \int_0^y \frac{du}{e_Y(x,u)} \right) \\
= \frac{e_Y(x,0)S_X(x)}{e_Y(x,y)} \exp \left( - \int_0^y \frac{du}{e_Y(x,u)} \right).
\]

The variance of the residual life function has generated considerable interest in the past few years. The following results can be found in Gupta and Kirmani (2000).

Noting that \( E[(X - x)^2|X > x, Y > y] = 2 \int_x^\infty e_X(u,y)S(u,y)du S(x,y) \) and \( E[(Y - y)^2|X > x, Y > y] = 2 \int_y^\infty e_Y(x,v)S(x,v)dv S(x,y) \), the components of the variance of the residual life function are \( V(x,y) = [V_X(x,y), V_Y(x,y)] \), where
\[
V_X(x,y) = E \left[ (X - x)^2 | (X,Y) \geq (x,y) \right] - \left\{ E \left[ (X - x) | (X,Y) \geq (x,y) \right] \right\}^2 \\
= \frac{2 \int_x^\infty e_X(u,y)S(u,y)du}{S(x,y)} \left( \frac{e_X^2(x,y)}{S(x,y)} - e_X^2(x,y) \right),
\]
\[
V_Y(x,y) = E \left[ (Y - y)^2 | (X,Y) \geq (x,y) \right] - \left\{ E \left[ (Y - y) | (X,Y) \geq (x,y) \right] \right\}^2 \\
= \frac{2 \int_y^\infty e_Y(x,v)S(x,v)dv}{S(x,y)} \left( \frac{e_Y^2(x,y)}{S(x,y)} - e_Y^2(x,y) \right).
\]

Let \( \gamma(x,y) = [\gamma_X(x,y), \gamma_Y(x,y)] \) be the square of the residual coefficient of variation, which is defined as the ratio of the components of the variance to the correspond-
ing mrl function. Then, the components of the vector of the residual coefficient of variation are given by:

\[ \gamma_X(x, y) = \frac{\sqrt{V_X(x, y)}}{e_X(x, y)}, \]

\[ \gamma_Y(x, y) = \frac{\sqrt{V_Y(x, y)}}{e_X(x, y)}. \]

Hence knowing \( \gamma_X(x, y), \gamma_Y(x, y) \) and the variance of the residual life, one can obtain \( e_X(x, y) \) and \( e_Y(x, y) \) and hence the bivariate distribution. This shows that \( \gamma(x, y) \) characterizes the bivariate distribution when the variance of the residual life function is known.

### 3.1.3 Necessary and sufficient conditions for a bivariate function to be a bivariate mrl function

In the univariate case Swartz (1973) has obtained the necessary and sufficient conditions for a non-negative function \( e \) to be mrl of some life-time distribution. This subsection characterizes the bivariate mrl of an absolutely continuous random vector. Kulkarni and Rattihalli (1996) have obtained the necessary and sufficient conditions for a non-negative function \( e : [0, \infty)^2 \rightarrow [0, \infty)^2 \) to be a valid bivariate mrl function. Their conditions are given below:

(i) \( \frac{\partial}{\partial y} h_X(x, y) = \frac{\partial}{\partial x} h_Y(x, y) \), for all \((x, y) \geq 0\). It follows that a solution to

\[ \nabla \{- \log(S(x, y))\} = \{h_X(x, y), h_Y(x, y)\} \]

exists if this condition holds.
(ii) Both \( \lim_{x \to -\infty} \left\{ \int_{0}^{x} \frac{du}{e^{x} (u, y)} \right\} \) and \( \lim_{y \to -\infty} \left\{ \int_{0}^{y} \frac{du}{e^{y} (x, u)} \right\} \) are infinite. Moreover, both
\[
\lim_{x \to -\infty} \left\{ \frac{1}{e^{x} (x, y)} \exp \left[ - \int_{0}^{x} \frac{du}{e^{u} (u, y)} \right] \right\} \quad \text{and} \quad \lim_{y \to -\infty} \left\{ \frac{1}{e^{y} (x, y)} \exp \left[ - \int_{0}^{y} \frac{du}{e^{u} (x, u)} \right] \right\}
\]
exist. These conditions are required so that the survival function \( S(x, y) \) is nondecreasing.

(iii) \( h_{X}(x, y) \cdot h_{Y}(x, y) \geq \frac{\partial h_{X}(x, y)}{\partial y} \) for all \((x, y) \geq (0, 0)\). This condition ensures that the pdf corresponding to \( S(x, y) \) is nonnegative (i.e. \( f(x, y) = \frac{\partial S(x, y)}{\partial x \partial y} \geq 0 \)).

### 3.2 The Empirical Bivariate MRL Function

#### 3.2.1 Definition

Let \((X_{1i}, X_{2i})\), \(i = 1, 2, \ldots, n\) be independent and identically distributed random vectors representing the lifetimes of \(n\) pairs of individuals and let \(S_{n}(x_{1}, x_{2})\) denote the empirical bivariate survival function. The empirical bivariate mrl function at a point \((x_{1}, x_{2})\) is given by

\[
\hat{e}(x_{1}, x_{2}) = (\hat{e}_{X_{1}}(x_{1}, x_{2}), \hat{e}_{X_{2}}(x_{1}, x_{2}))
\]

where

\[
\hat{e}_{X_{1}}(x_{1}, x_{2}) = \frac{\int_{-\infty}^{x_{1}} S_{n}(u, x_{2}) du}{S_{n}(x_{1}, x_{2})} = \frac{\sum_{j=1}^{n} (X_{1j} - x_{1}) I_{\{X_{1j} > x_{1}, X_{2j} > x_{2}\}}}{\sum_{j=1}^{n} I_{\{X_{1j} > x_{1}, X_{2j} > x_{2}\}}} \quad \text{(3.2.1)}
\]

\[
\hat{e}_{X_{2}}(x_{1}, x_{2}) = \frac{\int_{-\infty}^{x_{2}} S_{n}(x_{1}, u) du}{S_{n}(x_{1}, x_{2})} = \frac{\sum_{j=1}^{n} (X_{2j} - x_{2}) I_{\{X_{1j} > x_{1}, X_{2j} > x_{2}\}}}{\sum_{j=1}^{n} I_{\{X_{1j} > x_{1}, X_{2j} > x_{2}\}}} \quad \text{(3.2.2)}
\]
Thus, \( \hat{e}_{X_i}(x_1, x_2) \) for \( i = 1, 2 \) is the \( ith \) component of the average of the residual lifetimes from observations surviving beyond \( (x_1, x_2) \). Here we assume that the distribution of \( (X_1, X_2) \) has finite first order moments. Note that (3.2.1) and (3.2.2) are defined only when \( S_n(x_1, x_2) > 0 \) which reflects the fact that if we have no observations beyond \( (x_1, x_2) \), then the data provides no information regarding the bivariate mrl function at \( (x_1, x_2) \). When \( S_n(x_1, x_2) = 0 \), \( \hat{e}_{X_i}(x_1, x_2) \) can be defined arbitrarily. For the sake of convenience in calculation, we define \( \hat{e}_{X_i}(x_1, x_2) = 0 \) for \( i = 1, 2 \) whenever \( S_n(x_1, x_2) = 0 \). Thus, (3.2.1) and (3.2.2) can be written as \( \hat{e}_{X_1}(x_1, x_2) = \frac{\int_{\frac{1}{2}}^{\infty} S_n(u, x_2) du}{S_n(x_1, x_2)} \cdot I(S_n(x_1, x_2) > 0) \) and \( \hat{e}_{X_2}(x_1, x_2) = \frac{\int_{\frac{1}{2}}^{\infty} S_n(x_1, u) du}{S_n(x_1, x_2)} \cdot I(S_n(x_1, x_2) > 0) \) where \( I(S_n(x_1, x_2)) = 1 \) if \( S_n(x_1, x_2) > 0 \) and zero otherwise.

In what follows, the expected value and variance of the bivariate empirical mrl estimator will be given. These results were derived by Zahedi (1985).

\[
E[\hat{e}_{X_1}(x_1, x_2)] = \sum_{j=1}^{n} E[\hat{e}_{X_1}(x_1, x_2) | nS_n(x_1, x_2) = j] P(nS_n(x_1, x_2) = j) \quad (3.2.3)
\]

\[
= \sum_{j=1}^{n} E \left[ \frac{1}{j} \int_{\frac{1}{2}}^{\infty} nS_n(u, x_2) du | nS_n(x_1, x_2) = j \right] P(nS_n(x_1, x_2) = j)
\]

\[
= \sum_{j=1}^{n} P(nS_n(x_1, x_2) = j) \int_{\frac{1}{2}}^{\infty} E \left[ S_n(u, x_2) | nS_n(x_1, x_2) = j \right] du.
\]

Note that the last expression of the equality is obtained by interchanging the order of expectation and integration. From the definition of \( S_n(x_1, x_2) \) it follows that \( nS_n(x_1, x_2) \sim Bin(n, S(x_1, x_2)) \); that is \( nS_n(x_1, x_2) \) is a binomial random variable with parameter \( n \) and \( S(x_1, x_2) \). Therefore, we have \( nS_n(u, x_2) \sim Bin(n, S(u, x_1)) \).
and hence \( \{ nS_n(u, x_2) | nS_n(x_1, x_2) = j \} \sim Bin \left( j, \frac{S(u, x_2)}{S(x_1, x_2)} \right) \). This implies that

\[
P(nS_n(x_1, x_2) = j) = \binom{n}{j} S^j(x_1, x_2)[1 - S(x_1, x_2)]^{n-j}
\]  

(3.2.4)

and

\[
E[nS_n(u, x_2) | nS_n(x_1, x_2) = j] = \frac{jS(u, x_2)}{S(x_1, x_2)}.
\]  

(3.2.5)

Substituting (3.2.4) and (3.2.5) in (3.2.3) we get

\[
E[\hat{\epsilon}_{X_i}(x_1, x_2)] = \sum_{j=1}^{n} \binom{n}{j} S^j(x_1, x_2)[1 - S(x_1, x_2)]^{n-j} \int_{x_1}^{\infty} \frac{S(u, x_2)du}{S(x_1, x_2)}
\]

\[
= e_{X_i}(x_1, x_2)[1 - P(nS_n(x_1, x_2) = 0)]
\]

\[
= e_{X_i}(x_1, x_2)[1 - (1 - S(x_1, x_2))^n].
\]

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

\[
E[\hat{\epsilon}_{X_i}(x_1, x_2)] = e_{X_i}(x_1, x_2)[1 - (1 - S(x_1, x_2))^n].
\]  

(3.2.6)

When \( E[X_i^2] < \infty \), the variance of \( \hat{\epsilon}_{X_i}(x_1, x_2) \) is given by

\[
Var[\hat{\epsilon}_{X_i}(x_1, x_2)] = \{1 - S(x_1, x_2)\}^n \{1 - [1 - S(x_1, x_2)]^n\} e_{X_i}(x_1, x_2)
\]

(3.2.7)

\[
+ Var[X_i - x_i | (X_1, X_2) \geq (x_1, x_2)] \sum_{j=1}^{n} \frac{1}{j} B(n, j, S(x_1, x_2))
\]

where \( B(n, j, S(x_1, x_2)) = \binom{n}{j} [S(x_1, x_2)]^j [1 - S(x_1, x_2)]^{n-j} \).

### 3.2.2 Asymptotic properties

We next concentrate on the asymptotic properties of the empirical bivariate mrl function. We present three theorems, of Kulkarni and Rattihalli (2002), that extend
Yang's (1978) results to the bivariate case. It turns out that $\hat{e}(x_1, x_2)$ is asymptotically unbiased. The bias tends to 0 at an exponential rate with increasing sample size, and uniform strong consistency of $\hat{e}(x_1, x_2)$ over every bounded rectangular subset of the support is established. Weak convergence of $\hat{e}(x_1, x_2)$ to a bivariate Gaussian process is also presented.

Let $\mathcal{X} = \{(x_1, x_2) : S(x_1, x_2) > 0\}$. The following theorems establish the asymptotic unbiasedness, uniform strong consistency, and weak convergence (after a proper normalization) to a bivariate Gaussian process of $\hat{e}(x_1, x_2)$.

**Theorem 3.2.1.** (Zahedi (1985)). When $E[X_i^2] < \infty$, then $\text{Var}[\hat{e}_{X_i}(x_1, x_2)] \to 0$ as $n \to \infty$.

**Proof.** Since $E[X_i^2] < \infty$, $\text{Var}[X_i - x_i | (X_1, X_2) \geq (x_1, x_2)] < \infty$ and the proof follows, from equation (3.2.7), by showing $\lim_{n \to 0} \sum_{j=1}^{n} \frac{1}{j} B(n, j, S(x_1, x_2)) = 0$. Note that $B(n, j, S(x_1, x_2)) = \binom{n}{j}[S(x_1, x_2)]^j[1 - S(x_1, x_2)]^{n-j}$. Note that,

$$
\sum_{j=1}^{n} \frac{1}{j} B(n, j, S(x_1, x_2)) \leq \sum_{j=0}^{n} \frac{2}{j+1} B(n, j, S(x_1, x_2)) = \sum_{j=0}^{n} \frac{2\binom{n}{j}}{j+1} S(x_1, x_2)^j[1 - S(x_1, x_2)]^{n-j}.
$$

By letting $k = j + 1$, the above expression reduces to:

$$
\sum_{j=1}^{n} \frac{1}{j} B(n, j, S(x_1, x_2)) = \frac{2}{(n+1)S(x_1, x_2)} \sum_{k=1}^{n+1} \binom{n+1}{k} S(x_1, x_2)^k(1 - S(x_1, x_2))^{n+1-k} = \frac{2\left[1 - [1 - S(x_1, x_2)]^{n+1}\right]}{(n+1)S(x_1, x_2)}.
$$
Since $0 \leq S(x_1, x_2) \leq 1$, it follows that

$$\lim_{n \to \infty} \sum_{j=1}^{n} \frac{1}{j} B(n, j, S(x_1, x_2)) \leq \lim_{n \to \infty} \frac{2\{1 - [1 - S(x_1, x_2)]^{n+1}\}}{(n+1)S(x_1, x_2)} = 0.$$ \hfill \Box

**Theorem 3.2.2.** The bivariate empirical mrl function is an asymptotically unbiased estimator of the true bivariate mrl function. That is

$$E[\hat{e}_{X_i}(x_1, x_2)] \to e_{X_i}(x_1, x_2) \text{ for } i = 1, 2$$

as $n \to \infty$ for all $(x_1, x_2) \in \mathcal{X}$.

**Proof.** From (3.2.6) we know that the mean of $\hat{e}_{X_i}(x_1, x_2)$ is given by $E[\hat{e}_{X_i}(x_1, x_2)] = \{1 - (1 - S(x_1, x_2))^n\} e_{X_i}(x_1, x_2)$. It follows that the bias of the empirical estimator is $-e_{X_i}(x_1, x_2)(1 - S(x_1, x_2))^n$ and hence $\hat{e}_{X_i}(x_1, x_2)$ is asymptotically unbiased, with bias decaying exponentially to zero as $n \to \infty$. \hfill \Box

**Lemma 3.2.3.** For $i = 1, 2$ let $X_{ni}^* = \max \{X_{n1}, X_{n2}, \ldots, X_{ni}\}$. Suppose that $X_{ni}^* \to \infty$ w.p.1 and $X_{ni}^* \sup_{(0,0) \leq (x_1, x_2) < (\infty, \infty)} |S_n(x_1, x_2) - S(x_1, x_2)| \to 0$ w.p.1. as $n \to \infty$ and for $i = 1, 2$. Then as $n \to \infty$,

$$\int_0^\infty |S_n(u, x_2) - S(u, x_2)| \, du \to 0, \text{ w.p.1} \quad (3.2.8)$$

and

$$\int_0^\infty |S_n(x_1, u) - S(x_1, u)| \, du \to 0 \text{ w.p.1.} \quad (3.2.9)$$
Proof. For brevity, we will only prove (3.2.8). The proof of (3.2.9) follows from the arguments of (3.2.8). Now, let us rewrite (3.2.8) as

\[
\int_0^\infty |S_n(u, x_2) - S(u, x_2)| \, du = \int_0^{X_{n1}^*} |S_n(u, x_2) - S(u, x_2)| \, du \\
+ \int_{X_{n1}^*}^\infty |S_n(u, x_2) - S(u, x_2)| \, du \\
\leq X_{n1}^* \times \left\{ \sup_{(0,0) \leq (x_1, x_2) < (\infty, \infty)} |S_n(x_1, x_2) - S(x_1, x_2)| \right\} \\
+ \int_{X_{n1}^*}^\infty S_n(u, x_2) \, du + \int_{X_{n1}^*}^\infty S(u, x_2) \, du.
\]

By the assumption of the lemma and by the definition of the bivariate empirical survival function respectively, both the first and second terms on the right side of the above inequality converge to zero with probability one. On the other hand, the third term may be rewritten as

\[
\int_{X_{n1}^*}^\infty S(u, x_2) \, du = \int_0^\infty S(u, x_2) \cdot I_{[X_{n1}^*, \infty]}(u) \, du.
\]

Note that \( S(u, x_2) \cdot I_{[X_{n1}^*, \infty]} \to 0 \) w.p.1 as \( n \to \infty \) and \( S(u, x_2) \cdot I_{[X_{n1}^*, \infty]} \leq S(u, x_2) \).

Since \( \int_0^\infty S(u, x_2) \, du \leq \int_0^\infty S(u) \, du \) and \( \int_0^\infty S(u) \, du = E[X_1] < \infty \), by the dominated convergence theorem, the third term converges to zero with probability one. Thus the result follows. \( \square \)

For fixed \((b_1, b_2) \in \mathcal{X}\), let \( D = [0, b_1] \times [0, b_2] \). The following theorem establishes the strong consistency of the bivariate empirical mrl function.

**Theorem 3.2.4.** Under the assumptions of Lemma 3.2.3 the bivariate empirical mrl
function is strongly consistent. That is,

\[ \| \hat{e}(x_1, x_2) - e(x_1, x_2) \| \to 0 \]

with probability one.

Proof. Since the consistency of \( \hat{e} \) is equivalent to the consistency of both marginals, it is sufficient to prove the above theorem for one of the marginals. Here we will prove the consistency of the first component.

\[
\left| \hat{e}_{X_1}(x_1, x_2) - e_{X_1}(x_1, x_2) \right| = \left| \frac{\int_{x_1}^{\infty} S_n(u, x_2)du}{S_n(x_1, x_2)} - \frac{\int_{x_1}^{\infty} S(u, x_2)du}{S(x_1, x_2)} \right|
\]

\[
= \left| \frac{S(x_1, x_2) \int_{x_1}^{\infty} S_n(u, x_2)du - S_n(x_1, x_2) \int_{x_1}^{\infty} S(u, x_2)du}{S_n(x_1, x_2)S(x_1, x_2)} \right|
\]

\[
\leq \frac{\int_{x_1}^{\infty} |S_n(u, x_2) - S(u, x_2)| du}{S_n(x_1, x_2)}
\]

\[
+ \frac{|S_n(x_1, x_2) - S(x_1, x_2)|}{S_n(x_1, x_2)S(x_1, x_2)} \int_{x_1}^{\infty} S(u, x_2)du.
\]

Since \( S^{-1}(x_1, x_2) < \infty \) for all \((x_1, x_2) \in \mathcal{X}\), by Lemma 3.2.3 the first term of the right side of the above inequality converge to zero with probability one. The convergence of the second term to zero follows from the convergence result of \( S_n(x_1, x_2) \) and the fact that \( \int_{x_1}^{\infty} S(u, x_2)du \) is finite. This completes the proof. Note that Kulkarni and Rattihalli (2002) showed the strong uniform convergence of the bivariate empirical mrl function on closed rectangles without the conditions of Lemma 3.2.3.

Now let \( \{Z(x_1, x_2), (x_1, x_2) > 0\} \) denote the bivariate Gaussian process which is obtained as the weak limit of the process \( \{\sqrt{n} [S_n(x_1, x_2) - S(x_1, x_2)], (x_1, x_2) > 0\} \).
For all \((x_1, x_2) \in \mathcal{X}\), let us denote the bivariate empirical mrl process by \(\hat{Z}(x_1, x_2)\) where

\[
\hat{Z}(x_1, x_2) = \sqrt{n} \left[ \hat{e}(x_1, x_2) - e(x_1, x_2) \right]
\]

\[
= \sqrt{n} \left[ \hat{e}_1(x_1, x_2) - e_1(x_1, x_2), \hat{e}_2(x_1, x_2) - e_2(x_1, x_2) \right]
\]

The following result establishes the weak convergence of \(\hat{Z}(x_1, x_2)\) for all \((x_1, x_2) \in D\).

**Theorem 3.2.5.** The process \(\sqrt{n} \left[ \hat{e}(x_1, x_2) - e(x_1, x_2), (x_1, x_2) \in D \right]\) converges weakly to a zero-mean bivariate Gaussian process \((W_1(x_1, x_2), W_2(x_1, x_2))\) where

\[
W_1(x_1, x_2) = S^{-1}(x_1, x_2) \int_{x_1}^{\infty} Z(u, x_2) du - S^{-2}(x_1, x_2) Z(x_1, x_2) \int_{x_1}^{\infty} S(u, x_2) du
\]

and

\[
W_2(x_1, x_2) = S^{-1}(x_1, x_2) \int_{x_2}^{\infty} Z(x_1, u) du - S^{-2}(x_1, x_2) Z(x_1, x_2) \int_{x_2}^{\infty} S(x_1, u) du
\]

with covariance function given by

\[
Var \{W_1(x_1, x_2)\} = E \left[ \left( \int_{x_1}^{\infty} Z(u, x_2) du \right)^2 \right] + \frac{\int_{x_1}^{\infty} S(u, x_2) du}{S^4(x_1, x_2)} E \left[ Z^2(x_1, x_2) \right] - 2 \int_{x_1}^{\infty} S(u, x_2) du \frac{S(u, x_2)}{S^3(x_1, x_2)} E \left[ Z(x_1, x_2) \int_{x_1}^{\infty} Z(u, x_2) du \right],
\]

\[
Var \{W_2(x_1, x_2)\} = E \left[ \left( \int_{x_2}^{\infty} Z(x_2, u) du \right)^2 \right] + \frac{\int_{x_2}^{\infty} S(x_1, u) du}{S^4(x_1, x_2)} E \left[ Z^2(x_1, x_2) \right] - 2 \int_{x_2}^{\infty} S(x_1, u) du \frac{S(x_1, u)}{S^3(x_1, x_2)} E \left[ Z(x_1, x_2) \int_{x_1}^{\infty} Z(x_1, u) du \right],
\]
\[
\text{Cov} \{W_1(x_1, x_2), W_2(y_1, y_2)\} = \frac{E \left[ \int_{y_2}^{\infty} Z(u, x_2)du \int_{y_1}^{\infty} Z(y_1, u)du \right]}{S(x_1, x_2)S(y_1, y_2)} \\
- \frac{\int_{y_2}^{\infty} S(y_1, u)du}{S(x_1, x_2)S^2(y_1, y_2)} E \left[ Z(y_1, y_2) \int_{y_1}^{\infty} Z(u, x_2)du \right] \\
- \frac{\int_{y_1}^{\infty} S(u, x_2)du}{S^2(x_1, x_2)S(y_1, y_2)} E \left[ Z(x_1, x_2) \int_{y_2}^{\infty} Z(y_1, u)du \right] \\
+ \frac{\int_{y_1}^{\infty} S(u, x_2)du \int_{y_2}^{\infty} S(y_1, u)du}{S^2(x_1, x_2)S^2(y_1, y_2)} E [Z(x_1, x_2)Z(y_1, y_2)]
\]

for all \((x_1, x_2) \in \mathcal{X}, (y_1, y_2) \in \mathcal{X}\) such that \(x_1 \leq y_1\) and \(x_2 \leq y_2\).

**Proof.** Let \(\{Z(x_1, x_2), (x_1, x_2) > 0\}\) denote the bivariate Gaussian process which is obtained as the weak limit of \(Z_n(x_1, x_2) = \{\sqrt{n} [S_n(x_1, x_2) - S(x_1, x_2)], (x_1, x_2) > 0\}\).

For fixed \((x_1, x_2) \in \mathcal{X}\), we have

\[
\hat{Z}_1(x_1, x_2) = \sqrt{n} \{\hat{e}_{X_1}(x_1, x_2) - e_{X_1}(x_1, x_2)\} \tag{3.2.11}
\]

\[
= S_n^{-1}(x_1, x_2) \int_{x_1}^{\infty} \sqrt{n} \{S_n(u, x_2) - S(u, x_2)\} du \\
- \frac{\sqrt{n} [S_n(x_1, x_2) - S(x_1, x_2)]}{S_n(x_1, x_2)S(x_1, x_2)} \int_{x_1}^{\infty} S(u, x_2)du \\
= S_n^{-1}(x_1, x_2) \int_{x_1}^{\infty} Z_n(u, x_2)du - \frac{Z_n(x_1, x_2)}{S_n(x_1, x_2)S(x_1, x_2)} \int_{x_1}^{\infty} S(u, x_2)du
\]

and

\[
\hat{Z}(x_1, x_2) = \sqrt{n} \{\hat{e}_{X_2}(x_1, x_2) - e_{X_2}(x_1, x_2)\} \tag{3.2.12}
\]

\[
= S_n^{-1}(x_1, x_2) \int_{x_2}^{\infty} \sqrt{n} \{S_n(x_1, u) - S(x_1, u)\} du \\
- \frac{\sqrt{n} [S_n(x_1, x_2) - S(x_1, x_2)]}{S_n(x_1, x_2)S(x_1, x_2)} \int_{x_2}^{\infty} S(x_1, u)du
\]
\[ = S_n^{-1}(x_1, x_2) \int_{x_2}^{\infty} Z_n(x_1, u)du - \frac{Z_n(x_1, x_2)}{S_n(x_1, x_2)} \int_{x_2}^{\infty} S(x_1, u)du. \]

For fixed \((a, b) \in \mathcal{X}\), let \(D([0, a] \times [0, b])\) be the space of functions on the rectangle \([0, a] \times [0, b]\) that are right continuous and have left-hand limits. Let \(d\) be the Skorohod metric on \(D([0, a] \times [0, b])\) and for \(i = 1, 2\) let us define a map \(H_i : D([0, a] \times [0, b]) \to D([0, a] \times [0, b])\) by

\[ H_1(Z(x_1, x_2)) = S^{-1}(x_1, x_2) \int_{x_1}^{\infty} Z(u, x_2)du - \frac{Z(x_1, x_2)}{S^2(x_1, x_2)} \int_{x_2}^{\infty} S(u, x_2)du \]

and

\[ H_2(Z(x_1, x_2)) = S^{-1}(x_1, x_2) \int_{x_2}^{\infty} Z(x_1, u)du - \frac{Z(x_1, x_2)}{S^2(x_1, x_2)} \int_{x_2}^{\infty} S(x_1, u)du \]

for \(Z(x_1, x_2) \in D([0, a] \times [0, b])\). Then \(H_1\) and \(H_2\) are continuous maps with respect to \(d\) (See Yang (1978)). Thus by the continuity theorem of Billingsley (1968), Theorem B.1.2, the result follows.

Note that Kulkarni and Rattihalli (2002) showed weak convergence of \(\hat{Z}(x_1, x_2)\) for all \((x_1, x_2) \in \mathcal{X}\) where \(\mathcal{X} = \{(x_1, x_2) : S(x_1, x_2) > 0\}\).

### 3.3 Tests and Confidence Sets for MRL Functions

In this section we develop a test and confidence procedures for comparing mrl functions from two populations or treatment groups. Let \(X_i = (X_1, X_2), i = 1, 2, \ldots, n\) be a random vector in \(\mathbb{R}^2_+\) with corresponding survival and mrl functions \(S(x_1, x_2)\)
and $e(x_1, x_2)$ respectively. Moreover, let $e^0(x_1, x_2)$ be the bivariate mrl function of the standard population (treatment group) and assume that $e^0(x_1, x_2)$ is known for all $(x_1, x_2) \in \mathcal{X}$ where $\mathcal{X} = \{(x_1, x_2) : S(x_1, x_2) > 0\}$. We will consider the problem of testing the hypothesis $H_0: e(x_1, x_2) = e^0(x_1, x_2)$ and a related problem of finding the confidence region for $e(x_1, x_2)$ for all $(x_1, x_2) \in \mathcal{X}$.

Recall that in the previous section it was shown that $\sqrt{n}[\hat{e}(x_1, x_2) - e(x_1, x_2), (x_1, x_2) \in \mathbb{R}_+^2]$ converges weakly to a zero mean bivariate Gaussian process with some covariance matrix $\Sigma$, where $\Sigma$ is a $2 \times 2$ matrix and is given in Section 2 of Chapter 4. This implies that under $H_0$, $n[(\hat{e}(x_1, x_2) - e^0(x_1, x_2))^t \Sigma^{-1} [\hat{e}(x_1, x_2) - e^0(x_1, x_2)]]$ has a central chi-square distribution with 2 degrees of freedom and hence the test rejects $H_0: e(x_1, x_2) = e^0(x_1, x_2)$ whenever

$$n [\hat{e}(x_1, x_2) - e^0(x_1, x_2)]^t \Sigma^{-1} [\hat{e}(x_1, x_2) - e^0(x_1, x_2)] \geq \chi^2_{2,\alpha}$$  \hspace{1cm} (3.3.1)$$

where $\chi^2_{2,\alpha}$ is a constant such that $P(\chi^2_2 \geq \chi^2_{2,\alpha}) = \alpha$. The test has power function which increases monotonically (see, e.g. Giri (1996)) with the non-centrality parameter $n[e(x_1, x_2) - e^0(x_1, x_2)]^t \Sigma^{-1}[e(x_1, x_2) - e^0(x_1, x_2)]$. Thus the power function of the test given in (3.3.1) has the minimum value $\alpha$ (level of significance) when $e(x_1, x_2) = e^0(x_1, x_2)$ and its power is greater than $\alpha$ when $e(x_1, x_2) \neq e^0(x_1, x_2)$. For a given bivariate empirical mrl function $\hat{e}(x_1, x_2)$, consider the inequality

$$n [\hat{e}(x_1, x_2) - e(x_1, x_2)]^t \Sigma^{-1} [\hat{e}(x_1, x_2) - e(x_1, x_2)] \leq \chi^2_{2,\alpha}.$$  \hspace{1cm} (3.3.2)$$

Under $H_0$, the probability is $(1 - \alpha)$ that the empirical bivariate mrl function satisfies (3.3.2). Thus the set of values of $e(x_1, x_2)$ satisfying (3.3.2) gives an asymptotic
confidence region for \( e(x_1, x_2) \) with confidence coefficient \( 1 - \alpha \), and represents the interior and the surface of an ellipsoid with center \( \hat{e}(x_1, x_2) \), with shape depending on \( \Sigma \) and size depending on \( \Sigma \) and \( \chi^2_{2,\alpha} \).

Let \((X_{1i}, X_{2i})\) and \((Y_{1i}, Y_{2i})\), \(i = 1, 2, \ldots, n\), be independent random vectors with survival functions \( S_1(x_1, x_2) \) and \( S_2(x_1, x_2) \), and mrl functions \( e_1(x_1, x_2) \) and \( e_2(x_1, x_2) \) respectively. Assume both \( e_1(x_1, x_2) \) and \( e_2(x_1, x_2) \) are unknown. We now consider the problem of testing the hypothesis \( H_0 : e_1(x_1, x_2) - e_2(x_1, x_2) = 0 \) for some \((x_1, x_2) \in X\) and the problem of setting a confidence region with confidence coefficient \((1 - \alpha)\). As before, it turns out that \( n^{1/2} [\hat{e}_j(x_1, x_2) - e_j(x_1, x_2), (x_1, x_2) \in \mathbb{R}^2_j] \) converges weakly to a zero mean bivariate Gaussian process with some covariance matrix \( \Sigma_j \) for \( j = 1, 2 \), where \( \Sigma_j \) is a \( 2 \times 2 \) matrix and is given in Section 3 of Chapter 4. Thus under \( H_0 \),

\[
n [\hat{e}_1(x_1, x_2) - \hat{e}_2(x_1, x_2)]^T \Sigma^{-1} [\hat{e}_1(x_1, x_2) - \hat{e}_2(x_1, x_2)] \text{ for } \Sigma = \Sigma_1 + \Sigma_2 \text{ is distributed as chi-square with 2 degrees of freedom. Therefore, an asymptotic size } \alpha \text{ test rejects } H_0 \text{ if and only if}
\]

\[
n [\hat{e}_1(x_1, x_2) - \hat{e}_2(x_1, x_2)]^T \Sigma^{-1} [\hat{e}_1(x_1, x_2) - \hat{e}_2(x_1, x_2)] \geq \chi^2_{2,\alpha}. \tag{3.3.3}
\]

The test given in (3.3.3) has a power function which increases monotonically with non-centrality parameter \( n [e_1(x_1, x_2) - e_2(x_1, x_2)]^T \Sigma^{-1} [e_1(x_1, x_2) - e_2(x_1, x_2)] \); its power is greater than \( \alpha \) whenever \( e_1(x_1, x_2) \neq e_2(x_1, x_2) \), and the power function attains its minimum value \( \alpha \) whenever \( e_1(x_1, x_2) = e_2(x_1, x_2) \). The asymptotic confidence region for \( e_1(x_1, x_2) - e_2(x_1, x_2) \) with confidence coefficient \((1 - \alpha)\) is given by the set of
values \( e_1(x_1, x_2) - e_2(x_1, x_2) \) satisfying

\[
\begin{align*}
    n \left[ \hat{\epsilon}_1(x_1, x_2) - \hat{\epsilon}_2(x_1, x_2) \right] - (e_1(x_1, x_2) - e_2(x_1, x_2)) &\right] I(\Gamma) \leq \chi^2_{2,\alpha}.
\end{align*}
\] (3.3.4)

3.4 Estimating the Bivariate MRL Function under Censoring

We now turn our attention to the censored data case. Let \((X_i, Y_i), i = 1, 2, \ldots, n\) be independent and identically distributed pairs of failure times with survival function \(S(x, y) = P(X > x, Y > y)\) and let \(C_i, i = 1, 2, \ldots, n\) be independent and identically distributed censoring times with survival function \(G(t) = P(C > t)\). Suppose that the two sequences \((X_i, Y_i)_{i=1}^n\) and \((C_i)_{i=1}^n\) are independent. In the univariate random-censoring-from-the-right model, the \((X_i, Y_i)\) are censored on the right by the single censoring variable \(C_i\), so that the data consists of the random vectors \((\tilde{X}_i, \tilde{Y}_i, \delta_i^x, \delta_i^y), i = 1, 2, \ldots, n\), where \(\tilde{X}_i = X_i \land C_i, \tilde{Y}_i = Y_i \land C_i, \delta_i^x = I(X_i \leq C_i)\) and \(\delta_i^y = I(Y_i \leq C_i)\). The survival function of the observed pairs \(\{\tilde{X}_i, \tilde{Y}_i\}_{i=1}^n\) is \(S(x, y)\tilde{G}(x \lor y)\). This is a simple consequence of the independence between \((X, Y)\) and \(C\). Thus, under univariate censoring, it is natural to estimate the survival function \(S(x, y)\) by

\[
\hat{S}_n(x, y) = \frac{n^{-1} \sum_{i=1}^n I(\tilde{X}_i > x, \tilde{Y}_i > y)}{\tilde{G}(x \lor y)}
\] (3.4.1)
where the numerator is the empirical estimator for the survival function of the observed pairs and the denominator is the product-limit estimator for \( G(.) \). Note that \( G(.) \) is estimated from the data \((\tilde{C}_i, \delta_i^x) (i = 1, \ldots, n)\) by the Kaplan-Meier estimator, where \( \tilde{C}_i = C_i \wedge (X_i \vee Y_i) = \tilde{X}_i \vee \tilde{Y}_i, \delta_i^x = I\{C_i \leq (X_i \vee Y_i)\} = 1 - \delta_i^x \delta_i^y \). Let \( \{c_k, k \geq 1\} \) be the distinct time points at which censoring occurs, and let \( d_k \) be the number of censored observations at \( c_k \). Then the product-limit estimator for \( \tilde{G}(.) \) is given by:

\[
\tilde{G}(x \vee y) = \prod_{k: c_k \leq (x \vee y)} \frac{n_k - d_k}{n_k}
\]

where \( n_k = \sum_{i=1}^{n} I(\tilde{C}_i \geq c_k) \). Lin and Ying (1993) considered the estimator (3.4.1) and showed that \( \hat{S}_n(x, y) \) is strongly uniformly consistent, and upon proper normalization, converges weakly to a zero mean Gaussian process for all \((x, y) \in [0, \tau]^2\), where \( \tau \) satisfies \( S(\tau, \tau) \tilde{G}(\tau) > 0 \). Using \( \hat{S}_n(x, y) \) one can estimate \( e(x, y) = (e_1(x, y), e_2(x, y)) \), under univariate censoring, by \( \hat{e}_n(x, y) = (\hat{e}_{1,n}(x, y), \hat{e}_{2,n}(x, y)) \) where

\[
\hat{e}_{i,n}(x, y) = \left\{ \begin{array}{ll}
\hat{S}_n(x, y)^{-1} \int_x^\infty \hat{S}_n(u, y) du, & \text{if } i = 1, \\
\hat{S}_n(x, y)^{-1} \int_y^\infty \hat{S}_n(x, v) dv, & \text{if } i = 2.
\end{array} \right.
\]

(3.4.2)

In the absence of censoring \( \hat{e}_n(x, y) \) reduces to the empirical bivariate mrl function. Sohn et al (1996) showed uniform consistency and weak convergence of \( \hat{e}_n(x, y) \) on bounded rectangles and under the assumption that both \( \sqrt{n} \int_{\tilde{X}} S(u, y) du \) and \( \sqrt{n} \int_{\tilde{Y}} S(x, v) dv \) converge to zero in probability, where \( \tilde{X}^* = \max(\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n) \) and \( \tilde{Y}^* = \max(\tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_n) \). More specifically, they proved the following Theorem:

**Theorem 3.4.1.** Suppose that \( \sqrt{n} \int_{\tilde{X}} S(u, y) du \to 0 \) w.p.1. Then, as \( n \to \infty \)

(i) \( \sup_{(x,y) \in [0,\tau]^2} |\hat{e}_{1,n}(x, y) - e_1(x, y)| \to 0 \) w.p.1.
(ii) $\sqrt{n}\{\hat{e}_{1,n}(x,y) - e_1(x,y), (x,y) \in \mathbb{R}^2_+\} \Rightarrow W_1(x,y)$, where $W_1(.,.)$ is a mean zero bivariate Gaussian process.

Note that the asymptotic properties of the second component of the estimator defined in (3.4.2) can be obtained from the above Theorem by simply replacing $\hat{X}^*$ by $\hat{Y}^*$ and $S(u,y)$ by $S(x,v)$.

### 3.5 Simulation Results

Simulation studies were carried out to examine the properties of the bias and mean squared error of the empirical estimator as a function of various sample sizes 15, 30 and 45. Although formulas for the bias and variance of the empirical mean residual life function exist, the variance is difficult to compute. Therefore, simulations were carried out to assess the behaviour of the estimator. Each simulation consisted of a series of 10000 trials. Several bivariate distributions were used for the simulation study:

- Bivariate Gumbel distribution with density function

$$f(x, y; \theta) = [(1 + \theta x)(1 + \theta y) - \theta] \exp(-x - y - \theta xy)$$

where $x, y > 0; 0 \leq \theta \leq 1$. $\theta$ measures the correlation between the random variables $X$ and $Y$. The correlation is zero for $\theta = 0$, and it decreases to $-0.43$ as $\theta$ increases to 1.
• Bivariate pareto distribution with density function

\[ f(x_1, x_2, \theta_1, \theta_2, a) = a(a + 1)(\theta_1 \theta_2)^{(a+1)}(\theta_2 x_1 + \theta_1 x_2 - \theta_1 \theta_2)^{-(a+2)} \]

where \( x_1 \geq \theta_1 > 0, \ x_2 \geq \theta_2 > 0, \ a > 0 \). The correlation between \( X_1 \) and \( X_2 \) goes to 0 as \( a \to \infty \) and becomes close to 1 as \( a \to 0 \).

• Bivariate Morgenstern (1956) distribution with density function

\[ f(u_1, u_2, \alpha) = 1 + \alpha(2u_1 - 1)(2u_2 - 1) \]

where \( 0 \leq u_1, u_2 \leq 1, \ -1 \leq \alpha \leq 1 \). The dependence structure between \( U_1 \) and \( U_2 \) is controlled by \( \alpha \). The correlation between \( U_1 \) and \( U_2 \) is \( 1/\alpha \).

• Bivariate Sarmanov (1966) distribution with density function

\[ f(x_1, x_2, \alpha) = e^{-x_1}e^{-x_2}\{1 + \alpha(2e^{-x_1} - 1)(2e^{-x_2} - 1)\} \]

where \( 0 \leq x_1, x_2 < \infty, \ -1 \leq \alpha \leq 1 \). The correlation between \( X_1 \) and \( X_2 \) is given by \( 1/\alpha \).

Surface plots for the density of the above four distributions are given in Figure 3.1 and simulation results are plotted in Appendix C. Careful observation of the simulation results reveals the following patterns:

• The bias and mean squared error of the empirical estimator decrease with increasing sample size, as expected.
Figure 3.1: (a) Gumbel density with $\theta = 0.5$; (b) Pareto density with $a = 3$, $\theta_1 = 0.8$, $\theta_2 = 0.4$; (c) Morgenstern density with $\alpha = 1$ and (d) Sarmanov density with $\alpha = -1$. 

- The bias of the empirical estimator increase in \( x(y) \) for fixed \( y(x) \). The increment is more prominently reflected at the extreme values of \( x \) and \( y \) and for small \( n \).

- In the case of the Gumbel distribution the mean squared error of the first (second) component of the bivariate empirical function decreases in \( x(y) \) for fixed \( y(x) \) and the mean squared error of both components becomes small at large values of \( x \) and \( y \). However, in the case of the Morgenstern distribution, the mean squared error of the first (second) component of the bivariate empirical function increases in \( x(y) \) for fixed \( y(x) \). Otherwise the mean squared error of both components become small at large values of \( x \) and \( y \).

- In the case of the Sarmanov distribution, the mean squared error of the empirical mrl function increases in \( x \) and \( y \). Though the pattern is not clearly predictable, the mean squared error of the empirical mrl estimator in the case of the pareto distribution also increases in \( x \) and \( y \).

In the cases of Gumbel and pareto distributions, our simulation results agree with the simulation results of Kulkarni and Rattihalli (2002). In their simulation studies, Kulkarni and Rattihalli (2002) have not considred the Sarmanov and Morgenstern distributions.
Chapter 4

Estimation of Two Bivariate

Ordered MRL Functions

4.1 Introduction

Situation frequently arise in practice in which mrl functions of two populations must be ordered. For example, if a mechanical device is improved, the mrl function for the improved device should not be less that of the original device. Also, mrl functions for medical patients are often ordered depending on the status of concomitant variables. The Diabetic Retinopathy Study (DRS) data given in Chapter 5 is a good example where a natural ordering among mrl functions exists. In this data set, the survival times of both left and right eyes are given for two groups of patients: juvenile and adult diabetics. Thus, it seems natural to assume that the mrl for the juvenile diabetics is
longer than the mrl of the adult diabetics. A detailed explanation of the DRS data, together with its analysis, is given in Chapter 5.

The main contribution of this thesis are the results covered in this chapter. The results provide estimators for the bivariate mean residual lifetime function when a partial ordering between mean residual functions obtains. The finite sample properties of these estimators as well as their asymptotic distributions are delineated. The organization of the Chapter is as follows: Section 4.2 considers the problem of non-parametric estimation of a bivariate mrl function when it is bounded from above by another known mrl function. It is demonstrated that the new estimator is a projection in a sense to be made precise later. As a consequence, the new estimator dominates the empirical mrl function in terms of mean squared error and renders it inadmissible with respect to a wide class of loss functions. Section 4.3 considers the same problem except that the bounding function is an unknown mrl function. In both Section 4.2 and Section 4.3, the asymptotic theory of the estimators is developed and results from simulation studies for both cases are discussed in Section 4.7. Section 4.4 discusses briefly the "mirror image" problem where the mrl function of interest is bounded from below by another mrl function which can be known or unknown. Section 4.5 is an extensions of Section 4.2 in the censored data case and Section 4.6 deals with tests and confidence sets for mrl functions.

In the sequel, \( \|x_1, x_2\| \) will denote the norm of \((x_1, x_2)\) defined by \( \|(x_1, x_2)\| = \max \{|x_1|, |x_2|\} \). Let \( a = (a_1, a_2) \) and \( b = (b_1, b_2) \) be vectors. Then \( a \wedge b = \min(a, b) = \)
(a_1 \land b_1, a_2 \land b_2), a \lor b = \max(a, b) = (a_1 \lor b_1, a_2 \lor b_2) \text{ and } a \leq b \text{ will denote the order in } \mathbb{R}^2 \text{ defined by } (a_1 \leq b_1, a_2 \leq b_2). \text{ Moreover, for functions } f = (f_1, f_2) \text{ with } f_i : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+ \text{ where } \mathbb{R}_+^2 = \{(x, y) : x \geq 0 \text{ and } y \geq 0\} \text{ and } \mathbb{R}_+ = \{x : x \geq 0\}, \text{ we define } \|f\|_p \text{ by } \|f\|_p = \left( \int_0^\infty \int_0^\infty \max(f_1^p, f_2^p) \, dx_1 \, dx_2 \right)^{\frac{1}{p}} \text{ for } p \geq 1.

### 4.2 The One Sample Problem

#### 4.2.1 The estimator under mrl ordering

Suppose that \( X = (X_1, X_2) \) and \( Y = (Y_1, Y_2) \) are random vectors with finite means representing lifetimes of two populations with distribution functions \( F(x_1, x_2) \) and \( G(y_1, y_2) \); survival functions \( \overline{F}(x_1, x_2) \) and \( \overline{G}(y_1, y_2) \); and mrl functions \( e(x_1, x_2) \) and \( m(y_1, y_2) \) respectively. Let \( \mathcal{X} = \{(x_1, x_2) : \overline{F}(x_1, x_2) > 0 \text{ and } \overline{G}(x_1, x_2) > 0\} \). Consider the problem of estimating \( e(x_1, x_2) \) when \( e(x_1, x_2) \leq m(x_1, x_2) \) for all \((x_1, x_2)\), where \( m(x_1, x_2) \) is a known mean residual function. Motivated by the work of Rojo (1995), Rojo and Ma (1996) and Rojo (2004) we propose as an estimator of \( e(x_1, x_2) \) the following:

\[
e_n^*(x_1, x_2) = \hat{e}_n(x_1, x_2) \land m(x_1, x_2) \quad (4.2.1)
\]

\[
= (\hat{e}_{1,n}(x_1, x_2) \land m_1(x_1, x_2), \hat{e}_{2,n}(x_1, x_2) \land m_2(x_1, x_2))
\]

\[
= \frac{1}{2} \left\{ \left[ \hat{e}_{1,n}(x_1, x_2) + m_1(x_1, x_2) - |\hat{e}_{1,n}(x_1, x_2) - m_1(x_1, x_2)| \right] \right. \\
\left. \left. \left[ \hat{e}_{2,n}(x_1, x_2) + m_2(x_1, x_2) - |\hat{e}_{2,n}(x_1, x_2) - m_2(x_1, x_2)| \right] \right\}
\]
where $\hat{e}_n(x_1, x_2)$ is the bivariate empirical mrl function. Thus, our estimator modifies the empirical mrl only when it violates the restriction, in which case, it replaces the empirical mrl by the “benchmark” mrl function $m(x_1, x_2)$. As a consequence of this, it is not difficult to show that $e^*_n$ is the projection of the empirical estimator onto the set of bivariate mrl functions bounded above by $m(x_1, x_2)$, in a way to be made precise later. It is clear that $e^*_n$ is more negatively biased than $\hat{e}_n$. Nevertheless, $e^*_n$ has uniformly smaller mean squared error than $\hat{e}_n$ as stated in the next result.

**Theorem 4.2.1.** The estimator $e^*_n(x_1, x_2)$ has smaller mean squared error than the empirical mrl estimator $\hat{e}_n(x_1, x_2)$ for every $(x_1, x_2)$. That is,

$$
E \left\{ \sum_{j=1}^{2} [(e^*_{j,n}(x_1, x_2) - e_j(x_1, x_2))^2] \right\} \leq E \left\{ \sum_{j=1}^{2} [(\hat{e}_{j,n}(x_1, x_2) - e_{X_j}(x_1, x_2))^2] \right\}.
$$

**Proof.** Let

$$
Q_1 = \{ (x_1, x_2) : \hat{e}_{1,n}(x_1, x_2) > m_1(x_1, x_2), \hat{e}_{2,n}(x_1, x_2) \leq m_2(x_1, x_2) \}
$$

$$
Q_2 = \{ (x_1, x_2) : \hat{e}_{1,n}(x_1, x_2) \leq m_1(x_1, x_2), \hat{e}_{2,n}(x_1, x_2) > m_2(x_1, x_2) \}
$$

$$
Q_{12} = \{ (x_1, x_2) : \hat{e}_{1,n}(x_1, x_2) > m_1(x_1, x_2), \hat{e}_{2,n}(x_1, x_2) > m_2(x_1, x_2) \}
$$

$$
Q = \{ (x_1, x_2) : \hat{e}_{1,n}(x_1, x_2) \leq m_1(x_1, x_2), \hat{e}_{2,n}(x_1, x_2) \leq m_2(x_1, x_2) \}.
$$

Then for $j = 1$,

$$
|e^*_{1,n}(x_1, x_2) - e_1(x_1, x_2)| = |m_1(x_1, x_2) - e_1(x_1, x_2)|I_{Q_1 \cup Q_{12}}
$$

$$
+ |\hat{e}_{1,n}(x_1, x_2) - e_1(x_1, x_2)|I_{Q_2 \cup Q}
$$

$$
\leq |\hat{e}_{1,n}(x_1, x_2) - e_1(x_1, x_2)|I_{Q_1 \cup Q_{12}}
$$

$$
+ |\hat{e}_{1,n}(x_1, x_2) - e_1(x_1, x_2)|I_{Q_2 \cup Q}.
$$
\[ = |\hat{e}_{1,n}(x_1, x_2) - e_1(x_1, x_2)|. \]

A similar statement holds for \( e_{2,n}^* \), and therefore, for \( j = 1, 2 \),

\[ |e_{j,n}^*(x_1, x_2) - e_j(x_1, x_2)| \leq |\hat{e}_{j,n}(x_1, x_2) - e_j(x_1, x_2)|. \] (4.2.2)

The result then follows immediately.

\[ \square \]

An examination of the proof of Theorem 4.2.1 immediately shows that the empirical mrl function \( \hat{e}_n(x_1, x_2) \) is rendered inadmissible with respect to any loss function of the form \( L(e, \hat{e}) = v(||e - \hat{e}||) \) with \( v(0) = 0 \) and \( v(x) \) nondecreasing on \((0, \infty)\), since it is dominated in risk by the estimator \( e_n^*(x_1, x_2) \). As it turns out, \( e_n^*(x_1, x_2) \) is the projection of the empirical mrl onto the convex set of mrl functions \( k(x_1, x_2) \) such that \( k(x_1, x_2) \leq m(x_1, x_2) \). The interpretation of our estimator as a projection onto an appropriate convex set is provided by the following result.

**Theorem 4.2.2.** Let \( A \) be the convex set of all mrl functions bounded above by a known mrl function \( m(x_1, x_2) \). That is, let \( A = \{ k(x_1, x_2) : k \) is a mrl and \( k(x_1, x_2) \leq m(x_1, x_2) \) for all \( (x_1, x_2) \in X \} \). Let \( e_n^*(x_1, x_2) = \hat{e}_n(x_1, x_2) \wedge m(x_1, x_2) \). Then,

(i) For any \( k \in A \),

\[ \sup_{0 \leq (x_1, x_2) \in X} \| e_n^*(x_1, x_2) - \hat{e}_n(x_1, x_2) \| \leq \sup_{0 \leq (x_1, x_2) \in X} \| k(x_1, x_2) - \hat{e}_n(x_1, x_2) \|. \]

(ii) For \( p \geq 1 \) and for all \( k \in A \)

\[ \| e_n^* - \hat{e}_n \|_p \leq \| k - \hat{e}_n \|_p. \]
Proof. Let \( A = \{ k(x_1, x_2) : k \text{ is a mrl and } k(x_1, x_2) \leq m(x_1, x_2) \text{ for all } (x_1, x_2) \in \mathcal{X} \} \).

We prove (i) first. Note that
\[
e^*_n(x_1, x_2) - \hat{e}_n(x_1, x_2) = \begin{cases} 
0 & \text{if } \hat{e}_n(x_1, x_2) \leq m(x_1, x_2) \\
\hat{e}_n(x_1, x_2) \wedge m(x_1, x_2) - \hat{e}_n(x_1, x_2) & \text{otherwise.}
\end{cases}
\]

Now if \( \hat{e}_n(x_1, x_2) \leq m(x_1, x_2) \), then \( \hat{e}_n(x_1, x_2) \wedge m(x_1, x_2) < \hat{e}_n(x_1, x_2) \) with \( \hat{e}_n(x_1, x_2) \wedge m(x_1, x_2) \in A \). Moreover, letting \( e^*_n(x_1, x_2) = (e^*_{1,n}(x_1, x_2), e^*_{2,n}(x_1, x_2)) = \hat{e}_n(x_1, x_2) \wedge m(x_1, x_2) \), it follows easily that for all \( k(x_1, x_2) = (k_1(x_1, x_2), k_2(x_1, x_2)) \in A \),
\[
|e^*_{i,n}(x_1, x_2) - \hat{e}_{i,n}(x_1, x_2)| \leq |k_i(x_1, x_2) - \hat{e}_{i,n}(x_1, x_2)| \text{ for } i = 1, 2.
\]

Therefore, for each \( k \in A \),
\[
\sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \| e^*_n(x_1, x_2) - \hat{e}_n(x_1, x_2) \| \leq \sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \| k(x_1, x_2) - \hat{e}_n(x_1, x_2) \|.
\]

The proof of (ii) follows immediately from the above arguments.

\( \square \)

### 4.2.2 Asymptotic properties of the estimator

Theorem 4.2.1 attests to the superiority of the estimator defined in (4.2.1) when compared to the empirical mrl function for any finite sample size. We next concentrate on the asymptotic properties of \( e^*_n \). It turns out that \( e^*_n \) is asymptotically unbiased, uniformly strongly consistent, and converges weakly to a Gaussian process. We first discuss the bias of \( e^*_n \). In what follows let \( \sigma^2_1(x_1, x_2) = E[(X_1 - e_1(x_1, x_2))^2 | X_1 > x_1, X_2 > x_2] \), and \( \sigma^2_2(x_1, x_2) = E[(X_2 - e_2(x_1, x_2))^2 | X_1 > x_1, X_2 > x_2] \). The asymptotic unbiasedness of \( e^*_n \) is a direct consequence of the dominated convergence theorem,
but under the additional assumptions of \( \sigma_1^2(x, y) \) and \( \sigma_2^2(x, y) \) being finite, a bound on the difference of the biases of \( \hat{e}_n \) and \( e_n^* \) can be obtained.

**Theorem 4.2.3.** The estimator defined by (4.2.1) is asymptotically unbiased. That is, \( E[e_n^*(x_1, x_2)] \to e(x_1, x_2) \) as \( n \to \infty \) for all \( (x_1, x_2) \in X \). Moreover, when both \( \sigma_1^2(x_1, x_2) \) and \( \sigma_2^2(x_1, x_2) \) are finite for any \( x_1 \) and \( x_2 \), \( E[\hat{e}_n(x_1, x_2)] - E[e_n^*(x_1, x_2)] \leq \sigma_1^2(x_1, x_2)/\{(n+1)S(x_1, x_2)(m_1(x_1, x_2) + x_1 - e_1(x_1, x_2))\} \) and \( E[\hat{e}_n(x_1, x_2)] - E[e_n^*(x_1, x_2)] \leq \sigma_2^2(x_1, x_2)/\{(n+1)S(x_1, x_2)(m_2(x_1, x_2) + x_2 - e_2(x_1, x_2))\} \) so that the \( \text{Bias}(e_n^*) \to 0 \) exponentially as well.

**Proof.** It follows from the definition of \( e_n^*(x_1, x_2) \) that \( e_n^*(x_1, x_2) \leq m(x_1, x_2) \) with \( m(x_1, x_2) < \infty \) for all \( (x_1, x_2) \in X \). From Theorem 1 of Kulkarni and Rattihali (2002), it follows that \( \hat{e}_n(x_1, x_2) \to e(x_1, x_2) \), and as a consequence, \( e_n^*(x_1, x_2) = \min(\hat{e}_n(x_1, x_2), m(x_1, x_2)) \to e(x_1, x_2) \) almost surely for all \( (x_1, x_2) \in X \). Hence, by the dominated convergence theorem

\[
E[e_n^*(x_1, x_2)] \to e(x_1, x_2) \quad \text{as} \quad n \to \infty
\]

for all \( (x_1, x_2) \in X \). To prove the second part, as in Kulkarni and Rattihali (2002), let \( S_k \) be all subsets of size \( k \) of \( \{1, \ldots, n\} \), and for each \( f_k \in S_k \), define \( A_{f_k} = \{(X_j > x, Y_j > y), j \in f_k; (X_j \leq x \text{ or } Y_j \leq y), j \not\in f_k \} \). Then,

\[
E[\min(\hat{e}_{1n}, m_1)] = \int_0^{m_1} P(\hat{e}_{1n} > t) dt = \sum_{k=1}^n \sum_{f_k \in S_k} \int_0^{m_1} P(\hat{e}_{1n} > t|A_{f_k}) P(A_{f_k}) dt.
\]

Therefore,
\[ E[\hat{e}_{1n}] - E[\min(\hat{e}_{1n}, m_1)] \]
\[ = \sum_{k=1}^{n} \sum_{f_k \in S_k} P(A_{f_k}) \int_{m_1}^{\infty} P(\hat{e}_{1n} > t | A_{f_k}) dt \]
\[ = \sum_{k=1}^{n} \sum_{f_k \in S_k} P(A_{f_k}) \int_{m_1}^{\infty} P(\frac{1}{k} \sum_{j \in f_k} (X_j - x > t | A_{f_k}) dt \]
\[ = \sum_{k=1}^{n} \sum_{f_k \in S_k} P(A_{f_k}) \int_{m_1}^{\infty} \frac{\sigma^2_1(x, y)}{k} \int_{m_1}^{\infty} \frac{dt}{(t + x - e_1(x, y))^2} \]
\[ \leq \sum_{k=1}^{n} \sum_{f_k \in S_k} P(A_{f_k}) \frac{\sigma^2_1(x, y)}{k} \frac{\sigma^2_1(x, y)}{k(m_1(x, y) + x - e_1(x, y))} \]
\[ = \frac{\sigma^2_1(x, y)}{(m_1(x, y) + x - e_1(x, y))} \sum_{k=1}^{n} \frac{\binom{n}{k} S(x, y)^k (1 - S(x, y))^{n-k}}{k} \]
\[ \leq \frac{2\sigma^2_1(x, y)}{(n + 1)(m_1(x, y) + x - e_1(x, y))S(x, y)}. \]

A similar proof works for the case of \( E[\hat{e}_{2n}] - E[e_{2n}^*] \).

Thus, although \( e_{n}^* \) is more negatively biased than \( \hat{e}_n \), the bias of \( e_{n}^* \) also goes to zero. The following result states the strong uniform convergence of \( e_{n}^* \) on closed rectangles.

**Theorem 4.2.4.** The estimator defined by (4.2.1) is strongly uniformly consistent on finite rectangles. That is, for fixed \((b_1, b_2) \in X \) and \( D = [0, b_1] \times [0, b_2] \),

\[ \sup_{(x_1, x_2) \in D} ||e_{n}^*(x_1, x_2) - e(x_1, x_2)|| \to 0 \]

with probability one as \( n \to \infty \).
Proof. Since the consistency of $e^*_n$ is equivalent to the consistency of both marginals, it is sufficient to show the strong uniform consistency for one of the marginals. The proof immediately follows from (4.2.2) and from Theorem 2 of Kulkarni and Rattihalli (2002).

Kulkarni and Rattihalli (2002) showed that the process

$$\hat{Z}_n(x_1, x_2) = \{\sqrt{n} [\hat{e}(x_1, x_2) - e(x_1, x_2)] , (x_1, x_2) \in \mathcal{X}\}$$

converges weakly to a bivariate Gaussian process with mean vector $\mathbf{0}$ and covariance matrix $\Sigma$, where $\Sigma$ is a $2 \times 2$ matrix with its elements defined as follows:

(i) $\Sigma_{(11)}(x_1, x_2) = \frac{[\mu_{11}(x_1, x_2) - S(x_1, x_2)\{e_1(x_1, x_2) + x_1\}^2]}{S^2(x_1, x_2)}$,

(ii) $\Sigma_{(22)}(x_1, x_2) = \frac{[\mu_{22}(x_1, x_2) - S(x_1, x_2)\{e_2(x_1, x_2) + x_2\}^2]}{S^2(x_1, x_2)}$,

(iii) $\Sigma_{(12)}(x_1, x_2) = \Sigma_{(21)}(x_1, x_2) = \mu_{12}(x_1, x_2) - S(x_1, x_2)\{e_1(x_1, x_2) + x_1\}\{e_2(x_1, x_2) + x_2\}$,

where $\mu_{ij}(x_1, x_2) = E[X_i^1X_i^2I(X_1 > x_1, X_2 > x_2)]$.

Since the estimator $e^*_n(x_1, x_2)$ is pointwise closer to $e(x_1, x_2)$ than $\hat{e}_n(x_1, x_2)$ is to $e(x_1, x_2)$ for every $n$, it is expected that $e^*_n$ suitably normalized can also be shown to converge weakly to a Gaussian process. For that purpose, define $Z^*_n$ by

$$Z^*_n(x_1, x_2) = \sqrt{n} |e^*(x_1, x_2) - e(x_1, x_2)| \quad (4.2.3)$$

$$= \sqrt{n} |\hat{e}_n(x_1, x_2) \wedge m(x_1, x_2) - e(x_1, x_2)|$$

$$= \sqrt{n} |\hat{e}_n(x_1, x_2) - e(x_1, x_2)| \wedge \sqrt{n} |m(x_1, x_2) - e(x_1, x_2)|$$
\[ = \hat{Z}_n(x_1, x_2) \land \sqrt{n} [m(x_1, x_2) - e(x_1, x_2)]. \]

The following result establishes the weak convergence of \( Z_n^* \) to a bivariate Gaussian process.

**Theorem 4.2.5.** Let \((X_1, X_2)\) and \((Y_1, Y_2)\) have continuous mean residual lifetime functions given by \(e(x_1, x_2)\) and \(m(x_1, x_2)\) respectively, and let \(\bar{F}(x_1, x_2)\) and \(\bar{G}(x_1, x_2)\) denote their corresponding survival functions. Define \(\mathcal{X} = \{(x, y) : \bar{F}(x, y) > 0 \text{ and } \bar{G}(x, y) > 0\}\) and let \(e(x_1, x_2) \leq m(x_1, x_2)\) where \(m(x_1, x_2)\) is known. Let \(\{Z(x_1, x_2), x_1 > 0, x_2 > 0\}\) denote the bivariate Gaussian process obtained in Kulka-rni and Rattishalli (2002) as the weak limit of the bivariate empirical mean residual life process.

(i) If \(e < m\) on \(\mathcal{X}\), then

\[ Z_n^* \Rightarrow Z \text{ on } \mathcal{X}. \]

(ii) If \(e = m\) on \(\mathcal{X}\), then

\[ Z_n^* \Rightarrow Z \land 0 \text{ on } \mathcal{X}. \]

(iii) If for \(i = 1\) or \(i = 2\), \(e_i(x_0, y_0) = m_i(x_0, y_0)\) for some \((x_0, y_0) \in \mathcal{X}\) and \(e_i(x, y) < m_i(x, y)\) for all \((x, y)\) in the line segment \(\alpha(x_0, y_0) + (1 - \alpha)(x_1, y_1), 0 < \alpha < 1\) for some \((x_1, y_1) \in \mathcal{X}\), then \(\{Z_n^*(x, y), (x, y) \in \mathcal{X}\}\) does not converge weakly.

**Proof.** (i) Define \(Z_n(x_1, x_2) = \sqrt{n}(\hat{e}_n(x_1, x_2) - e(x_1, x_2))\) and consider

\[ Z_n^*(x_1, x_2) = \sqrt{n}(e^*(x_1, x_2) - e(x_1, x_2)) \quad (4.2.4) \]

\[ = \sqrt{n}(\hat{e}_n(x_1, x_2) - e(x_1, x_2)) \land \sqrt{n}(m(x_1, x_2) - e(x_1, x_2)). \]
Suppose first that \( \inf_{(x_1, x_2) \in \mathcal{X}} (m(x_1, x_2) - e(x_1, x_2)) > 0 \). Then, \( \sqrt{n}(m_i(x_1, x_2) - e_i(x_1, x_2)) \) converges uniformly to \( \infty \) for \( i = 1, 2 \) and, since \( \sup_{(x_1, x_2) \in \mathcal{X}} \{ \sqrt{n}(\hat{e}_{in}(x_1, x_2) - e_i(x_1, x_2)) \} = O_p, i = 1, 2, P(\sup_{(x_1, x_2) \in \mathcal{X}} \| Z_n^*(x_1, x_2) - Z_n(x_1, x_2) \| > \epsilon) \to 0 \) as \( n \to \infty \), and hence \( \{ Z_n^*(x_1, x_2), (x_1, x_2) \in \mathcal{X} \} \Rightarrow \{ Z(x_1, x_2), (x_1, x_2) \in \mathcal{X} \} \).

Now consider the case when \( e < m \) but \( \inf_{(x_1, x_2) \in \mathcal{X}} (m_i(x_1, x_2) - e_i(x_1, x_2)) = 0 \) for \( i = 1 \) or \( i = 2 \) or both. The proof hinges on the same idea as in the case that \( \inf_{(x_1, x_2) \in \mathcal{X}} (m(x_1, x_2) - e(x_1, x_2)) > 0 \) except that we apply it to an increasing sequence of closed and bounded \( T_i \), with \( T_i \uparrow \mathcal{X} \). Consider a sequence \( \alpha_n \downarrow 0 \), and define \( T_i = \{(x_1, x_2) \in \mathcal{X} : \hat{F}(x_1, x_2) \geq \alpha_i \} i = 1, 2, \cdots \). Since the \( T_i \)'s are closed and bounded \( \inf_{(x_1, x_2) \in T_i} (m(x_1, x_2) - e(x_1, x_2)) > 0 \). By the previous arguments \( Z_n^*|_{T_i} \Rightarrow Z|_{T_i} \) for each \( i = 1, 2, \cdots \) where \( Z_n^*|_{T_i} \) denotes the restriction of \( Z_n^* \) to \( T_i \). It follows that \( Z_n^* \Rightarrow Z \).

(ii) If \( e(x_1, x_2) = m(x_1, x_2) \) for all \( (x_1, x_2) \in \mathcal{X} \), \( \sqrt{n}(m_j(x_1, x_2) - e_j(x_1, x_2)) = 0 \) on \( \mathcal{X} \) for \( j = 1, 2 \). Hence, by the continuous mapping theorem for weak convergence \( Z_n^* \Rightarrow Z \wedge 0 \).

(iii) If \( Z_n^* \Rightarrow Z \), then by the continuous mapping theorem, the projection mappings \( e_{i,n}^*(x, y) = \sqrt{n}(\hat{e}_{i,n}(x, y) \wedge m_i(x, y) - e_i(x, y)) \) for \( i = 1, 2, \) must converge weakly to the projection mappings of \( Z \). It is now shown that under the conditions of the theorem under (iii), this is not possible. Without loss of generality, suppose that \( e_1(x_0, y_0) = m_1(x_0, y_0) \) for some \( (x_0, y_0) \in \mathcal{X} \), and \( e_1(x, y) < m_1(x, y) \) for all \( (x, y) \in (x_0, x_0 + \delta_1) \times (y_0, y_0 + \delta_2) \) where \( \delta_1 \geq 0, \delta_2 \geq 0 \) and \( \delta_1 \vee \delta_2 > 0 \). We show
that \( e_{1,n}^* \) is not tight on \([x_0, x_0 + \delta_1] \times [y_0, y_0 + \delta_2]\) and hence cannot converge weakly, which in turn implies that \( e_{1,n}^* \) cannot converge weakly. The proof follows Rojo (1995).

Suppose then that \( \delta_1 > 0 \), and for small \( \gamma > 0 \) consider

\[
\sup_{(x_0, y_0) \leq (s_1, s_2) \leq (x_0 + \gamma, y_0)} |e_{1,n}^*(s_1, y_0) - e_{1,n}^*(x_0, y_0)|
\]

\[
= \| e_{1,n}^*(s_1, y_0) - e_{1,n}^*(x_0, y_0) \|_{(x_0, y_0)}^{(x_0 + \gamma, y_0)}
\]

\[
= \sqrt{n} |\hat{e}_{1,n}(s_1, y_0) \wedge m_1(s_1, y_0) - e_1(s_1, y_0) |
\]

\[
- (\hat{e}_{1,n}(x_0, y_0) - e_1(x_0, y_0)) \wedge 0 |_{(x_0, y_0)}^{(x_0 + \gamma, y_0)}.
\]

Now, for \( s_0 = x_0 + \frac{\min(\delta_1, \gamma)}{2} \), eventually with probability one, \( m_1(s_0, y_0) > \hat{e}_{1,n}(s_0, y_0) \).

Therefore, eventually with probability one,

\[
\sqrt{n} |e_{1,n}^*(s_1, y_0) - e_{1,n}^*(x_0, y_0)|_{(x_0, y_0)}^{(x_0 + \gamma, y_0)} \geq \sqrt{n} |\hat{e}_{1,n}(s_0, y_0) - e_1(s_0, y_0) |
\]

\[
- (\hat{e}_{1,n}(x_0, y_0) - e_1(x_0, y_0)) \wedge 0 |
\]

\[
= \sqrt{n} |\max(0, e_1(x_0, y_0) - \hat{e}_{1,n}(x_0, y_0)) |
\]

\[
+ (\hat{e}_{1,n}(s_0, y_0) - e_1(s_0, y_0)) |.
\]

Therefore,

\[
\lim_{n \to \infty} P \left\{ \sqrt{n} |e_{1,n}^*(s_1, y_0) - e_{1,n}^*(x_0, y_0)|_{(x_0, y_0)}^{(x_0 + \gamma, y_0)} \geq \epsilon \right\}
\]

\[
\geq \lim_{n \to \infty} P \left\{ \sqrt{n} |\max(0, e_1(x_0, y_0) - \hat{e}_{1,n}(x_0, y_0)) |
\right.
\]

\[
+ \hat{e}_{1,n}(s_0, y_0) - e_1(s_0, y_0) | \geq \epsilon \}
\]

\[
\geq \lim_{n \to \infty} P \left\{ |\max(0, \sqrt{n}|e_1(x_0, y_0) - \hat{e}_{1,n}(x_0, y_0)) | \right.
\]

\[
+ \hat{e}_{1,n}(s_0, y_0) - e_1(s_0, y_0) | \geq \epsilon \}
\]
\[
\begin{align*}
+ \sqrt{n}[\bar{e}_{1,n}(s_0, y_0) - e_1(s_0, y_0)] \geq \epsilon \\
\geq \lim_{n \to \infty} P \left\{ \sqrt{n}[\bar{e}_{1,n}(s_0, y_0) - e_1(s_0, y_0)] \geq \epsilon \right\} = 1 - \Phi\left( \frac{\epsilon}{\sigma^2} \right)
\end{align*}
\]
where \( \sigma^2 = \text{Var}(Z(s_0, y_0)) \), and \( \Phi \) denotes the standard normal distribution. It follows that \( \{e_n^*(x_1, x_2), (x_1, x_2) \in \mathcal{D}\} \) is not tight. \( \square \)

Note that part (ii) of the above theorem provides the tools for testing the null hypothesis that \( e = m \) against the alternative that \( e < m \). To test this hypothesis, let

\[
D_n^- = \max \left\{ \sup_{(x_1, x_2)} [m_1(x_1, x_2) - e_{1,n}^*(x_1, x_2)], \sup_{(x_1, x_2)} [m_2(x_1, x_2) - e_{2,n}^*(x_1, x_2)] \right\}
\]

and reject the hypothesis that \( e = m \) if \( D_n^- > K_\alpha \) where \( K_\alpha \) is the \( (1 - \alpha) \times 100 \) quantile of the distribution of \( \max \{\sup_{\mathcal{D}} \min(Z_1, 0), \sup_{\mathcal{D}} \min(Z_2, 0)\} \).

4.3 The Two Sample Problem

4.3.1 The estimator under mrl ordering

In this section we consider the problem of nonparametric estimation of a bivariate mrl function \( e \) when it is bounded above by an unknown mrl function \( m \). Let \((X_{1i}, X_{2i}), i = 1, \ldots, n_1\) and \((Y_{1j}, Y_{2j}), j = 1, \ldots, n_2\), be random vectors with finite means representing the lifetimes of two populations with distribution functions \( F_1(x, y) \) and \( F_2(x, y) \); survival functions \( S^e(x, y) \) and \( S^m(x, y) \); and mrl functions \( e_1(x, y) \) and \( e_2(x, y) \) respectively. Given \((X_{1i}, X_{2i}), i = 1, \ldots, n_1\) and \((Y_{1j}, Y_{2j}), j = 1, \ldots, n_2\)
it is of interest to estimate $e$ or $m$, or both, subject to the restriction that $e \leq m$.

Consider the empirical survival functions $S_{n_1}^e$ and $S_{n_2}^m$. One possible approach is to define the “pooled” empirical survival function $\hat{S}^* = (n_1 S_{n_1}^e + n_2 S_{n_2}^m)/(n_1 + n_2)$ and obtain the corresponding mean residual life function defined by $\hat{e}^* = (\hat{e}_1^*, \hat{e}_2^*)$ where

$$\hat{e}_1^*(x, y) = \frac{\int_y^\infty \left[ n_1 S_{n_1}^e(u, y) + n_2 S_{n_2}^m(u, y) \right] du}{n_1 S_{n_1}^e(x, y) + n_2 S_{n_2}^m(x, y)} \quad (4.3.1)$$

$$= w_1(x, y) \hat{e}_1(x, y) + w_2(x, y) \hat{m}_1(x, y)$$

and similarly,

$$\hat{e}_2^*(x, y) = \frac{\int_x^\infty \left[ n_1 S_{n_1}^e(x, v) + n_2 S_{n_2}^m(x, v) \right] dv}{n_1 S_{n_1}^e(x, y) + n_2 S_{n_2}^m(x, y)} \quad (4.3.2)$$

$$= w_1(x, y) \hat{e}_2(x, y) + w_2(x, y) \hat{m}_2(x, y)$$

where $\hat{e} = (\hat{e}_1(x, y), \hat{e}_2(x, y))$ and $\hat{m} = (\hat{m}_1(x, y), \hat{m}_2(x, y))$ are the empirical mean residual life functions corresponding to $S_{n_1}^e$ and $S_{n_2}^m$ respectively, and

$$w_1(x, y) = \frac{n_1 S_{n_1}^e(x, y)}{n_1 S_{n_1}^e(x, y) + n_2 S_{n_2}^m(x, y)} \quad (4.3.3)$$

$$w_2(x, y) = \frac{n_2 S_{n_2}^m(x, y)}{n_1 S_{n_1}^e(x, y) + n_2 S_{n_2}^m(x, y)}.$$

Unfortunately, although this approach still provides better estimators than the empirical mrl function in terms of mean squared error, by choosing a different set of weights $w_i, i = 1, 2$, one can improve on the resulting estimators derived from (4.3.1)–(4.3.3).

Note that, alternatively, we could define our estimators for $e$ and $m$ as follows:

$$e^*(x, y) = \min(\hat{e}(x, y), \hat{m}(x, y)) \quad (4.3.4)$$

$$m^*(x, y) = \max(\hat{e}(x, y), \hat{m}(x, y)).$$
One drawback of the estimators defined by (4.3.4) is that although they are strongly uniformly consistent on closed and bounded rectangles when both \( n_1, n_2 \to \infty \), \( e^* \) fails to be consistent if \( n_2 \to \infty \) and \( n_1 \) remains bounded. (A similar phenomenon has been observed by Rojo (2004) in the estimation of stochastically ordered distribution functions).

To circumvent the above problems, we propose estimators similar to those suggested by (4.3.1), (4.3.2), and (4.3.3), except that we pool the empirical mean residual lifetime functions using weights \( w_i = \frac{n_i}{n_1 + n_2} \quad i = 1, 2 \), so that our "benchmark" mrl function is now defined by the pooled estimate

\[
\hat{e}_p(x, y) = w_1 \hat{e}(x, y) + w_2 \hat{m}(x, y)
\]

and our estimators of \( e \) and \( m \) are given respectively by

\[
\hat{e}^*(x, y) = \min(\hat{e}_p(x, y), \hat{e}(x, y))
\]

\[
\hat{m}^*(x, y) = \max(\hat{e}_p(x, y), \hat{m}(x, y)).
\]

These estimators have the property, as will be demonstrated, that \( \hat{e}^*(x, y) \) converges to \( e(x, y) \) almost surely and uniformly on closed bounded rectangles when \( n_1 \to \infty \); similarly, \( \hat{m}^*(x, y) \) is strongly uniformly consistent on closed bounded rectangles when \( n_2 \to \infty \). In addition, because of the simplicity of the weights \( w_i, i = 1, 2 \), the finite sample and asymptotic theory of these estimators can be obtained. Moreover, results from simulation work suggest that the estimators defined by (4.3.1)–(4.3.3) are easily dominated in mean squared error by the estimators (4.3.5) and (4.3.6). For these
reasons, we focus our attention on the estimators \( \hat{\epsilon}^* \) and \( \hat{m}^* \) defined by (4.3.5) and (4.3.6).

It follows immediately from (4.3.5) and (4.3.6) that \( \hat{\epsilon}^* \leq \hat{\epsilon}_p \leq \hat{m}^* \) and, for \( i = 1, 2 \),

\[
\hat{\epsilon}^*_i(x, y) = \hat{\epsilon}_i(x, y)I_{\{\hat{\epsilon}_i(x, y) \leq \hat{m}_i(x, y)\}} + \hat{\epsilon}_{ip}(x, y)I_{\{\hat{\epsilon}_i(x, y) > \hat{m}_i(x, y)\}} \tag{4.3.7}
\]

\[
\hat{m}^*_i(x, y) = \hat{m}_i(x, y)I_{\{\hat{m}_i(x, y) \geq \hat{\epsilon}_i(x, y)\}} + \hat{\epsilon}_{ip}(x, y)I_{\{\hat{m}_i(x, y) < \hat{\epsilon}_i(x, y)\}}. \tag{4.3.8}
\]

This representation leads immediately to the study of the finite-sample and asymptotic properties of our estimators.

The following two theorems show that \( \hat{\epsilon}^*(x, y) \) is the projection of the empirical mrl onto the convex set of mean residual life functions \( k(x, y) \) such that \( k(x, y) \leq \hat{\epsilon}_p(x, y) \) and \( \hat{m}^*(x, y) \) is the projection of the empirical mrl onto the convex set of mean residual life functions \( k(x, y) \) such that \( k(x, y) \geq \hat{\epsilon}_p(x, y) \). These results parallel the results obtained in the one-sample case.

**Theorem 4.3.1.** Let \( A \) be the set of all mrl functions bounded above by the mrl function \( \hat{\epsilon}_p(x_1, x_2) \). That is, let \( A = \{k(x_1, x_2) : k(x_1, x_2) \leq \hat{\epsilon}_p(x_1, x_2) \text{ for all } (x_1, x_2) \in \mathcal{X}\} \). Let \( \hat{\epsilon}^*(x_1, x_2) = \hat{\epsilon}_p(x_1, x_2) \land \hat{\epsilon}(x_1, x_2) \). Then,

(i) For any \( k \in A \),

\[
\sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|\hat{\epsilon}^*(x_1, x_2) - \hat{\epsilon}(x_1, x_2)\| \leq \sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|k(x_1, x_2) - \hat{\epsilon}(x_1, x_2)\|
\]

(ii) For any \( k \in A \) and for \( p \geq 1 \),

\[
\|\hat{\epsilon}^* - \hat{\epsilon}\|_p \leq \|k - \hat{\epsilon}\|_p.
\]
Proof. Only (i) will be proven. The proof of (ii) follows immediately from the arguments used in (i). Let \( \hat{e}^* = (\hat{e}^*_1, \hat{e}^*_2) \) and \( \hat{e}^*_p = (\hat{e}^*_1, \hat{e}^*_2) \). The result follows immediately if it can be shown that \( |\hat{e}^*_i(x_1, x_2) - \hat{e}_i(x_1, x_2)| \leq |\hat{e}^*_p(x_1, x_2) - \hat{e}_i(x_1, x_2)| \) for \( i = 1, 2 \).

Now, from equation (4.3.4) we know that, for \( i = 1, 2 \),

\[
\hat{e}^*_i(x_1, x_2) - \hat{e}_i(x_1, x_2) = \begin{cases} 
\hat{e}_p(x_1, x_2) - \hat{e}_i(x_1, x_2) & \text{if } \hat{e}^*_i(x_1, x_2) \geq \hat{m}^*_i(x_1, x_2) \\
0 & \text{otherwise.}
\end{cases}
\]

But when \( \hat{e}_i(x_1, x_2) \geq \hat{m}^*_i(x_1, x_2) \), \( \hat{m}_i(x_1, x_2) \leq \hat{e}_p(x_1, x_2) \leq \hat{e}^*_i(x_1, x_2) \). Since \( k(x_1, x_2) \leq \hat{e}_p(x_1, x_2) \), this implies that for \( i = 1, 2 \),

\[
|\hat{e}^*_i(x_1, x_2) - \hat{e}_i(x_1, x_2)| = |\hat{e}_p(x_1, x_2) - \hat{e}_i(x_1, x_2)| \leq |k_i(x_1, x_2) - \hat{e}_i(x_1, x_2)|.
\]

Hence,

\[
\sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|\hat{e}^*(x_1, x_2) - \hat{e}(x_1, x_2)\| \leq \sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|k(x_1, x_2) - \hat{e}(x_1, x_2)\|
\]

for all \( k(x_1, x_2) \in A \).

A similar results hold for the estimator \( \hat{m}^* \) as given in the following theorem.

**Theorem 4.3.2.** Let \( B \) be the set of all mrl functions bounded below by the mrl function \( \hat{e}_p(x_1, x_2) \). That is, let \( B = \{h(x_1, x_2) : h(x_1, x_2) \geq \hat{e}_p(x_1, x_2) \text{ for all } (x_1, x_2) \in \mathcal{X}\} \). Let \( \hat{m}^*(x_1, x_2) = \hat{m}(x_1, x_2) \lor \hat{e}_p(x_1, x_2) \). Then

(i) For any \( h \in B \),

\[
\sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|\hat{m}^*(x_1, x_2) - \hat{m}(x_1, x_2)\| \leq \sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|h(x_1, x_2) - \hat{m}(x_1, x_2)\|.
\]

(ii) For any \( h \in B \) and for \( p \geq 1 \),

\[
\|\hat{m}^* - \hat{m}\|_p \leq \|h - \hat{m}\|_p.
\]
Proof. The proof is the same as the previous proof. □

4.3.2 Asymptotic properties of the estimators

We now turn our attention to the asymptotic properties of \( \hat{e}^* \) and \( \hat{m}^* \). Asymptotic unbiasedness and normality as well as strong uniform consistency of \( \hat{e}^* \) and \( \hat{m}^* \) are established in the following results.

Theorem 4.3.3. The estimators \( \hat{e}^* \) and \( \hat{m}^* \) are asymptotically unbiased. That is,

\[
E[\hat{e}^*(x_1, x_2)] \rightarrow e(x_1, x_2) \text{ as } n_1 \rightarrow \infty \text{ and }
E[\hat{m}^*(x_1, x_2)] \rightarrow m(x_1, x_2) \text{ as } n_2 \rightarrow \infty \text{ for all } (x_1, x_2) \in \mathcal{X}.
\]

Proof. The proof proceeds by showing uniform integrability, and almost sure convergence to zero, of an appropriate sequence of statistics. Consider,

\[
E[\hat{e}^*_1(x, y)] = E \left[ \min \left( \hat{e}_1(x, y), \frac{n_1}{n_1 + n_2} \hat{e}_1(x, y) + \frac{n_2}{n_1 + n_2} \hat{m}_1(x, y) \right) \right]
\]

\[
= E[\hat{e}_1(x, y)] + \frac{n_2}{n_1 + n_2} E \left[ \min \left( 0, \hat{m}_1(x, y) - \hat{e}_1(x, y) \right) \right]. \tag{4.3.9}
\]

If \( n_1 \rightarrow \infty \) while \( n_2 \) remains bounded, \( E[\hat{e}_1(x, y)] \rightarrow e_1(x, y) \) while the last term in (4.3.9) converges to zero. On the other hand, if both \( n_1 \) and \( n_2 \rightarrow \infty \), since \( \hat{m}_1(x, y) - \hat{e}_1(x, y) \) converges to \( m_1(x, y) - e_1(x, y) \geq 0 \) with probability one, and hence \( \min(0, \hat{m}_1(x, y) - \hat{e}_1(x, y)) \) converges to zero with probability one, it is enough to show that the sequence \( \min(0, \hat{m}_1 - \hat{e}_1) \) is uniformly integrable. This follows easily after writing

\[
E \left[ \min(0, \hat{m}_1(x, y) - \hat{e}_1(x, y)) | I_{\min(0, \hat{m}_1 - \hat{e}_1) > 1} \right]
\]
\[
E \left[ \max(0, \hat{e}_1 - \hat{m}_1) I_{\{\max(0, \hat{e}_1 - \hat{m}_1) > 1\}} \right]
\]

and noticing that \( \hat{e}_1 - \hat{m}_1 \) converges almost surely to \( e_1 - m_1 \leq 0 \). Therefore, eventually, with probability one, \( I_{\{\max(0, \hat{e}_1 - \hat{m}_1) > 1\}} = 0 \) and the result follows.

The strong uniform convergence of both \( \hat{e}^* \) and \( \hat{m}^* \) on closed bounded rectangles follows almost immediately from the representations (4.3.7) and (4.3.8) and the fact that \( \hat{e} \) and \( \hat{m} \) converge strongly and uniformly to \( e \) and \( m \), respectively, on closed bounded rectangles. To establish the strong uniform convergence, let \((b_1, b_2) \in \mathcal{X}\) be fixed, and let \(D = [0, b_1] \times [0, b_2]\).

**Theorem 4.3.4.** The estimators \( \hat{e}^* \) and \( \hat{m}^* \) are uniformly strongly consistent on \( D \).

That is

\[
\sup_{(x_1, x_2) \in D} \| \hat{e}^*(x_1, x_2) - e(x_1, x_2) \| \to 0 \text{ as } n_1 \to \infty \text{ and }
\sup_{(x_1, x_2) \in D} \| \hat{m}^*(x_1, x_2) - m(x_1, x_2) \| \to 0 \text{ as } n_2 \to \infty.
\]

**Proof.** Since the strong uniform consistency of the bivariate mrl vector is equivalent to the strong uniform consistency of both marginals, it is sufficient to prove the theorem for the marginals. Here, we will show the proof for the first component of the mrl vector, the proof for the second component is exactly the same. Let \( D = [0, b_1] \times [0, b_2]\) where \( b_1, b_2 > 0 \), and consider

\[
\sup_{(x, y) \in D} | \hat{e}^*_1(x, y) - e_1(x, y) | = \sup_{(x, y) \in D} | \min(\hat{e}_1(x, y), \hat{e}_{1p}(x, y)) - e_1(x, y) |
\]

\[
= \sup_{(x, y) \in D} \left| \min(\hat{e}_1(x, y) - e_1(x, y), \frac{n_1}{n_1 + n_2}(\hat{e}_1(x, y)) \right|
\]

\[
= \sup_{(x, y) \in D} \left| \frac{n_1}{n_1 + n_2}(\hat{e}_1(x, y)) \right|
\]
Now let $n_1 \to \infty$ and $\frac{n_2}{n_1 + n_2} \to \alpha$, $0 \leq \alpha \leq 1$. Since, $\hat{e}$ and $\hat{m}$ are strongly and uniformly consistent on $D$ for $e$ and $m$ respectively, it follows that $\min \left( \hat{e}_1(x,y) - e_1(x,y) \right)$ converges, uniformly on $D$, with probability one to $\min(0, \alpha(m_1(x,y) - e_1(x,y))) = 0$, and the result follows. A similar proof yields the result for $\hat{m}^*$.

Shifting attention to the asymptotic behavior of the mrl processes, define the bivariate mrl processes $Z^*_e$ and $Z^*_m$ by

$$
Z^*_e(x_1, x_2) = \sqrt{n_1} \left[ \hat{e}^*(x_1, x_2) - e(x_1, x_2) \right]
$$

$$
= \sqrt{n_1} \left[ \hat{e}(x_1, x_2) \wedge \hat{e}_p(x_1, x_2) - e(x_1, x_2) \right] \quad \text{and} \quad (4.3.10)
$$

$$
Z^*_m(x_1, x_2) = \sqrt{n_2} \left[ \hat{m}^*(x_1, x_2) - m(x_1, x_2) \right]
$$

$$
= \sqrt{n_2} \left[ \hat{m}(x_1, x_2) \vee \hat{e}_p(x_1, x_2) - m(x_1, x_2) \right]. \quad (4.3.11)
$$

Similar to the one-sample case, it is possible to study the weak convergence of the processes defined by (4.3.10) and (4.3.11).

The following theorem provides the asymptotic theory of the mean residual life processes $Z^*_e$ and $Z^*_m$. In what follows, let $C_1(x, y)$ and $C_2(x, y)$ represent the covariance functions obtained by plugging into (11a)-(11e) of Kulkarni and Rattihali (2002), the quantities corresponding to the mean residual life functions $e$ and $m$. Note that both $C_1(x, y)$ and $C_2(x, y)$ are $2 \times 2$ matrices and their elements are given by:

$$
(1) \quad C_1(x, y)_{(11)} = \frac{[\omega_0(x, y) - S^*(x, y) \cdot e_1(x, y) + \omega_1(x, y)]^2}{(S^*(x, y))^2},
$$
(2) \( C_1(x, y)_{(22)} = \frac{[\mu_{22}(x, y) - S^e(x, y)\{e_2(x, y) + y\}]^2}{(S^e(x, y))^2} \),

(3) \( C_1(x, y)_{(12)} = C_1(x, y)_{(21)} = \mu_{11}(x, y) - S^e(x, y)[\{e_1(x, y) + x\}{e_2(x, y) + y}] \),

(4) \( C_2(x, y)_{(11)} = \frac{[\mu_{11}(x, y) - S^m(x, y)\{m_1(x, y) + x\}]^2}{(S^m(x, y))^2} \),

(5) \( C_2(x, y)_{(22)} = \frac{[\mu_{22}(x, y) - S^m(x, y)\{m_2(x, y) + y\}]^2}{(S^m(x, y))^2} \),

(6) \( C_2(x, y)_{(12)} = C_2(x, y)_{(21)} = \mu_{11}(x, y) - S^m(x, y)[\{m_1(x, y) + x\}{m_2(x, y) + y}] \),

where \( \mu_{ij}(x, y) = E[X_1^iX_2^j I(X_1 > x, X_2 > y)] \) and \( \mu'_{ij}(x, y) = E[Y_1^iY_2^j I(Y_1 > x, Y_2 > y)] \).

**Theorem 4.3.5.** Let \( S^e(x, y) \) and \( S^m(x, y) \) be continuous survival functions and let \( \mathcal{X} = \{(x, y) : S^e(x, y) > 0 \text{ and } S^m(x, y) > 0 \} \) denote bivariate mean zero Gaussian processes with covariance functions \( C_1(x, y) \) and \( C_2(x, y) \). Suppose that \( \frac{n_2}{n_1} \to \alpha \) with \( 0 \leq \alpha \leq \infty \) and \( e(x, y) \leq m(x, y) \) for all \( (x, y) \in \mathcal{X} \) where both \( e \) and \( m \) are unknown.

(i) Suppose that \( e \leq m \) on \( \mathcal{X} \). Then

- if \( n_1 \to \infty \) with \( \alpha = 0 \), \( Z^*_e \Rightarrow Z_e \); if \( n_1 \to \infty \) with \( \alpha = \infty \), \( Z^*_e \Rightarrow Z_e^- \).

- if \( n_2 \to \infty \) with \( \alpha = \infty \), \( Z^*_m \Rightarrow Z_m \); if \( n_2 \to \infty \) with \( \alpha = 0 \), \( Z^*_m \Rightarrow Z_m^+ \),

where \( f^+ = \max(0, f) \), \( f^- = \min(0, f) \).

(ii) If \( e < m \) on \( \mathcal{X} \), then as \( n_1 \to \infty \) (\( n_2 \to \infty \))

\[ Z^*_e \Rightarrow Z_e \ (Z^*_m \Rightarrow Z_m) \]
(iii) If \( e = m \) on \( \mathcal{X} \) with \( 0 < \alpha < \infty \), then

\[
Z_e^* \Rightarrow \min(Z_e, \frac{\sqrt{\alpha}}{1+\alpha} Z_m + \frac{1}{1+\alpha} Z_e) \quad \text{on} \quad \mathcal{X}
\]

\[
Z_e^m \Rightarrow \max(Z_m, \frac{\sqrt{\alpha}}{1+\alpha} Z_e + \frac{\alpha}{1+\alpha} Z_m) \quad \text{on} \quad \mathcal{X}.
\]

(iv) If for \( i = 1 \) or \( i = 2 \), \( e_i(x_0, y_0) = m_i(x_0, y_0) \) for some \( (x_0, y_0) \in \mathcal{X} \) and \( e_i(x, y) < m_i(x, y) \) for all \( (x, y) \) in some line segment \( \{(x, y) : (x, y) = \alpha(x_0, y_0) + (1 - \alpha)(x_1, y_1), 0 < \alpha < 1\} \) for some \( (x_1, y_1) \neq (x_0, y_0) \), then \( Z_e^* \) and \( Z_e^m \) do not converge weakly.

**Proof.** (i) Note that

\[
Z_e^* = \sqrt{n_1} \min \left( 0, \frac{n_2}{n_1 + n_2} \{ (\hat{m} - m) + (m - e) + (\hat{e} - e) \} \right)
\]

\[
+ \sqrt{n_1(\hat{e} - e)}.
\]

Consider first the case with \( \alpha = 0 \). Since \( \sup_{(x,y)\in\mathcal{X}} \sqrt{n_2} \| \hat{m} - m \| = O_p, \sqrt{n_1} \| \hat{e} - e \| = O_p, \) and \( (m - e) \geq 0 \), it follows that the first term in (4.3.12) converges uniformly with probability one to \( 0 \). Therefore, \( Z_e^* \Rightarrow Z_e \). If \( \alpha = \infty \), a similar analysis yields the result that \( Z_e^* \Rightarrow \min(0, -Z_e) = Z_e^- + Z_e^+ \). The results for the process \( Z_e^m \) are obtained in a similar fashion.

(ii) Let \( K_{n_1, n_2}(x, y) = \sqrt{n_1(\frac{n_1}{n_1 + n_2})} \min(0, (\hat{m} - m) + (m - e) + (\hat{e} - e)) \). Note that because of (i), it is enough to consider the case where \( 0 < \alpha < \infty \). The proof exploits the expression for \( K_{n_1, n_2}(x, y) \). Suppose first that \( \inf_{(x,y)\in\mathcal{X}} (m(x, y) - e(x, y)) > 0 \) so that \( \sqrt{n_1} \sqrt{n_2}(m - e) \) converges to \( \infty \) uniformly on \( \mathcal{X} \). The result follows immediately after observing that, for \( 0 < \alpha < \infty \), \( \sup_{(x,y)\in\mathcal{X}} \sqrt{n_1(\frac{n_1}{n_1 + n_2})} \| \sqrt{n_2}(\hat{m} - m) \| = \)
$O_p$ and \( \sup_{(x,y) \in \mathcal{X}} \left( \frac{n_2}{n_1 + n_2} \right) \| \sqrt{n_1} (\hat{e} - e) \| = O_p \), and therefore, with probability one, \( \sup_{(x,y) \in \mathcal{X}} \| K_{n_1,n_2}(x,y) \| \to 0 \). It follows that \( Z^*_e \Rightarrow Z_e \).

If, on the other hand, \( \inf_{(x,y) \in \mathcal{X}} (m(x,y) - e(x,y)) = 0 \), we proceed as in the proof of Theorem 2.5. Choose \( \beta_k \downarrow 0 \) and define \( T_k = \{ (x,y) \in \mathcal{X} : S(x,y)^{\star} \geq \beta_k \text{ and } S(x,y)^m \geq \beta_k \} \) so that \( T_k \uparrow \mathcal{X} \) and \( \inf_{(x,y) \in T_k} (m(x,y) - e(x,y)) > 0 \). Arguments similar to those used above yield the result that \( Z^*_e |_{T_k} \Rightarrow Z_e |_{T_k} \). The weak convergence of \( Z^*_e \) to \( Z_e \) follows. Similar arguments show that \( Z^*_m \Rightarrow Z_m \).

(iii) Only the proof for \( Z^*_e \) is given here, the proof for \( Z^*_m \) following from similar arguments. Consider the independent bivariate processes \( \left\{ \hat{W}_1(x,y), (x,y) \in \mathcal{X} \right\} = \left\{ \sqrt{n_1} (\hat{e}(x,y) - e(x,y)), (x,y) \in \mathcal{X} \right\} \) and \( \left\{ \hat{W}_2(x,y), (x,y) \in \mathcal{X} \right\} = \left\{ \sqrt{n_2} (\hat{m}(x,y) - m(x,y)), (x,y) \in \mathcal{X} \right\} \); each process converging weakly to its appropriate Gaussian processes limit \( W_i \). It follows that \( \left\{ \hat{W}_1(x,y), \hat{W}_2(x,y); (x,y) \in \mathcal{X} \right\} \Rightarrow \{ W_1(x,y), W_2(x,y); (x,y) \in \mathcal{X} \} \). Since the map \( \min(f,g) \rightarrow \min(f,af + bg) \) is continuous, it follows from the continuous mapping theorem that \( Z^*_e \Rightarrow \min \left( Z_e, \frac{\sqrt{a}}{1+a} Z_m + \frac{1}{1+a} Z_e \right) \).

Similar arguments yield the result for \( Z^*_m \).

(iv) Without loss of generality we suppose that \( e(x_0, y_0) = m(x_0, y_0) \) for some \((x_0, y_0)\) in \( \mathcal{X} \) and \( e(x_0, y) < m(x_0, y) \) for all \( y \in (y_0, y_1) \). Since weak convergence of \( Z^*_e \) implies weak convergence of the projection mappings, it is enough to show that the process \( \left\{ \sqrt{n_1} (\hat{e}_1(x,y) - e_1(x,y)), (x,y) \in \mathcal{X} \right\} \) cannot converge weakly. But this follows immediately from arguments similar to those used in the proof of Theorem 4.2.5. \( \square \)
4.4  The Mirror Image of the One Sample Problem

4.4.1  The estimator under mrl ordering

In this section we will consider the "mirror image" of the problem discussed in Section 4.2. Here we will assume that the mrl function is bounded from below by another known mrl function. Suppose that \( \mathbf{X} = (X_1, X_2) \) and \( \mathbf{Y} = (Y_1, Y_2) \) are random vectors with finite means representing lifetimes of two populations with distribution functions \( F(x_1, x_2) \) and \( G(y_1, y_2) \), survival functions \( \overline{F}(x_1, x_2) \) and \( \overline{G}(y_1, y_2) \), and mrl functions \( e(x_1, x_2) \) and \( m(y_1, y_2) \) respectively. Let \( \mathcal{X} = \{ (x_1, x_2) : \overline{F}(x_1, x_2) > 0 \text{ and } \overline{G}(x_1, x_2) > 0 \} \). Moreover, let us assume that \( e(x_1, x_2) \geq m(x_1, x_2) \) for all \( (x_1, x_2) \in \mathcal{X} \) where \( m(x_1, x_2) \) is a known mean residual lifetime function. The estimator of the mrl function under this ordered restriction is given by:

\[
\hat{e}^{*}(x_1, x_2) = \hat{e}(x_1, x_2) \vee m(x_1, x_2) \\
= \begin{cases} 
\frac{1}{2} \left[ \hat{e}_1(x_1, x_2) + m_1(x_1, x_2) + |\hat{e}_1(x_1, x_2) - m_1(x_1, x_2)| \right], \\
\frac{1}{2} \left[ \hat{e}_2(x_1, x_2) + m_2(x_1, x_2) + |\hat{e}_2(x_1, x_2) - m_2(x_1, x_2)| \right] 
\end{cases}
\]

where \( \hat{e}(x_1, x_2) \) is the bivariate empirical mrl function. It can easily be seen that \( \hat{e}^{*} \) is less negatively biased than \( \hat{e} \). Moreover, \( \hat{e}^{*} \) has uniformly smaller mean squared error than \( \hat{e} \) as proved in the next result.

**Theorem 4.4.1.** For every \( (x_1, x_2) \in \mathcal{X} \), the estimator \( \hat{e}^{*}(x_1, x_2) \) has smaller mean squared error than the empirical mrl estimator \( \hat{e}(x_1, x_2) \).
Proof. The proof of Theorem 4.4.1 is similar to the proof of Theorem 4.2.1. Let

\[
Q_1 = \{(x_1, x_2) : \hat{e}_{1,n}(x_1, x_2) > m_1(x_1, x_2), \hat{e}_{2,n}(x_1, x_2) \leq m_2(x_1, x_2)\}
\]

\[
Q_2 = \{(x_1, x_2) : \hat{e}_{1,n}(x_1, x_2) \leq m_1(x_1, x_2), \hat{e}_{2,n}(x_1, x_2) > m_2(x_1, x_2)\}
\]

\[
Q_{12} = \{(x_1, x_2) : \hat{e}_{1,n}(x_1, x_2) > m_1(x_1, x_2), \hat{e}_{2,n}(x_1, x_2) > m_2(x_1, x_2)\}
\]

\[
Q = \{(x_1, x_2) : \hat{e}_{1,n}(x_1, x_2) \leq m_1(x_1, x_2), \hat{e}_{2,n}(x_1, x_2) \leq m_2(x_1, x_2)\}.
\]

Then for \(j = 1\),

\[
\left|\hat{e}_{1,n}(x_1, x_2) - e_{1,n}(x_1, x_2)\right| = \left|\hat{e}_{1}(x_1, x_2) - e_{1}(x_1, x_2)\right| I_{Q_{1} \cup Q_{12}}
\]

\[
+ \left|m_{1,n}(x_1, x_2) - e_{1}(x_1, x_2)\right| I_{Q_{2} \cup Q_{12}}
\]

\[
\leq \left|\hat{e}_{1,n}(x_1, x_2) - e_{1,n}(x_1, x_2)\right| I_{Q_{1} \cup Q_{12}}
\]

\[
+ \left|m_{1,n}(x_1, x_2) - e_{1}(x_1, x_2)\right| I_{Q_{2} \cup Q_{12}}
\]

\[
= \left|\hat{e}_{1,n}(x_1, x_2) - e_{1}(x_1, x_2)\right|.
\]

A similar statement holds for \(e_{2,n}^*\), and therefore, for \(j = 1, 2\),

\[
\left|e_{j,n}(x_1, x_2) - e_{j}(x_1, x_2)\right| \leq \left|\hat{e}_{j,n}(x_1, x_2) - e_{j}(x_1, x_2)\right|.
\]

(4.4.2)

The result then follows immediately. \(\square\)

As it was the case of Section 4.2, it turns out that \(\hat{e}^*(x_1, x_2)\) is the projection of the empirical mrl onto the convex set of mrl functions \(k(x_1, x_2)\) such that \(k(x_1, x_2) \geq m(x_1, x_2)\). The following result provides the interpretation of our estimator as a projection onto an appropriate convex set.
Theorem 4.4.2. Let $A$ be the convex set of all mrl functions bounded below by a known mrl function $m(x_1, x_2)$. That is, let $A = \{k(x_1, x_2) : k$ is a mrl and $k(x_1, x_2) \geq m(x_1, x_2)$ for all $(x_1, x_2) \in \mathcal{X}\}$. Let $\hat{e}^*(x_1, x_2) = \hat{e}(x_1, x_2) \vee m(x_1, x_2)$. Then,

(i) For any $k \in A$,

$$\sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|\hat{e}^*(x_1, x_2) - \hat{e}(x_1, x_2)\| \leq \sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|k(x_1, x_2) - \hat{e}(x_1, x_2)\|$$

(ii) For $p \geq 1$ and for all $k \in A$

$$\|\hat{e}^* - \hat{e}\|_p \leq \|k - \hat{e}\|_p.$$

Proof. Here we will prove (i) first. From (4.4.1) we have the following:

$$\hat{e}^*(x_1, x_2) - \hat{e}(x_1, x_2) = \begin{cases} 
0 & \text{if } \hat{e}(x_1, x_2) \geq m(x_1, x_2) \\
m(x_1, x_2) - \hat{e}(x_1, x_2) & \text{otherwise.}
\end{cases}$$

But when $\hat{e}(x_1, x_2) < m(x_1, x_2)$, $\hat{e}(x_1, x_2) < m(x_1, x_2) \leq k(x_1, x_2) \in A$ and this is equivalent to $\|m(x_1, x_2) - \hat{e}(x_1, x_2)\| \leq \|k(x_1, x_2) - \hat{e}(x_1, x_2)\|$. This implies that $\|\hat{e}^*(x_1, x_2) - \hat{e}(x_1, x_2)\| \leq \|k(x_1, x_2) - \hat{e}(x_1, x_2)\|$. Hence, for all $k(x_1, x_2) \in A$,

$$\sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|\hat{e}^*(x_1, x_2) - \hat{e}(x_1, x_2)\| \leq \sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|k(x_1, x_2) - \hat{e}(x_1, x_2)\|.$$

The proof of (ii) follows from the arguments in (i).

4.4.2 Asymptotic properties of the estimator

The results of the previous section show that pointwise and for every $n$, the estimator $\hat{e}^*$ is closer to $e$ than the empirical mrl is to $e$. Thus, in this case, as it was the case in
Section 4.2, our estimator renders the empirical mean residual life function estimator $\hat{e}_n$ inadmissible with respect to a large class of loss functions. The following three results delineate the asymptotic distribution theory of $\hat{e}^*$. It turns out that $\hat{e}^*$ is asymptotically unbiased, uniformly strongly consistent and converges weakly to a bivariate Gaussian process.

**Theorem 4.4.3.** The estimator in (4.4.1) is asymptotically unbiased. That is,

$$E[\hat{e}^*(x_1, x_2)] \rightarrow e(x_1, x_2) \text{ as } n \rightarrow \infty$$

for all $(x_1, x_2) \in \mathcal{X}$.

**Proof.** Suppose $e(x_1, x_2) \geq m(x_1, x_2)$ for all $(x_1, x_2) \in \mathcal{X}$. Since $\hat{e}^*(x_1, x_2) = m(x_1, x_2)$ $\forall \hat{e}(x_1, x_2)$, then

$$E[\hat{e}^*(x_1, x_2) - e(x_1, x_2)] = E[e(x_1, x_2) + \min\{\hat{e}(x_1, x_2), m(x_1, x_2)\}]$$

$$= E[\min\{e(x_1, x_2) - m(x_1, x_2), e(x_1, x_2) - \hat{e}(x_1, x_2)\}] = 0.$$

But $\min\{e(x_1, x_2) - m(x_1, x_2), e(x_1, x_2) - \hat{e}(x_1, x_2)\}$ converges almost surely to zero since $e(x_1, x_2) - \hat{e}(x_1, x_2)$ converges almost surely to zero. Moreover, $\min\{e(x_1, x_2) - m(x_1, x_2), e(x_1, x_2) - \hat{e}(x_1, x_2)\} \leq e(x_1, x_2) - m(x_1, x_2)$. Hence, it follows from the dominated convergence theorem that $E[\hat{e}^*(x_1, x_2) - e(x_1, x_2)] \rightarrow 0$.

**Theorem 4.4.4.** The estimator in (4.4.1) is strongly uniformly consistent on any finite rectangle. That is, for fixed $(b_1, b_2) \in \mathcal{X}$ and $D = [0, b_1] \times [0, b_2]$,

$$\sup_{(x_1, x_2) \in D} \|\hat{e}^*(x_1, x_2) - e(x_1, x_2)\| \rightarrow 0$$
with probability one as \( n \rightarrow \infty \).

**Proof.** The proof of Theorem 4.4.4 follows immediately from (4.4.2) and Theorem 2 of Kulkarni and Rattihalli (2002).

The following result states that \( \hat{e}^* \) suitably normalized converges weakly to a Gaussian process. Let \( Z^*_n \) be defined by:

\[
Z^*_n(x_1, x_2) = \sqrt{n} \left[ \hat{e}^*(x_1, x_2) - e(x_1, x_2) \right]
= \sqrt{n} \left[ \hat{e}(x_1, x_2) \vee m(x_1, x_2) - e(x_1, x_2) \right].
\]

**Theorem 4.4.5.** Let \( \mathcal{X} = \{(x, y) : \bar{F}(x, y) > 0 \text{ and } \bar{G}(x, y) > 0\} \) and let \( e(x_1, x_2) \geq m(x_1, x_2) \) where \( m(x_1, x_2) \) is known. Let \( \{Z(x_1, x_2), \ x_1 > 0, x_2 > 0\} \) denote the bivariate Gaussian process obtained in Kulkarni and Rattihalli (2002) as the weak limit of the bivariate empirical process.

(i) If \( e > m \) on \( \mathcal{X} \), then

\[
Z^*_n \Rightarrow Z \text{ on } \mathcal{X}.
\]

(ii) If \( e = m \) on \( \mathcal{X} \), then

\[
Z^*_n \Rightarrow Z \vee 0 \text{ on } \mathcal{X}.
\]

(iii) If \( e(x_0, y_0) = m(x_0, y_0) \) for some \( (x_0, y_0) \in \mathcal{X} \) and \( e(s_1, s_2) > m(s_1, s_2) \) for all \( (s_1, s_2) \in (x_0, x_0 + \delta_1] \times (y_0, y_0 + \delta_2] \) where \( (\delta_1, \delta_2) > 0 \), then \( Z^*_n \) does not converge weakly.
Proof. (i) 

$$Z_n^* = \sqrt{n} [\hat{e}(x_1, x_2) - e(x_1, x_2)] \vee \sqrt{n} [m(x_1, x_2) - e(x_1, x_2)]$$

But $e(x_1, x_2) > m(x_1, x_2)$ for all $(x_1, x_2) \in \mathcal{X}$ implies that $\sqrt{n} [m(x_1, x_2) - e(x_1, x_2)] \to -\infty$ uniformly. And we know that $\sqrt{n} [\hat{e}(x_1, x_2) - e(x_1, x_2)] \Rightarrow Z$. Hence, by the continuous mapping theorem

$$Z_n^* \Rightarrow Z \vee -\infty = Z$$

(ii) When $e(x_1, x_2) = m(x_1, x_2)$ for all $(x_1, x_2) \in \mathcal{X}$, then $\sqrt{n} [m(x_1, x_2) - e(x_1, x_2)] = 0$. This implies that $Z_n^* = \sqrt{n} [\hat{e}(x_1, x_2) - e(x_1, x_2)] \vee 0$. Then as in (i) by the continuous mapping theorem,

$$Z_n^* \Rightarrow Z \vee 0$$

(iii) Without loss of generality suppose that there is $(\delta_1, \delta_2) > (0, 0)$ such that $e(s_1, s_2) > m(s_1, s_2)$ for all $(s_1, s_2) \in (x_0, x_0 + \delta_1] \times (y_0, y_0 + \delta_2]$. Note that $\hat{e}(s_1, s_2) \to e(s_1, s_2) > m(s_1, s_2)$ with probability one as $n \to \infty$, this implies that $\hat{e}(s_1, s_2) > m(s_1, s_2)$ eventually with probability one. On this event, $e^*(s_1, s_2) = \hat{e}(s_1, s_2)$. When $e(x_0, y_0) = m(x_0, y_0)$ for some $(x_0, y_0) \in \mathcal{X}$, then $Z_n^*(x_0, y_0) = \sqrt{n} [\hat{e}(x_0, y_0) - e(x_0, y_0)] \vee 0$. Thus, for $\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} > 0$ and for $a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \neq 0$,

$$\lim_{n \to \infty} P \left\{ \sup_{(x_0, y_0) < (s_1, s_2) \leq (x_0 + \delta_1, y_0 + \delta_2)} |Z_n^*(s_1, s_2) - Z_n^*(x_0, y_0)| < \epsilon \right\}$$

$$\leq \lim_{n \to \infty} P \{ |Z_n^*(s_1, s_2) - Z_n^*(x_0, y_0)| < \epsilon \}$$
\[ \lim_{n \to \infty} P \left\{ \sqrt{n} \left[ e^*(s_1, s_2) - e(s_1, s_2) \right] \right\} \]

\[ - \sqrt{n} \left[ \{ \hat{e}(x_0, y_0) - e(x_0, y_0) \} \lor 0 \right] < \epsilon \}

\[ \leq \lim_{n \to \infty} P \left\{ \sqrt{n} \left[ \hat{e}(s_1, s_2) - e(s_1, s_2) \right] \right\} \]

\[ - \sqrt{n} \left[ \{ \hat{e}(x_0, y_0) - e(x_0, y_0) \} \lor 0 \right] < \epsilon \}

\[ \leq \lim_{n \to \infty} P \left\{ \sqrt{n} \left[ \hat{e}(s_1, s_2) - e(s_1, s_2) \right] < \epsilon \right\} \]

\[ = \lim_{n \to \infty} P \left\{ \frac{a^*}{\sqrt{a'a}} \sum_{n=0}^{-\frac{1}{2}} \sqrt{n} \left[ \hat{e}(s_1, s_2) - e(s_1, s_2) \right] < \frac{a^*}{\sqrt{a'a}} \sum_{n=0}^{-\frac{1}{2}} \epsilon \right\} \]

\[ = \Phi \left( \frac{a^*}{\sqrt{a'a}} \sum_{n=0}^{-\frac{1}{2}} \epsilon \right), \]

where \( \Sigma \) denotes the covariance matrix of the bivariate Gaussian process and \( \Phi \) denotes the standard normal c.d.f. This shows that \( Z^*_n \) is not tight on \( (x_0, x_0 + \delta_1) \times (y_0, y_0 + \delta_2) \) and can not converge weakly. \( \Box \)

### 4.5 The Censored Data Case

We now turn our attention to the censored data case. Let \( (X_i, Y_i), i = 1, 2, \ldots, n \) be independent and identically distributed pairs of failure times with survival function \( S(x, y) = P(X > x, Y > y) \) and let \( C_i, i = 1, 2, \ldots, n \) be \( n \) independent and identically distributed censoring times with survival function \( G(t) = P(C > t) \). Suppose that the two sequences \( \{(X_i, Y_i)\}_{i=1}^n \) and \( \{(C_i)\}_{i=1}^n \) are independent. In the random univariate censorship model from the right, the \( (X_i, Y_i) \) may be censored on the right by the single censoring variable \( C_i \), so that we observe the random vectors \( (\tilde{X}_i, \tilde{Y}_i, \delta_i^x, \delta_i^y), i = 1, 2, \ldots, n \), where \( \tilde{X}_i = X_i \land C_i, \tilde{Y}_i = Y_i \land C_i, \delta_i^x = I(X_i \leq C_i) \) and \( \delta_i^y = I(Y_i \leq C_i) \).
and \( \delta_i^U = I(Y_i \leq C_i) \). The survival function of the observed pairs \( \{\bar{X}_i, \bar{Y}_i\}_{i=1}^n \) is \( S(x, y) \hat{G}(x \vee y) \), which is a simple consequence of the independence between \( (X, Y) \) and \( C \). Thus, under the univariate censoring, it is natural to estimate the survival function \( S(x, y) \) by

\[
\hat{S}_n(x, y) = \frac{n^{-1} \sum_{i=1}^n I(\bar{X}_i > x, \bar{Y}_i > y)}{\hat{G}(x \vee y)} \tag{4.5.1}
\]

where the numerator is the empirical estimator for the survival function of the observed pairs and the denominator is the product-limit estimator for \( \hat{G}(\cdot) \). Lin and Ying (1993) have showed that \( \hat{S}_n(x, y) \) is strongly consistent, and upon proper normalization, converges weakly to a zero mean Gaussian process for all \( (x, y) \in [0, \tau]^2 \), where \( \tau \) satisfies \( S(\tau, \tau) \hat{G}(\tau) > 0 \). Using \( \hat{S}_n(x, y) \), one can estimate the bivariate mrl function \( e(x, y) = (e_1(x, y), e_2(x, y)) \) by \( \hat{e}_n(x, y) = (\hat{e}_{1,n}(x, y), \hat{e}_{2,n}(x, y)) \) where

\[
\hat{e}_{i,n}(x, y) = \begin{cases} \{\hat{S}_n(x, y)\}^{-1} \int_x^\infty \hat{S}_n(u, y) du, & \text{if } i = 1, \\ \{\hat{S}_n(x, y)\}^{-1} \int_y^\infty \hat{S}_n(x, v) dv, & \text{if } i = 2, \end{cases} \tag{4.5.2}
\]

Sohn et al (1996) showed uniform consistency and weak convergence of \( \hat{e}_n(x, y) \) on bounded rectangles under the assumption that \( \sqrt{n} \int_x^\tau S(u, y) du \) and \( \sqrt{n} \int_y^\tau S(x, v) dv \) converge to zero in probability, where \( \bar{X}^* = \max(\bar{X}_1, \ldots, \bar{X}_n) \) and \( \bar{Y}^* = \max(\bar{Y}_1, \ldots, \bar{Y}_n) \).

Now suppose that \( e(x, y) \leq m(x, y) \) for all \( (x, y) \in [0, \tau]^2 \) and for some known bivariate mrl function \( m(x, y) \). As an estimator of \( e(x, y) \) based on the censored data and under the constraint \( e(x, y) \leq m(x, y) \), we propose

\[
e_n^c(x, y) = \min(\hat{e}_n(x, y), m(x, y)) \tag{4.5.3}
\]
where $\hat{e}_n(x, y)$ is the estimator of $e(x, y)$ defined by (4.5.2). The consistency of $e_n^c(x, y)$ follows immediately from the consistency of $\hat{e}_n(x, y)$. To prove weak convergence of the proposed estimator, let us consider the process

$$Z_n^c(x, y) = \sqrt{n} \left\{ [e_n^c(x, y) - e(x, y)], (x, y) \in [0, \tau]^2 \right\}$$

$$= \sqrt{n} \left[ \min(\hat{e}_n(x, y), m(x, y)) - e(x, y) \right]$$

$$= \min \left( \sqrt{n}[\hat{e}_n(x, y) - e(x, y)], \sqrt{n}[m(x, y) - e(x, y)] \right)$$

The following result establishes the weak convergence of $Z_n^c(x, y)$.

**Theorem 4.5.1.** Let $(X_i, Y_i), i = 1, 2, \ldots, n$ be i.i.d. pairs of failure times with survival and mrl functions given by $S(x, y)$ and $e(x, y)$ respectively; $C_i, i = 1, 2, \ldots, n$ be independent and identically distributed censoring times with survival function $G(t)$. Suppose that the two sequences $\{(X_i, Y_i)\}_{i=1}^n$ and $\{(C_i)\}_{i=1}^n$ are independent and $e(x, y) \leq m(x, y)$ where $m(x, y)$ is known. Let $\{Z(x, y), (x, y) \in [0, \tau]^2\}$ denote the bivariate Gaussian process obtained in Sohn et al (1996) as the weak limit of the bivariate empirical mrl process under univariate censoring.

(i) If $e < m$ on $[0, \tau]^2$, then

$$Z_n^c \Rightarrow Z \text{ on } [0, \tau]^2.$$

(ii) If $e = m$ on $[0, \tau]^2$, then

$$Z_n^c \Rightarrow Z \land 0 \text{ on } [0, \tau]^2.$$

(iii) If for $i = 1$ or $i = 2$, $e_i(x_0, y_0) = m_i(x_0, y_0)$ for some $(x_0, y_0) \in [0, \tau]^2$ and
\( e_i(x, y) < m_i(x, y) \) for all \((x, y) \in [0, \tau)^2\) in the line segment \(\alpha(x_0, y_0) + (1 - \alpha)(x_1, y_1)\),

\( 0 < \alpha < 1 \) for some \((x_1, y_1) \in \mathcal{X}\), then \(Z_n^\ast\) does not converge weakly.

\textit{Proof.} The proof of this Theorem is similar to the proof of Theorem 4.2.5. \(\square\)

In the two sample case with censored data, similar arguments can be used to develop estimators of the mrl function under mrl ordering.

### 4.6 Tests and Confidence Sets for MRL Functions

In this section we develop a test and confidence procedures for comparing mrl functions from two populations or treatment groups. Let \(X_i = (X_{i1}, X_{i2})\), \(i = 1, 2, \ldots, n\) be a random vector in \(\mathbb{R}^2\) with corresponding survival and mrl functions \(S(x, y)\) and \(e(x, y)\) respectively. Moreover, let \(e^\circ(x, y)\) be the bivariate mrl function of the standard population (treatment group) and assume \(e^\circ(x, y)\) is known and defined for all \((x, y) \in \mathcal{X}\) where \(\mathcal{X} = \{(x, y) : S(x, y) > 0\}\). We will consider the problem of testing the hypothesis \(H_0: e(x, y) = e^\circ(x, y)\) and a related problem of finding the confidence region for \(e(x, y)\) for each \((x, y) \in \mathcal{X}\).

If \(e(x, y) \leq e^\circ(x, y)\) for all \((x, y) \in \mathcal{X}\), a reasonable estimator of the mrl function \(e\) is \(e_n^\ast(x, y) = \min(\hat{e}_n(x, y), e^\circ(x, y)\) where \(\hat{e}_n\) the empirical mrl function that corresponds to \(e\). Recall that in Section 4.2 we showed that under \(H_0\), the process \(Z_n^\ast(x, y) = \sqrt{n}\{[e_n^\ast(x, y) - e^\circ(x, y)], (x, y) \in \mathcal{X}\}\) converges to \(Z \land 0\) where \(Z = (Z_1, Z_2)\) is a zero mean bivariate Gaussian processes. Thus, an asymptotic size \(\alpha\) test for test-
ing $H_0 : e(x, y) = e^0(x, y)$ against the alternative $H_\alpha : e(x, y) \leq e^0(x, y)$, with strict inequality for some $(x, y)$, rejects $H_\alpha$ when

\[
D_n^- = \sqrt{n} \sup_{(x,y) \in X} [e^0(x,y) - e_n^*(x,y)]
\]

\[
= \sqrt{n} \sup_{(x,y) \in X} \max \{e_1^0(x,y) - e_{1,n}^*(x,y), e_2^0(x,y) - e_{2,n}^*(x,y)\} > K_\alpha, \quad \text{where}
\]

$K_\alpha$ is the $(1 - \alpha) \times 100$ quantile of the distribution $\max\{\sup_X Z_1 \wedge 0, \sup_X Z_2 \wedge 0\}$. The above test can be inverted to give a confidence band for $e(x_1, x_2)$. Note that (4.6.1) is equivalent to the statement

\[
P[\sup_{(x,y) \in X} \max \{e_1(x,y) - e_{1,n}^*(x,y), e_2(x,y) - e_{2,n}^*(x,y)\} \leq K_\alpha] = 1 - \alpha.
\]

Thus, we define $L_n(x, y) = 0$ and $U_n(x, y) = \max\{e_{1,n}^*(x,y), e_{2,n}^*(x,y)\} + K_\alpha$ and call the region between $L_n(x,y)$ and $U_n(x,y)$ a confidence set for $e(x,y)$, with associated confidence coefficient $1 - \alpha$. The above procedures can be easily extended to the case when the bounding mrl function $e^0(x,y)$ is unknown.

### 4.7 Simulation Results

Simulation studies were carried out to examine the properties of the proposed estimators. Various sample sizes 15, 30 and 45 were considered. Each simulation consisted of a series of 10000 trials. Several bivariate distributions were used for the simulation study:

1. Bivariate Gumbel distribution with density function

\[
f_{i}(x,y; \theta) = [(1 + \theta_i x)(1 + \theta_i y) - \theta_i] \exp(-x - y - \theta_i xy)
\]
where \( x, y > 0; 0 \leq \theta_i \leq 1, i = 1, 2. \)

2. Bivariate pareto distribution with density function

\[
f_i(x_1, x_2, \theta_1, \theta_2, a_i) = a_i (a_i + 1) (\theta_1 \theta_2)^{(a_i+1)} (\theta_2 x_1 + \theta_1 x_2 - \theta_1 \theta_2)^{-(a_i+2)}
\]

where \( x_1 \geq \theta_1 > 0, x_2 \geq \theta_2 > 0, a_i > 0, i = 1, 2. \)

3. Bivariate Morgenstern (1956) distribution with density function

\[
f_i(u_1, u_2, \alpha_i) = 1 + \alpha_i(2u_1 - 1)(2u_2 - 1)
\]

where \( 0 \leq u_1, u_2 \leq 1, -1 \leq \alpha_i \leq 1, i = 1, 2. \)

4. Bivariate Sarmanov (1966) distribution with density function

\[
f_i(x_1, x_2, \alpha_i) = e^{-x_1}e^{-x_2}\left\{1 + \alpha_i(2e^{-x_1} - 1)(2e^{-x_2} - 1)\right\}
\]

where \( 0 \leq x_1, x_2 < \infty, -1 \leq \alpha_i \leq 1, i = 1, 2. \)

The parameters of the four distributions are chosen such that \( e_1(x, y) \leq e_2(x, y) \) where \( e_i(x, y), i = 1, 2, \) is the mrl function that corresponds to \( f_i(x, y). \) Simulation results are plotted in Appendix D for the ratio of the mean squared error of the restricted estimator to the empirical estimator. A careful examination of the simulation results reveals the following: (1) In the case of the one sample problem the mean squared error of the proposed estimator is uniformly smaller than that of the empirical estimator. Moreover, the ratio of the mean squared error of the proposed estimator to the mean squared error of the empirical estimator increases in \( x(y) \) for fixed \( y(x). \) The increment
is more prominently reflected at the extreme values of \( x \) and \( y \) and for small \( n \). (2) The above remarks also hold true in the two-sample problem when estimating the smaller of the two mean residual life functions. Unlike as in the case of the minimum estimator of the two sample problem, for the maximum estimator the ratio of the mean squared error of the proposed estimator to the empirical estimator is not always less than one. There are few points in the support set for which this ratio is either one or few decimal points larger than one. In the “mirror image” of the one sample problem the ratio of the mean squared error of the proposed estimator to the empirical estimator is close to zero. (3) For small sample sizes and large \( x \) and \( y \) values, the proposed estimators have small \(|\text{bias}|\) for all the distributions from which we simulated and in the three cases which we considered.

It is clear from the simulations that the new estimators dominate the empirical mean residual life function uniformly for all the cases examined in the simulation study. This conclusion is not surprising in the one sample case in view of Theorem 4.2.1 and Theorem 4.4.1, although the gain in mean squared error in the case of the estimator presented in Section 4.4 is substantially more than the gain obtained in connection with the problem discussed Section 4.2. What is truly surprising is the uniformity of the gain in the two-sample problem. It is easy to understand that knowing \( m(x, y) \) provides a “benchmark” that our estimators take advantage of to calibrate themselves. But the simulation work shows that even in the absence of this benchmark, the estimators can still beat the empirical mean residual function.
Chapter 5

Application to a Real Data Set

In this chapter we will use real data sets to illustrate the estimators developed in
the previous chapters. We should emphasize that our motivation for choosing these
data sets is simply to illustrate our estimators. It is understood that in some of
the examples considered here the assumption that one mean residual life function is
uniformly larger than a second mean residual life function may need closer scrutinizing
and a valid justification must be provided to imply such an assumption. Nevertheless,
the assumption seems to be a good approximation. The first data set that we use is
from the acute myelogenous leukemia (AML) study that was conducted at Stanford
University in 1977. We will obtain an estimator (under censoring) of mrl function
for this data set. The second data set involves a clinical trial carried out on bladder
cancer patients. We will use this data set to illustrate the empirical bivariate mrl
function. Finally, we will use a data set from the Diabetic Retinopathy Study (DRS)
to obtain the estimators of the bivariate mrl function under mrl ordering. In the DRS diabetes study, patients are classified into two general groups by the age at the onset: juvenile (< 20) and adult diabetics. Hence, it seems natural to assume that the mrl for the juvenile diabetics should be longer than the mrl for adult diabetics.

5.1 AML Maintenance Study

A clinical trial to evaluate the efficacy of maintenance chemotherapy for acute myelogenous leukemia (AML) was conducted by Embury et al. (1977) at Stanford University. After reaching a state of remission through treatment by chemotherapy, the patients who entered the study were randomized into two groups. The first group received chemotherapy; the second group or control group did not. The objective of the trial was to see if maintenance chemotherapy prolonged the time until relapse, that is, increased the length of remission. For preliminary analysis during the course of the trial the data were as follows (+ indicates a censored value):

Table 5.1: Data for AML maintenance study

<table>
<thead>
<tr>
<th>Group</th>
<th>Length of complete remission (in weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintained</td>
<td>9,13,13+,18,23,28+,31,34,45+,48,161+</td>
</tr>
<tr>
<td>Nonmaintained</td>
<td>5,5,8,8,12,16+,23,27,30,33,45,45</td>
</tr>
</tbody>
</table>

The Kaplan-Meier survival curves and the corresponding mrl function are plotted in Figure 5.1 and Figure 5.2 respectively. Although both figures show that the
Figure 5.1: Kaplan-Meier curves for the AML study.

Figure 5.2: Mean residual functions for the AML study.
maintained group survived longer than the nonmaintained group, Figure 5.2 shows us the time points where the remission times of the two groups are close (or far apart). Moreover, Figure 5.2 shows that the empirical mrl function of the maintained group decreases in time. This suggests that the distribution of $X$, where $X$ is the length of remission for the maintained group, is a member of DMRL. But the empirical mrl function of the nonmaintained group, as it is shown in Figure 5.2, first increases and then show a steep decrease in time. Thus, this suggests that the distribution of $Y$, where $Y$ is the length of remission for the nonmaintained group, is a member of IDMRL.

5.2 Bladder Cancer Data Set

In this section we will use the multiple tumor recurrence data for patients with bladder cancer (Andrews and Herzberg (1985)). Andrews and Herzberg (1985) presented the data as follows: "The data were obtained in a randomized clinical trial conducted by the Veterans Administration Co-operative Urological Research Group (VACURG). All patients had superficial bladder tumors when they entered the trial. These tumors were removed transurethrally and patients were assigned randomly to one of these treatments: placebo pills, pyridoxine (vitamin $B_6$) pills, or periodic instillation of a chemotherapeutic agent, thiotepa, into the bladder. At subsequent follow-up visits any tumors noticed were removed and the treatment was continued. The goal of the study was to determine the effect of treatment on the frequency of tumor recurrence,
where the word recurrence refers to a visit at which one or more tumors are found on
the bladder regardless of whether these are thought to be recurrences or new tumors.
This term is not to be confused with the number of tumors present at any single visit
because the tumors are often multiple. An analysis of perhaps secondary importance
would compare treatments with respect to numbers and size of tumors.”. As we
said earlier, our motivation for choosing this data sets is not to analyze it but to
demonstrate our estimator. For this purpose, let us denote the time (in months) to
first and second recurrence of a tumor by $X$ and $Y$ respectively.

The data set for those patients assigned to the placebo, pyridoxine, and thiotepa-
treatment groups are respectively given in Table 5.2, Table 5.3 and Table 5.4. The
bivariate empirical mrl functions are calculated for each treatment group and are
displayed in Table 5.5, Table 5.6 and Table 5.7. Besides, we have also plotted the
surface plots of the bivariate empirical mrl functions for the data tabulated in Tables
(5.5-5.7).

Table 5.2: Lifetimes in months for bladder tumor patients undergoing placebo pills
treatment

<table>
<thead>
<tr>
<th>Patient i</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>12</td>
<td>10</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>28</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>$Y_i$</td>
<td>16</td>
<td>15</td>
<td>9</td>
<td>10</td>
<td>15</td>
<td>26</td>
<td>30</td>
<td>17</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Patient i</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>$X_i$</td>
<td>12</td>
<td>9</td>
<td>16</td>
<td>3</td>
<td>9</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>$Y_i$</td>
<td>15</td>
<td>17</td>
<td>19</td>
<td>6</td>
<td>11</td>
<td>15</td>
<td>15</td>
<td>14</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>
Figure 5.3: Surface plots for the empirical mrl function for patients on the placebo (top), pyridoxine (center) and thiotepa (bottom) treatments.
Table 5.3: Lifetimes in months for bladder tumor patients undergoing pyridoxine treatment

<table>
<thead>
<tr>
<th>Patient i</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>8</td>
<td>44</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>$Y_i$</td>
<td>4</td>
<td>3</td>
<td>10</td>
<td>9</td>
<td>7</td>
<td>6</td>
<td>20</td>
<td>15</td>
<td>47</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 5.4: Lifetimes in months for bladder tumor patients undergoing thiotepa treatment

<table>
<thead>
<tr>
<th>Patient i</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>1</td>
<td>17</td>
<td>6</td>
<td>26</td>
<td>22</td>
<td>4</td>
<td>24</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>$Y_i$</td>
<td>3</td>
<td>19</td>
<td>12</td>
<td>35</td>
<td>23</td>
<td>16</td>
<td>26</td>
<td>27</td>
<td>20</td>
<td>24</td>
</tr>
</tbody>
</table>

The bivariate empirical mrl functions and their corresponding surface plots indicate that if the time to first recurrence is large, then the second recurrence is likely to occur after a long period. In contrast, if the first recurrence occur immediately, then the second recurrence also would be expected to occur soon. This is true for the three treatment groups. The bivariate empirical mrl functions and their corresponding surface plots also indicate that the recurrence of tumors is longer for the thiotepa treatment group and shorter for the placebo treatment group. Moreover, one can see that the distribution of $(X, Y)$ is a member of DMRL for the placebo group and DIMRL for the pyridoxine group. For the thiotepa treatment group, $\hat{e}_X(x, y)$ first increases in $x$ and then decreases for fixed values of $y$; but $\hat{e}_Y(x, y)$ decreases in $y$ for every fixed $x$. 
Table 5.5: The empirical mrl function for patients on the thiopeta treatment

<table>
<thead>
<tr>
<th>x\y</th>
<th>0</th>
<th>3</th>
<th>12</th>
<th>16</th>
<th>19</th>
<th>23</th>
<th>26</th>
<th>27</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10.70</td>
<td>11.77</td>
<td>12.50</td>
<td>13.71</td>
<td>13.16</td>
<td>13.75</td>
<td>13.50</td>
<td>26.00</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>20.50</td>
<td>19.44</td>
<td>11.75</td>
<td>8.85</td>
<td>6.83</td>
<td>5.00</td>
<td>5.00</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>12.12</td>
<td>12.12</td>
<td>13.14</td>
<td>14.83</td>
<td>14.60</td>
<td>17.00</td>
<td>25.00</td>
<td>25.00</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>21.87</td>
<td>18.87</td>
<td>11.28</td>
<td>8.50</td>
<td>6.60</td>
<td>5.33</td>
<td>9.00</td>
<td>8.00</td>
<td>9.00</td>
</tr>
<tr>
<td>2</td>
<td>12.71</td>
<td>12.71</td>
<td>14.16</td>
<td>16.60</td>
<td>17.00</td>
<td>16.00</td>
<td>25.00</td>
<td>25.00</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>22.14</td>
<td>19.14</td>
<td>11.83</td>
<td>9.40</td>
<td>8.00</td>
<td>5.33</td>
<td>9.00</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>15.00</td>
<td>15.00</td>
<td>18.25</td>
<td>18.25</td>
<td>20.00</td>
<td>21.00</td>
<td>22.00</td>
<td>22.00</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>23.00</td>
<td>20.00</td>
<td>13.75</td>
<td>9.75</td>
<td>9.00</td>
<td>7.50</td>
<td>9.00</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>16.25</td>
<td>16.25</td>
<td>16.25</td>
<td>16.25</td>
<td>18.00</td>
<td>19.00</td>
<td>20.00</td>
<td>20.00</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>25.75</td>
<td>22.75</td>
<td>13.75</td>
<td>9.75</td>
<td>9.00</td>
<td>7.50</td>
<td>9.00</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
<td>8.00</td>
<td>9.00</td>
<td>9.00</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>28.00</td>
<td>25.00</td>
<td>16.00</td>
<td>12.00</td>
<td>9.00</td>
<td>7.50</td>
<td>9.00</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>4.00</td>
<td>4.00</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>30.5</td>
<td>27.50</td>
<td>18.50</td>
<td>14.50</td>
<td>11.50</td>
<td>7.50</td>
<td>9.00</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>35.00</td>
<td>32.00</td>
<td>23.00</td>
<td>19.00</td>
<td>16.00</td>
<td>12.00</td>
<td>9.00</td>
<td>8.00</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 5.6: The empirical mrl function for patients on the pyridoxine treatment

<table>
<thead>
<tr>
<th>x \ y</th>
<th>0</th>
<th>3</th>
<th>4</th>
<th>6</th>
<th>7</th>
<th>9</th>
<th>10</th>
<th>14</th>
<th>15</th>
<th>20</th>
<th>47</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8.20&lt;sup&gt;a&lt;/sup&gt;</td>
<td>8.88</td>
<td>9.63</td>
<td>10.71</td>
<td>12</td>
<td>13.8</td>
<td>16.5</td>
<td>19.33</td>
<td>25</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>13.50&lt;sup&gt;b&lt;/sup&gt;</td>
<td>11.67</td>
<td>12.00</td>
<td>11.43</td>
<td>12.17</td>
<td>12.20</td>
<td>14</td>
<td>13.33</td>
<td>18.50</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>7.75</td>
<td>7.75</td>
<td>8.71</td>
<td>8.71</td>
<td>10</td>
<td>11.8</td>
<td>14.5</td>
<td>17.33</td>
<td>23</td>
<td>42</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>15.75</td>
<td>12.75</td>
<td>13.43</td>
<td>11.43</td>
<td>12.17</td>
<td>12.2</td>
<td>14</td>
<td>13.33</td>
<td>18.5</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>13.5</td>
<td>13.5</td>
<td>13.5</td>
<td>13.5</td>
<td>13.5</td>
<td>13.5</td>
<td>13.5</td>
<td>16.33</td>
<td>22</td>
<td>41</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>21</td>
<td>20</td>
<td>18</td>
<td>17</td>
<td>15</td>
<td>14</td>
<td>13.33</td>
<td>18.5</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>20</td>
<td>38</td>
<td>38</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25.33</td>
<td>22.33</td>
<td>21.33</td>
<td>19.33</td>
<td>18.33</td>
<td>16.33</td>
<td>15.33</td>
<td>17</td>
<td>32</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>44</td>
<td>43</td>
<td>41</td>
<td>40</td>
<td>38</td>
<td>37</td>
<td>33</td>
<td>32</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>44</td>
<td>43</td>
<td>41</td>
<td>40</td>
<td>38</td>
<td>37</td>
<td>33</td>
<td>32</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>44</td>
<td>43</td>
<td>41</td>
<td>40</td>
<td>38</td>
<td>37</td>
<td>33</td>
<td>32</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>44</td>
<td>43</td>
<td>41</td>
<td>40</td>
<td>38</td>
<td>37</td>
<td>33</td>
<td>32</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>44</td>
<td>43</td>
<td>41</td>
<td>40</td>
<td>38</td>
<td>37</td>
<td>33</td>
<td>32</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>44</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 5.7: The empirical mrl function for patients on the placebo treatment

<table>
<thead>
<tr>
<th>x \ y</th>
<th>0</th>
<th>6</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7.05&lt;sup&gt;a&lt;/sup&gt;</td>
<td>7.52</td>
<td>7.87</td>
<td>8.20</td>
<td>8.29</td>
<td>8.23</td>
<td>8.50</td>
<td>10.29</td>
<td>11.40</td>
<td>15.33</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>14.74&lt;sup&gt;b&lt;/sup&gt;</td>
<td>9.76</td>
<td>8.25</td>
<td>7.73</td>
<td>7.21</td>
<td>6.69</td>
<td>4.00</td>
<td>5.14</td>
<td>5.80</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>6.40</td>
<td>7.23</td>
<td>7.23</td>
<td>7.75</td>
<td>8.00</td>
<td>8.10</td>
<td>8.67</td>
<td>11.60</td>
<td>15.67</td>
<td>20</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>14.27</td>
<td>9.54</td>
<td>7.54</td>
<td>7.08</td>
<td>6.64</td>
<td>6.20</td>
<td>3.56</td>
<td>4.60</td>
<td>6</td>
<td>7.50</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>16.33</td>
<td>10.33</td>
<td>8.33</td>
<td>7.33</td>
<td>7.13</td>
<td>7</td>
<td>4.67</td>
<td>5.50</td>
<td>6</td>
<td>7.5</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>7.88</td>
<td>7.88</td>
<td>7.88</td>
<td>7.88</td>
<td>8.71</td>
<td>9.50</td>
<td>9.50</td>
<td>11.25</td>
<td>12.67</td>
<td>17.00</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>16.63</td>
<td>10.63</td>
<td>8.63</td>
<td>7.63</td>
<td>7.57</td>
<td>7.67</td>
<td>4.67</td>
<td>5.50</td>
<td>6</td>
<td>7.5</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>6.71</td>
<td>6.71</td>
<td>6.71</td>
<td>6.71</td>
<td>6.71</td>
<td>7.5</td>
<td>7.5</td>
<td>9.25</td>
<td>10.67</td>
<td>15</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>17.57</td>
<td>11.57</td>
<td>9.57</td>
<td>8.57</td>
<td>7.57</td>
<td>7.67</td>
<td>4.67</td>
<td>5.5</td>
<td>6</td>
<td>7.5</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>6.6</td>
<td>6.6</td>
<td>6.6</td>
<td>6.6</td>
<td>6.6</td>
<td>6.6</td>
<td>6.6</td>
<td>6.6</td>
<td>9.67</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>13</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>5</td>
<td>6.67</td>
<td>8.5</td>
<td>7.5</td>
<td>11</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>8.67</td>
<td>12</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>14</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>6</td>
<td>6.67</td>
<td>8.5</td>
<td>7.5</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>24.5</td>
<td>18.5</td>
<td>16.5</td>
<td>15.5</td>
<td>14.5</td>
<td>13.5</td>
<td>10.5</td>
<td>9.5</td>
<td>8.5</td>
<td>7.5</td>
<td>11</td>
</tr>
<tr>
<td>16</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>24</td>
<td>22</td>
<td>21</td>
<td>20</td>
<td>19</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>28</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
5.3 The Diabetic Retinopathy Data

The 197 patients in this data set represent a 50% random sample of the patients with high-risk diabetic retinopathy as defined by the Diabetic Retinopathy Study (DRS). A discussion of this study can be found in Huster et al (1989). Diabetic retinopathy is a complication associated with diabetes mellitus consisting of abnormalities in the microvasculature within the retina of the eye. It is the leading cause of new cases of blindness in patients under 60 years of age in the United States and is the major cause of visual loss elsewhere in many industrialized countries. The study was begun in 1971 to study the effectiveness of laser photocoagulation in delaying the onset of blindness in patients with diabetic retinopathy. Patients with diabetic retinopathy in both eyes and visual acuity of 20/100 or better in both eyes were eligible for the study. One eye of each patient was randomly selected for treatment and the other eye was observed without treatment. For each eye, the event of interest was the time from initiation of treatment to the time when visual acuity dropped below 5/200 two visits in a row (call it “blindness”). Thus there is a built-in lag time of approximately 6 months (visits were every 3 months). Survival times in this data set are therefore the actual time to blindness in months, minus the minimum possible time to event (6.5 months). Huster et al (1989) stated that “The primary question of the DRS study was to assess the effectiveness of the laser photocoagulation treatment. Secondary questions were whether the survival times for the eyes of a patient were related and whether the treatment and type of diabetes were related. Diabetes can be classified
into two general groups by the age at the onset: juvenile (< 20 years) and adult diabetes”. For a complete description of the DRS study refer to the DRS Research Group (1976).

In the DRS study censoring was caused by death, dropout, or the end of the study. But in this work, we will consider the uncensored cases only because we can easily calculate the estimator of the bivariate survival function in the absence of censoring.

For each uncensored case i, the survival times of the treated \((X_i)\) and untreated \((Y_i)\) eyes are given in Table 5.8 and Table 5.9.

### Table 5.8: Survival times (months) for adult diabetics.

<table>
<thead>
<tr>
<th>Patient, i</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_i)</td>
<td>38.57</td>
<td>1.33</td>
<td>21.9</td>
<td>13.87</td>
<td>48.3</td>
<td>9.9</td>
<td>8.3</td>
</tr>
<tr>
<td>(Y_i)</td>
<td>30.83</td>
<td>5.77</td>
<td>25.63</td>
<td>25.8</td>
<td>5.73</td>
<td>9.9</td>
<td>8.3</td>
</tr>
</tbody>
</table>

Figure 5.5 and Figure 5.6 show the surface plots of the empirical and the restricted estimators respectively, where it is evident that the new estimators modify the empirical mrl function only in the region where the order is violated. Moreover, Figure 5.5 and Figure 5.6 indicate that the average remaining lifetime of both treated and untreated eyes of the juvenile group is larger than that of the adult group. This can be clearly seen from Figure 5.6.
Table 5.9: Survival times (months) for juvenile diabetics.

<table>
<thead>
<tr>
<th>Patient, i</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>6.9</td>
<td>1.63</td>
<td>13.83</td>
<td>35.53</td>
<td>14.8</td>
<td>6.2</td>
<td>22</td>
<td>1.7</td>
</tr>
<tr>
<td>$Y_i$</td>
<td>20.17</td>
<td>10.27</td>
<td>5.67</td>
<td>5.90</td>
<td>33.9</td>
<td>1.73</td>
<td>30.2</td>
<td>1.7</td>
</tr>
<tr>
<td>Patient, i</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>$X_i$</td>
<td>43.03</td>
<td>6.53</td>
<td>42.17</td>
<td>48.43</td>
<td>9.6</td>
<td>7.6</td>
<td>1.8</td>
<td>9.9</td>
</tr>
<tr>
<td>$Y_i$</td>
<td>1.77</td>
<td>18.7</td>
<td>42.17</td>
<td>14.3</td>
<td>13.33</td>
<td>14.27</td>
<td>34.57</td>
<td>21.57</td>
</tr>
<tr>
<td>Patient, i</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td>$X_i$</td>
<td>13.77</td>
<td>0.83</td>
<td>1.97</td>
<td>11.3</td>
<td>30.4</td>
<td>19</td>
<td>5.43</td>
<td>46.63</td>
</tr>
<tr>
<td>$Y_i$</td>
<td>13.77</td>
<td>10.33</td>
<td>11.07</td>
<td>2.1</td>
<td>13.97</td>
<td>13.80</td>
<td>13.57</td>
<td>42.43</td>
</tr>
</tbody>
</table>

Figure 5.4: Empirical survival curves for DRS data.
Figure 5.5: Empirical mrl function for the DRS data: juvenile (top) and adult (bottom).
Figure 5.6: Estimators of the mrl function under mrl ordering for the DRS data: juvenile (top) and adult (bottom).
Chapter 6

Conclusions

6.1 Summary

The main contributions of this thesis to the estimation of mrl functions are summarized as follows: Taking into account the natural ordering that may exist between mrl functions, we proposed nonparametric estimators (restricted estimators) of the bivariate mrl function under mrl ordering. That is, we have proposed estimators for two mrl functions, $e_1$ and $e_2$, under the order restriction that $e_1 \leq e_2$ when $e_2$ is known or unknown. We have proved that the restricted estimators are strongly uniformly consistent and asymptotically unbiased. We have also showed their weak convergence on $\mathbb{R}_+^2$. It has been shown that the restricted estimators are projections of the empirical mrl function onto an appropriate convex set of mrl functions. Simulation studies were carried out for various sample sizes and various bivariate distributions. The
simulation results indicate that the proposed estimators are superior to the empirical estimators in terms of mean squared error. Tests and confidence sets for the bivariate mrl function were also developed using the proposed estimators.

6.2 Extension to the Multivariate Case

Extensions of the results developed in Chapter 3 and Chapter 4 to the multivariate case are straight-forward. In what follows, we assume that \( \mathcal{T} = (T_1, T_2, \ldots, T_k) \) denote the survival times of \( k \) individuals (not necessarily independent). The corresponding distribution and survival functions of \( \mathcal{T} \) will be denoted by \( F(t_1, \ldots, t_k) \) and \( S(t_1, \ldots, t_k) \), respectively. Let us assume that \( E(T_i) < \infty \), for \( i = 1, \ldots, k \). This assumption guarantees the existence of the multivariate mrl function. In view of the “wear-out” phenomena in nature, it is reasonable to assume that a component has a finite mean lifetime. Let \( t_i = (t_1, \ldots, t_i-1, t_{i+1}, \ldots, t_k) \). Then the multivariate mrl function at a point \( t = (t_1, \ldots, t_k) \) is given by

\[
e(t_1, t_2, \ldots, t_k) = \{ e_1(t_1, \ldots, t_k), \ldots, e_k(t_1, \ldots, t_k) \} \tag{6.2.1}
\]

\[
= E[T - t | T > t],
\]

where the \( i \)th component is given by:

\[
e_i(t_1, \ldots, t_k) = E[T_i - t_i | T > t] \tag{6.2.2}
\]

\[
= \frac{\int_{t_i}^{\infty} S(t_1, \ldots, t_{i-1}, u, t_{i+1}, \ldots, t_k) du}{S(t_1, \ldots, t_k)}.
\]
Let \((T_{ij}, \ldots, T_{kj}), j = 1, 2, \ldots, n\) be i.i.d. random vectors in \(\mathbb{R}_+^k\) and let \(S_n(t_1, \ldots, t_k)\) be the multivariate empirical survivor function. Then the multivariate empirical mrl function is given by

\[
\hat{e}_n(t_1, t_2, \ldots, t_k) = \{\hat{e}_{1,n}(t_1, \ldots, t_k), \ldots, \hat{e}_{k,n}(t_1, \ldots, t_k)\}, \tag{6.2.3}
\]

where

\[
\hat{e}_{i,n}(t_1, \ldots, t_k) = \frac{\int_{t_i}^{\infty} S_n(t_1, \ldots, t_{i-1}, u, t_{i+1}, \ldots, t_k) \, du}{S_n(t_1, \ldots, t_k)} \tag{6.2.4}
\]

\[
= \frac{\sum_{j=1}^{n} \left( T_{ij} - t_i \right) I(T_{ij} > t_1, T_{ij} > t_2, \ldots, T_{ij} > t_k)}{\sum_{j=1}^{n} I(T_{ij} > t_1, T_{ij} > t_2, \ldots, T_{ij} > t_k)}.
\]

The expected value of \(\hat{e}_{i,n}(t_1, \ldots, t_k)\) was derived by Zahedi (1985):

\[
E[\hat{e}_{i,n}(t_1, \ldots, t_k)] = \{1 - [1 - S(t_1, \ldots, t_k)]^n\} e_i(t_1, \ldots, t_k). \tag{6.2.5}
\]

Thus, \(\hat{e}_{i,n}(t_1, \ldots, t_k)\) is asymptotically unbiased with bias decaying exponentially to zero as \(n \to \infty\). Moreover, one can show that \(\hat{e}_{i,n}(t_1, \ldots, t_k)\) is strongly uniformly consistent and converges weakly to a \(k\)-variate Gaussian process. The proof of these results, which follows from the arguments used to prove the strong consistency and weak convergence of the bivariate empirical mrl function, will appear in another paper.

Now let us consider the problem of nonparametric estimation of a multivariate mrl function when it is bounded by another known (unknown) multivariate mrl function. Let \((X_i, X_{i1}, \ldots, X_{ik}), i = 1, \ldots, n_1\) and \((Y_j, Y_{j1}, \ldots, Y_{jk}), j = 1, \ldots, n_2\), be random vectors in \(\mathbb{R}_+^k\) with finite means representing the lifetimes of two populations with survival functions \(S_1(x_1, \ldots, x_k)\) and \(S_2(y_1, \ldots, y_k)\); and mrl functions \(e_1(x_1, \ldots, x_k)\)
and \( e_2(y_1, \ldots, y_k) \). Let \( \mathcal{X} = \{(x_1, \ldots, x_k) : S_1(x_1, \ldots, x_k) > 0 \text{ and } S_2(x_1, \ldots, x_k) > 0\} \).

Case-1: Suppose that \( e_2(x_1, \ldots, x_k) \) is known and \( e_1(x_1, \ldots, x_k) \leq e_2(x_1, \ldots, x_k) \) for all \( (x_1, \ldots, x_k) \in \mathcal{X} \). Then a reasonable estimator of \( e_1(x_1, \ldots, x_k) \) under this order restriction is given by

\[
e^*_1(x_1, \ldots, x_k) = \min\{\hat{e}_{1,n}(x_1, \ldots, x_k), e_2(x_1, \ldots, x_k)\}.
\] (6.2.6)

Case-2: Suppose that \( e_2(x_1, \ldots, x_k) \) is known and \( e_1(x_1, \ldots, x_k) \geq e_2(x_1, \ldots, x_k) \) for all \( (x_1, \ldots, x_k) \in \mathcal{X} \). Then a reasonable estimator of \( e_1(x_1, \ldots, x_k) \) under this order restriction is given by

\[
e^*_1(x_1, \ldots, x_k) = \max\{\hat{e}_{1,n}(x_1, \ldots, x_k), e_2(x_1, \ldots, x_k)\}.
\] (6.2.7)

Case-3: Suppose that \( e_2(x_1, \ldots, x_k) \) is unknown and \( e_1(x_1, \ldots, x_k) \leq e_2(x_1, \ldots, x_k) \) for all \( (x_1, \ldots, x_k) \in \mathcal{X} \). Then estimators of \( e_1(x_1, \ldots, x_k) \) and \( e_2(x_1, \ldots, x_k) \) under this order restriction are given by

\[
e^*_1(x_1, \ldots, x_k) = \min\{\hat{e}_{1,n}(x_1, \ldots, x_k), \hat{e}_p(x_1, \ldots, x_k)\} \quad \text{and} \quad \hat{e}_p(x_1, \ldots, x_k) = \frac{n_1}{n_1 + n_2} \hat{e}_{1,n}(x_1, \ldots, x_k) + \frac{n_2}{n_1 + n_2} \hat{e}_{2,n}(x_1, \ldots, x_k).
\] (6.2.8)

\[
e^*_2(x_1, \ldots, x_k) = \max\{\hat{e}_{2,n}(x_1, \ldots, x_k), \hat{e}_p(x_1, \ldots, x_k)\}.
\] (6.2.9)

where \( \hat{e}_p(x_1, \ldots, x_k) \) is the pooled mrl function defined by

\[
\hat{e}_p(x_1, \ldots, x_k) = \frac{n_1}{n_1 + n_2} \hat{e}_{1,n}(x_1, \ldots, x_k) + \frac{n_2}{n_1 + n_2} \hat{e}_{2,n}(x_1, \ldots, x_k).
\]

A detailed investigation of the properties of the estimators defined in (6.2.6 - 6.2.9) requires further research and will appear in another paper.
Appendix A

Properties of The Kaplan-Meier Estimator

Lawless (1982) describes the area of survival analysis in the following way: "A useful way of portraying ungrouped univariate survival data is to compute and graph the empirical survivor function or, equivalently, the empirical distribution function. This also provides a nonparametric estimate of the survivor or distribution function for the life distribution under study. If there are no censored observations in a sample of size \( n \), the empirical survival function (ESF) is defined as

\[
\hat{S}_n(t) = \frac{\text{Number of observations} > t}{n}, \quad t \geq 0. \tag{A.0.1}
\]

This is a step function that decreases by \( 1/n \) just after each observed lifetime if all observations are distinct. More generally, if there are \( d \) lifetimes equal to \( t \), the ESF drops by \( d/n \) just past \( t \).
When dealing with survival data, some modifications of (A.0.1) are necessary, since the number of lifetimes greater than or equal to \( t \) will not generally be known exactly. The modification of (A.0.1) described in Chapter 1 is called the Kaplan-Meier (1958) estimate of the survivor function. The estimate is defined as follows: suppose that there are observations on \( n \) individuals and that there are \( k (k \leq n) \) distinct times \( t_1 < t_2 \ldots < t_k \) at which deaths occur. The possibility of there being more than one death at \( t_j \) is allowed, and we let \( d_j \) represent the number of deaths at \( t_j \). In addition to the lifetimes \( t_1 < t_2 \ldots < t_k \), there are also censoring times \( L_i \) for individuals whose lifetimes are not observed. Let us denote the distribution functions of the failure and censoring times by \( F(t) \) \( (S(t) = 1 - F(t)) \) and \( G(t) \) \( (\tilde{G}(t) = 1 - G(t)) \) respectively. Let \( \bar{H}(t) = 1 - H(t) = S(t)\tilde{G}(t) \). The Kaplan-Meier estimate of \( S(t) \) is defined as

\[
S_{km}(t) = \prod_{j: t_j \leq t} \frac{n_j - d_j}{n_j}
\]  

(A.0.2)

where \( n_j \) is the number of individuals at risk at \( t_j \), that is, the number of individuals alive and uncensored just prior to \( t_j \). If a censoring time \( L_i \) and a lifetime \( t_j \) are recorded as equal, we adopt the convention that censoring times are adjusted an infinitesimal amount to the right so that \( L_i \) is considered to be infinitesimally larger than \( t_j \). In other words, any individuals with censoring times recorded as equal to \( t_j \) are included in the set of \( n_j \) individuals at risk at \( t_j \), as are individuals who die at \( t_j \). This convention is sensible, since an individual censored at time \( L \) almost certainly survives past \( L \). Another point about (A.0.2) concerns situations in which
the largest observed time in the sample is a censoring time rather than a lifetime. In this case the Kaplan-Meier estimator is taken as being defined only up to this last observation”. By examining the formula for $S_{km}(t)$, we can immediately see that: (1) $S_{km}(0) = 1$, (2) $S_{km}(t)$ is monotone decreasing, (3) $S_{km}(t)$ is piecewise constant, (4) $S_{km}(t)$ has jump discontinuities at the uncensored observations, and (5) $S_{km}(t)$ has no jump discontinuities at the censored observations.

To effectively assess results when using K-M estimates it is desirable to have an estimate of the variance of $S_{km}(t)$. The variance of $S_{km}(t)$ is estimated by

$$\text{Var}[S_{km}(t)] = S_{km}^2(t) \sum_{j: t_j \leq t} \frac{d_j}{n_j(n_j - d_j)} \quad (A.0.3)$$

$$= S_{km}^2 \sum_{k=1}^{j} \frac{d_k}{n_k(n_k - d_k)} \quad \text{for } t_j \leq t < t_{j+1}.$$

The above formula is referred to as Greenwood’s formula. When there is no censoring, (A.0.3) reduces to $\hat{S}_n[1 - \hat{S}_n]/n$.

The Kaplan-Meier estimator possesses several desirable large-sample properties. A thorough study of the properties of Kaplan-Meier estimator has been done by several authors. We shall outline a few pertinent results and refer the reader to these papers for more details. The important properties are that $S_{km}(t)$ is a consistent estimate of $S(t)$ and after properly normalized $S_{km}(t)$ converges to a Gaussian process. Shorack and Wellner (1986) discuss the following result:

$$\sup_{0 \leq t < \tau} |S_{km}(t) - S(t)| \to 0 \text{ with probability } 1, \quad (A.0.4)$$

where $\tau = H^{-1}(1)$. Stute and Wang (1993) extended the above result to the interval
[0, \tau] under the assumption that both \( S(t) \) and \( \tilde{G}(t) \) do not have jumps in common.

Let \( T < \infty \) satisfy \( S(T) > 0 \) and suppose that \( S(x) \) and \( \tilde{G}(x) \) are continuous, then Breslow and Crowley (1974) showed that the process

\[
\{ \sqrt{n}[S_{km}(t) - S(t)], \ 0 < t < T \} \Rightarrow W(t), \tag{A.0.5}
\]

where \( W(t) \) is a Brownian motion with covariance function

\[
U(t) = S(t)^2 \int_0^t \frac{|dS(u)|}{S^2(u)\tilde{G}(u)} \ 0 < t < T. \tag{A.0.6}
\]

Li (1997) showed that

\[
\sqrt{n}\left[ \int_t^\infty S_{km}(x)dx - \int_t^\infty S(x)dx \right] \Rightarrow Z(t), \tag{A.0.7}
\]

where \( Z(t) = \int_t^\infty S(x)W(x)dx \) is a normal variate having mean zero and variance given by

\[
\sigma_Z^2(t) = \int_t^\infty \frac{[\int_x^\infty S(u)du]^2dF(x)}{S(x)\tilde{H}(x)}. \tag{A.0.8}
\]
Appendix B

Definition of Weak Convergence and Gaussian Processes

In this appendix, we will state some of the results which we have used to prove weak convergence. Moreover, we will provide definitions and properties of Gaussian processes.

B.1 Weak Convergence and Tightness

Let $S$ be an arbitrary metric space and $\mathcal{F}$ be the $\sigma$-field generated by Borel sets in $S$. Suppose that $P_n$ and $P$ are probability measures on $(S, \mathcal{F})$. If $P_n$ and $P$ satisfy that $\int_S f dP_n \to \int_S f dP$ for every bounded, continuous real function $f$ on $S$, we say that $P_n$ converges weakly to $P$ and write $P_n \Rightarrow P$. Note that, since the integrals $\int f dP$ completely determine $P$, the sequence $\{P_n\}$ cannot converge weakly to two different
limits. Note also that weak convergence depends only on the topology of $S$, not on the specific metric that generates it: Two metrics generating the same topology give rise to the same classes $\mathcal{Y}$ and $C(S)$ and hence to the same notion of weak convergence. $C(S)$ denotes the set of all bounded, continuous real functions on $S$.

Weak convergence in function space is proved by proving weak convergence of the finite-dimensional distributions and then proving tightness. The following notion of tightness (Billingsley 1968) proves important both in the theory of weak convergence and its application:

**Theorem B.1.1.** Let $\{P_n\}$ be a sequence of probability measures on $(C(S), \mathcal{Y})$. The sequence $\{P_n\}$ is tight if and only if these two conditions hold:

(i) For each positive $\eta$, there exists an $a$ such that

\[ P_n\{x : |x(0)| > a\} \leq \eta, \text{ for } n \geq 1. \]

(ii) For each positive $\epsilon$ and $\eta$, there exists a $\delta$, with $0 < \delta < 1$, and an integer $n_0$ such that

\[ P_n\{x : \sup_{|s-t|<\delta} |x(s) - x(t)| \geq \epsilon\} \leq \eta, \text{ for } n \geq n_0 \]

where $x \in C(S)$.

If $h$ is measurable mapping of $S$ into another metric space $S'$ (with $\sigma$-field $\mathcal{Y}'$ of Borel sets), then each probability measure $P$ on $(S, \mathcal{Y})$ induces on $(S', \mathcal{Y}')$ a unique probability measure $Ph^{-1}$, defined by $Ph^{-1}(A) = P(h^{-1}A)$ for $A \in \mathcal{Y}'$. We need
conditions under which \( P_n \Rightarrow P \) implies \( P_n h^{-1} \Rightarrow Ph^{-1} \). One such condition is that \( h \) be continuous, since then \( f(h(x)) \) is bounded and continuous on \( S \) whenever \( f(y) \) is bounded and continuous on \( S' \), so that \( P_n \Rightarrow P \) implies \( \int f(h(x)) dP_n(dx) \rightarrow \int f(h(x)) P(dx) \), a relation which, upon transformation of the integrals becomes, \( \int f(y) P_nh^{-1}(dy) \rightarrow \int f(y) Ph^{-1}(dy) \). We can weaken the continuity assumption. Assume only that \( h \) is measurable and let \( D_h \) be the set of discontinuities of \( h \) (\( D_h \in \mathcal{S} \)). Then if \( P_n \Rightarrow P \) and \( P(D_h) = 0 \), then \( P_nh^{-1} \Rightarrow Ph^{-1} \). The above results are summarized in the following Theorem:

**Theorem B.1.2.** Let \( h \) be a mapping of \( S \) into \( S' \) and let \( D_h \) be the set of discontinuities of \( h \).

(i) If \( h \) is a continuous mapping of \( S \) into \( S' \), then \( P_n \Rightarrow P \) implies \( P_nh^{-1} \Rightarrow Ph^{-1} \).

(ii) If is a measurable function with \( P(D_h) = 0 \), then \( P_n \Rightarrow P \) implies \( P_nh^{-1} \Rightarrow Ph^{-1} \).

The above results can be also stated in terms of random variables (functions) as it is given in the following Corollary:

**Corollary B.1.3.** Let \( X \) be a random element of \( S \), thus \( h(X) \) is a random element of \( S' \) (we still assume \( h \) is measurable).

(i) If \( X_n \rightarrow X \) and \( P\{X \in D_h\} = 0 \), then \( h(X_n) \Rightarrow h(X) \).

(ii) If \( X_n \rightarrow X \) and if \( h \) is a continuous mapping of \( S \) into \( S' \), then \( h(X_n) \Rightarrow h(X) \).
B.2 Gaussian Processes

A stochastic process $X = \{X(t), t \in T\}$ is a collection of random variables defined on the same probability space $(\Omega, \zeta, P)$. That is, for each $t$ in the index set $T$, $X(t)$ is a random variable. We often interpret $t$ as time and call $X(t)$ the state of the process at time $t$. If the index set $T$ is a countable set, we call $X$ a discrete-time stochastic process, and if $T$ is a continuum, we call it a continuous-time process. Any realization of $X$ is called a sample path.

A continuous-time stochastic process $\{X(t), t \in T\}$ is said to have independent increments if for any $n$ and all $t_0 \leq t_1 \leq \ldots \leq t_n$, the random variables $X(t_1) - X(t_0)$, $X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})$ are independent. It is said to possess stationary increments if, for any $s > 0$, $X(t + s) - X(t)$ has the same distribution for all $t$. That is, it possesses independent increments if the changes in the process’ value over nonoverlapping time intervals are independent; and it possesses stationary increments if the distribution of the change in value between any two points depends only on the distance between those points.

A stochastic process $B = \{B(t), t \geq 0\}$ is said to be a Brownian motion process if the following hold:

1. $B(0) = 0$.

2. $B$ has stationary and independent increments.

3. For any $c > 0$ and for $0 \leq s < t$, the increment $B(t) - B(s)$ is normally
distributed with mean 0 and variance $c^2(t-s)$.

It can be proved that for Brownian motion, the sample paths are continuous with probability one. When $c = 1$, $B(t)$ is called standard Brownian motion. If $\{B(t), t \geq 0\}$ is a standard Brownian motion, then the process $\{Z(t), 0 \leq t \leq 1\}$ where $Z(t) = B(t) - tB(1)$ is called Brownian Bridge process.

A stochastic process $\{X(t), t \geq 0\}$ is called a Gaussian process if $X(t_1), \ldots, X(t_n)$ has a multivariate normal distribution for all $t_1, \ldots, t_n$. Since a multivariate normal distribution is completely determined by the marginal mean values and the covariance values, it follows that Brownian motion could also be defined as a Gaussian process having $E[X(t)] = 0$ and, for $s \leq t$,

$$\text{Cov}[X(s), X(t)] = \text{Cov}[X(s), X(s) + X(t) - X(s)]$$

$$= \text{Cov}[X(s), X(s)] + \text{Cov}[X(s), X(t) - X(s)]$$

$$= c^2s \text{ for } c > 0,$$

where the last equality follows from independent increments and $\text{Var}[X(s)] = s$. 
Appendix C

Simulation Results for Chapter Three

In this section we will display some of the simulation results for the bivariate empirical mrl function which was defined in Chapter 3. The surface plots of the mean squared error for the empirical mrl function are shown for each of the four distributions from which we simulated. The results given here are for (a) a bivariate Gumbel distribution with density function

\[ f(x, y; \theta) = [(1 + \theta x)(1 + \theta y) - \theta] \exp(-x - y - \theta xy), \ x, y > 0 \]

for \( \theta = 1 \), (b) bivariate pareto distribution with density function

\[ f(x_1, x_2, \theta_1, \theta_2, a) = a(a + 1)(\theta_1 \theta_2)^{-(a+1)}(\theta_2 x_1 + \theta_1 x_2 - \theta_1 \theta_2)^{-(a+2)}, \ x_1, x_2 \geq 1, \]

for \( a = 10, \theta_1 = \theta_2 = 1 \), (c) bivariate Morgenstern (1956) distribution with density function

128
\[ f(u_1, u_2, \alpha) = 1 + \alpha(2u_1 - 1)(2u_2 - 1), \hspace{1em} 0 \leq u_1, u_2 \leq 1, \]

for \( \alpha = 0.5 \) and \((d)\) bivariate Sarmanov (1966) distribution with density function

\[ f(x_1, x_2, \alpha) = e^{-x_1}e^{-x_2} \{1 + \alpha(2e^{-x_1} - 1)(2e^{-x_2} - 1)\}, \hspace{1em} 0 \leq x_1, x_2 < \infty, \]

for \( \alpha = 0.6 \) and for various sample sizes, \( n = 15, 30, \) and \( 45. \)

---

Figure C.1: Mean squared error for the empirical mrl function in the case of bivariate Gumbel distribution \( (\theta = 1). \)
Figure C.2: Mean squared error for the empirical mrl function in the case of Morgenstern distribution ($\alpha = 0.5$).
Figure C.3: Mean squared error for the empirical mrl function in the case of Sarmanov distribution ($\alpha = 0.6$).
Appendix D

Simulation Results for Chapter Four

In this section we will display some of the simulation results for the bivariate restricted mrl function which was defined in Chapter 4. The surface plots of the ratio the mean squared errors of the restricted estimator to the empirical estimator are shown for each of the four distributions from which we simulated and for the three cases which we have considered. The results given here are for various sample sizes, \( n = 15, 30, \) and 45 and for a bivariate distribution with density function

\[ f_i(x_1, x_2, \theta_i, a_i, \alpha_i), \text{ for } i = 1, 2 \]

where \( \theta_i, a_i, \) and \( \alpha_i \) are unknown parameters; \( (x_1, x_2) > 0 \) for the bivariate Gumbel, pareto, and Sarmanov (1966) distributions and \( 0 \leq x_1, x_2 \leq 1 \) for the Morgenstern (1956) distribution. The values of \( \theta_1 \) and \( \theta_2 \) are chosen so that \( e_1(x_1, x_2) \leq e_2(x_1, x_2) \)
for all \((x_1, x_2) \in \mathbb{R}_+^2\). Thus, in the bivariate Gumbel we used \(\theta_1 = 1\) and \(\theta_2 = 0.5\). In the pareto we choose \(\alpha_1 = 10\) and \(\alpha_2 = 5\), for both the Morgenstern (1956) and Sarmanov (1966) distributions we used \(\alpha_1 = 0.5\) and \(\alpha_2 = 0.6\).

\[\text{Figure D.1: Ratios of mean squared errors of the restricted estimator to the empirical estimator for two Gumbel distributions (}\theta_1 = 1, \theta_2 = 0.5\): (Left)-one sample problem, (Right)-mirror image problem.\]
Figure D.2: Ratios of mean squared errors of the restricted estimator to the empirical estimator for two Morgenstern distributions ($\alpha_1 = 0.5, \alpha_2 = 0.6$): (Left)-one sample problem, (Right)-mirror image problem.
Figure D.3: Ratios of mean squared errors of the restricted estimator to the empirical estimator for two Sarmanov distributions ($\alpha_1 = 0.5, \alpha_2 = 0.6$): (Left)-one sample problem, (Right)-mirror image problem.
Figure D.4: Two sample problem: Ratios of mean squared errors of the restricted estimator to the empirical estimator for two Gumbel distributions ($\theta_1 = 1, \theta_2 = 0.5$): (Left)-minimum estimator, (Right)-maximum estimator.
Figure D.5: Two Sample Problem: Ratios of mean squared errors of the restricted estimator to the empirical estimator for two Morgenstern distributions ($a_1 = 10, a_2 = 5$): (Left)-minimum estimator, (Right)-maximum estimator.
Figure D.6: Two Sample Problem: Ratios of mean squared errors of the restricted estimator to the empirical estimator for two Sarmanov distributions ($a_1 = 10, a_2 = 5$): (Left)-minimum estimator, (Right)-maximum estimator.
Bibliography


