1994

Life-table analysis for correlated response times

Shin-Soo Kang
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Biostatistics Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/11271

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

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 bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI 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.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600
Life-table analysis for correlated response times

by

Shin-Soo Kang

A Dissertation Submitted to the Graduate Faculty in Partial Fulfillment of the Requirements for the Degree of DOCTOR OF PHILOSOPHY

Department: Statistics
Major: Statistics

Approved:

In Charge of Major Work

For the Major Department

For the Graduate College

Iowa State University
Ames, Iowa
1994

Copyright © Shin-Soo Kang, 1994. All rights reserved.
Life-table analysis for
correlated response times

by

Shin-Soo Kang

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Statistics
Major: Statistics

Iowa State University
Ames, Iowa
1994

Copyright © Shin-Soo Kang, 1994. All rights reserved.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ....................................... viii

CHAPTER 1. INTRODUCTION ................................. 1

CHAPTER 2. MULTIVARIATE LIFE-TABLE ANALYSIS ............ 7
  2.1 Random Censorship Model .............................. 8
  2.1.1 Statistical model ...................................... 8
  2.1.2 Life-table quantities ............................... 9
  2.2 PL Estimate .................................................. 10
  2.2.1 Definition .............................................. 10
  2.2.2 Large sample properties of \( \hat{P}_h \) .............. 11
  2.2.3 Variance of the PL estimator ....................... 18
  2.3 AL Estimate ................................................. 24
  2.4 Flexibility for General Multivariate Response Times .. 25
  2.5 Example .................................................... 26
    2.5.1 Description of the angioplasty study ............... 26
    2.5.2 Life-table analysis ................................. 27

CHAPTER 3. MAXIMUM LIKELIHOOD ESTIMATION OF SURVIVAL PROBABILITIES ........................................... 31
  3.1 Likelihood Function ..................................... 31
3.1.1 Univariate case ........................................... 31
3.1.2 Bivariate case ............................................ 32

3.2 Likelihood Equations ....................................... 35
3.3 Existence and Uniqueness of MLE .......................... 39

CHAPTER 4. SIMULATION STUDY ............................... 44
4.1 Simulation of Bivariate Survival Data ...................... 44
   4.1.1 Generating bivariate failure times with Weibull marginal distribution ....................... 46
   4.1.2 Generating bivariate censoring times with uniform marginal distribution .................. 48
   4.1.3 Simulated data tables ................................... 51

4.2 Procedure .................................................. 54
   4.2.1 PL estimate ............................................ 54
   4.2.2 AL estimate ........................................... 56
   4.2.3 Maximum likelihood estimation ......................... 57

4.3 Results .................................................... 58
   4.3.1 Bias for the true survival probabilities ................. 59
   4.3.2 Compare $VA_h$ to $A_{vu}A_h$ and $VP_h$ to $A_{vu}P_h$ ........................................ 61
   4.3.3 Compare $VA_h$ to $A_{vm}A_h$ and $VP_h$ to $A_{vm}P_h$ ........................................ 62
   4.3.4 Relative efficiency of the MLE and PL estimators .......... 62

4.4 Summary ................................................... 63

CHAPTER 5. TESTS FOR EQUALITY OF TWO DISTRIBUTIONS WITH LIFE-TABLE MODEL .................. 71
5.1 Introduction ................................................ 71
5.2 Univariate Independent Response Times .......................... 73
5.3 Correlated Response Times where Treatments are applied to Different
Independent Sets of Cohorts ........................................... 77
5.4 Example for Section 5.3 .............................................. 79
5.5 Correlated Response Times where Different Treatments are Applied
within Individual Cohorts .............................................. 81

CHAPTER 6. CONCLUSION .................................................. 87

BIBLIOGRAPHY .............................................................. 89

APPENDIX A. THE FOURTEEN FORMULAS FOR THE SECOND PARTIAL DERIVATIVES OF THE LOGLIKELIHOOD
FUNCTION, $\frac{\partial^2 \Lambda}{\partial \theta_{wx} \partial \theta_{yz}}$ .............................................. 95

APPENDIX B. S-PLUS PROGRAMS FOR MAXIMUM LIKELIHOOD ESTIMATOR .............................................. 102

APPENDIX C. S-PLUS PROGRAM FOR THE LIFE TABLE ANALYSIS OF CORRELATED RESPONSES ...................... 126

APPENDIX D. ANGIOPLASTY DATA .......................................... 136
LIST OF TABLES

Table 2.1: Life-table for PL computed from angioplasty data .......... 28
Table 2.2: Life-table for AL computed from angioplasty data .......... 29
Table 4.1: Comparing \( \text{Corr}(X_1, X_2) \) and \( \text{Corr}(Y_1, Y_2) = w^2 \) .......... 47
Table 4.2: Comparison of \( \text{Corr}(W_1, W_2) \) and \( \text{Corr}(Y_1, Y_2) = w^2 \) .......... 50
Table 4.3: Comparing \( mB(x) \times 10^3 \) for 3 estimates ............... 65
Table 4.4: Comparing \( mB^2(x) \times 10^6 \) for 3 estimates ............... 66
Table 4.5: Comparison of \( VA_h \) to \( AvuA_h \) and \( VP_h \) to \( AvuP_h \) .......... 67
Table 4.6: Comparison of \( VA_h \) to \( AvmA_h \) and \( VP_h \) to \( AvmP_h \) .......... 68
Table 4.7: Relative efficiency of the MLE to the PL estimator .......... 69
Table 4.8: Comparison of true standard deviation of PL to AL .......... 70
Table 5.1: \( \chi^2 \) test for equality of two survivor functions using PL estimates .......... 80
Table 5.2: \( \chi^2 \) test for equality of two survivor functions using AL estimates ......................................................... 80
Table D.1: Angioplasty Data .............................................. 137
LIST OF FIGURES

Figure 2.1: Comparison of PL and AL survival curves for angioplasty data 30

Figure 4.1: Comparison of $\text{Corr}(X_1, X_2)$ and $\text{Corr}(Y_1, Y_2) = w^2$ 48

Figure 4.2: Accuracy of approximation (4.5) for the censoring probability when $\beta = 1$ and $\theta = 100.$ 50

Figure 4.3: Comparison of $\text{Corr}(W_1, W_2)$ and $\text{Corr}(Y_1, Y_2) = w^2$ 51

Figure 4.4: Survival probability when $\beta = 1$ and $\theta = 100.$ 52

Figure 4.5: Survival probability when $\beta = 3$ and $\theta = 100.$ 53

Figure 5.1: Comparison of PL survival curves for males and females 82

Figure 5.2: Comparison of PL survival curves for presence and absence of diabetes mellitus 82

Figure 5.3: Comparison of PL survival curves for presence and absence of myocardial infarction 83

Figure 5.4: Comparison of PL survival curves for presence and absence of hypertension 83

Figure 5.5: Comparison of AL survival curves for males and females 84

Figure 5.6: Comparison of AL survival curves for presence and absence of diabetes mellitus 84
Figure 5.7: Comparison of AL survival curves for presence and absence of myocardial infarction 85
Figure 5.8: Comparison of PL survival curves for presence and absence of hypertension 85
I would like to express my deepest gratitude to Professor Kenneth J. Koehler for his eagerness, encouragement, enthusiasm, and patience throughout the course of this research. My appreciation also goes to Professor William Q. Meeker, Professor Herbert T. David, Professor James L. Cornette, and Professor Sarah M. Nusser for their help, time, and kindness in serving on my program of study committee. Finally, but most importantly, I would like to appreciate my family, Sunhwa, Daegun and Minseon Lee, for their continuous support and encouragement during the studies in Iowa State University.
CHAPTER 1. INTRODUCTION

Survival data consisting of independent groups of correlated response times arise from a variety of situations, such as event times collected from husband and wife pairs, siblings, litter mates, distinct components of a machine, or repeated measurements on each individual subject. In this dissertation, we restrict our attention to situations where response times within groups are not distinguishable, and the marginal survival distributions are same for all individuals within any group. Furthermore, the data are interval censored, so exact event times are not observed, only the number of failures and number of censored individuals are observed within a finite set of time intervals. We are interested in estimating marginal survival probabilities and their variances and covariances from the observed counts. In this situation, distribution-free methods are developed for situations involving independent groups of correlated response times. This will be referred to as “life-table analysis for correlated response times”.

The life-table analysis has been used widely to summarize failure time or event time data without assuming any specific parametric distributions for response times. Kaplan and Meier (1958) introduced the product limit (PL) and the actuarial (AL) estimates of survival probabilities for univariate analysis with independent failure times. The PL estimate is consistent under the assumption that censoring only
occurs at the ends of time intervals. Breslow and Crowley (1974) showed that a necessary and sufficient condition for the consistency of the AL estimate of a survival probability is that $F^0$ satisfy

$$F^0(t) = 1 - \left[1/(1 + cH(t))\right]^{5}, \tag{1.1}$$

for some constant $c > 0$, where $H(t)$ is a cdf of iid censoring times and absolutely continuous with density $h(t)$ on a finite time interval, and $F^0(t)$ is a cdf of iid failure times. They suggest that the uniform distribution is a good approximation for the censoring distribution in many cases that satisfy (1.1).

The Greenwood (1926) formula is commonly used to get the variance of the PL and AL estimates of survival probabilities computed from independent response times. Suppose we have time intervals $A_h = (t_{h-1}, t_h]$, for $h = 1, 2, \ldots, m$. Then, define

$$q_h = \Pr(\text{an individual dies in } A_h \mid \text{he survives beyond } A_{h-1}) \quad \text{and}$$

$$P_h = \Pr(\text{an individual survives beyond } A_h).$$

Then the Greenwood formula for estimating the variance of $\hat{P}_h = (1 - \hat{q}_1)(1 - \hat{q}_2) \cdots (1 - \hat{q}_h)$ is

$$\text{Var}(\hat{P}_h) = \hat{P}_h^2 \sum_{i=1}^{h} \frac{\hat{q}_i}{N_i(1 - \hat{q}_i)}, \tag{1.2}$$

where $N_h$ is the number of observations "at risk" at time $t_{h-1}$, and $\hat{q}_h$ is an estimate of $q_h$. The PL estimator, for example, uses $\hat{q}_h = D_h/N_h$, where $D_h$ is the observed number of failures in $A_h$. The Greenwood formula is based on the assumptions that $\text{Cov}(\hat{q}_i, \hat{q}_j) = 0$, for $i < j$ and $\text{Var}(q_h) = q^2 h(1-q^2) / N_h$. Lawless (1982) suggested that (1.2) is a reasonable estimate of $\text{Var}(\hat{P}_h)$ when $E(N_h)$ is not too small and the number of time intervals is also not too small. When the data contain correlated
response times, we use the same life-table procedure to get estimates of the marginal survival probabilities that was described above for the case when all response times are independent. This provides appropriate estimates of survival probabilities, but (1.2) is no longer appropriate because the assumption that $\text{Cov}(\hat{g}_i, \hat{g}_j) = 0$ and $\text{Var}(\hat{g}_h) = \frac{q_h(1-q_h)}{N_h}$ are violated. In Chapter 2, we derive a modification of the Greenwood formula to appropriately account for within group correlations among response times.

Turnbull's (1974) non-parametric likelihood function is based on the product limit estimate with censored observations on the left and some on the right. He introduced an algorithm to get a "self-consistent estimator" and showed that the estimator is unique consistent maximum likelihood estimate under fairly general assumptions with existence of failures during any time interval. A concept of a "self-consistent estimator" was defined by Efron (1967). Campbell (1981) and Hanley and Parnes (1983) studied non-parametric maximum likelihood estimation for a bivariate survival function when the response times are interval censored. These methods assume that there is a clear distinction between each member of a pair, such as male and female siblings, right and left eyes. Campbell (1981) showed that the resulting maximum likelihood estimator is a self-consistent estimator. He also examined the existence and uniqueness of the estimator, and showed that the matrix of the second partial derivatives of the loglikelihood function is non-positive definite. Consequently, the loglikelihood is unimodal and the MLE is unique up to possible flat spots.

The consistency of the PL estimator was studied by Winter et al. (1978), Földes et al. (1980), and Földes and Rejtő (1981) for the univariate case with exact failure times. Winter et al. (1978) and Földes et al. (1980) proved the uniform consistency
of the PL estimator with ratio $o\left(\sqrt{\log n}/n^{1/4}\right)$ without imposing any continuity conditions on either the survival function for the failure time or the survivor function for the censoring times. Földes and Rejtő (1981) showed that uniformly almost sure consistency with rate $O(\sqrt{\log n}/\sqrt{n})$ assuming that the survival functions of failure times and censoring times are continuous, where $n$ is total number of responses. Horvath (1983) showed the consistency of a the multivariate PL estimator, computed from multivariate exact failure times, under the assumption that the joint survival function of multivariate failure times is continuous.

There is another approach based on counting processes, to prove consistency of the PL estimator. Fleming and Harrington (1991), and Andersen et al. (1993) considered $N(t)$, the number of observed failures in $[0, t]$, as univariate counting process and $Y(t)$ is the number at risk just prior to time $t$. They assumed that survivor function of a failure time random variable $T$ is absolutely continuous and the number at risk $Y(t)$ converges to $\infty$ as $n$ goes to $\infty$ to prove uniformly consistency of PL estimator using the Lenglart (1977) inequality.

Tsai and Crowley (1986) proposed a family of estimators of a bivariate survival function assuming that the survival function for the failure times and the survivor function for the censoring times are both absolutely continuous. These estimators are uniformly consistent under bivariate censoring and are self-consistent under univariate censoring. They showed that the estimators are always survival functions. Dabrowska (1988) proposed a new estimator of a multivariate survival function with a faster rate of convergence than the Tsai et al. (1986) estimator. Pruitt (1991) identified some cases in which negative mass is assigned by Dabrowska's estimator. Consequently, Dabrowska's estimator does not necessarily provide a proper multi-
variate survivor function.

Independence of failures and censoring was assumed in all papers reviewed above. This assumption is also used throughout this dissertation. In Chapter 2, PL and AL estimates that ignore correlations within groups are considered as point estimates of marginal survival probabilities, and a modification of the Greenwood formula is derived to estimate their variances. This approach does not require the estimation of a bivariate survival function, but we must estimate bivariate conditional probabilities of the form

\[ \Pr(X_{ij} \in (t_{h-1}, t_h], X_{ij'} \in (t_{h-1}, t_h] \mid X_{ij} > t_{h-1}, X_{ij'} > t_{h-1}), \]

where \( X_{ij} \) is the failure time for the \( j^{th} \) unit in \( i^{th} \) group. Furthermore, large sample properties of the PL and AL estimates are examined.

Maximum likelihood estimates are developed in Chapter 3 under the same assumption for the censoring mechanism as used for the PL estimates. Maximum likelihood estimates are more difficult to compute than the life-table estimates considered in Chapter 2 and they are not easily extended beyond the bivariate case, but they provide a reference for assessing the efficiency of the estimators considered in Chapter 2. The bias and relative efficiency of the various estimators are compared through simulation studies in Chapter 4. The effects of strength of correlation between responses within groups, the level of censoring, the number of time intervals, and sample size are examined. Estimates of variances from the Greenwood formula and the modified Greenwood formula are compared. Differences between these variance estimates depend mostly on the correlation between responses within groups.

In Chapter 5, we review tests for the equality of two survival distributions given independent and univariate event times. Then a test is developed for the situation
involving independent groups of correlated response times where entire groups are randomly assigned to different treatments.

The proposed methods are illustrated with a life-table analysis of survival times for angioplasty procedures used to reduce obstructions to blood flow in the legs of patients treated at the Iowa Heart Center in Des Moines, Iowa between 1987 and 1991. The analysis includes 159 angioplasty procedures performed on 115 patients. Different patients are assumed to respond independently of each other, but failure times for procedures performed on the same patient are allowed to be correlated.
CHAPTER 2. MULTIVARIATE LIFE-TABLE ANALYSIS

Life-table analysis is one of the oldest statistical methods used to analyze survival data and it is widely used in medical, actuarial, and industrial reliability studies. Little consideration, however, has been given to multivariate life-table analysis.

In this chapter, we consider the situation where there are groups, or cohorts of correlated event times, but an event time from one group is independent of any event time from any other group. This situation arises, for example, when time to appearance of a certain type of tumor is monitored for mice from the same litter, or time to restitution is monitored for more than one angioplasty procedure per patient.

We assume members of a group are not distinguishable from each other, and the marginal event time distribution is the same for each response. Under this situation, we consider the problem of estimating the common marginal survival probabilities through a non-parametric methods using life-tables. Estimates of variances and covariances of the estimated survival probabilities are obtained from an extension of the well known Greenwood formula. S-plus functions for computing life-table estimates are displayed in Appendix C.
2.1 Random Censorship Model

2.1.1 Statistical model

The bivariate case is mainly considered in this chapter for notational convenience, but more general multivariate cases involving groups with more than two response times, or groups with different numbers of response times can be derived in a similar manner.

Let \( \{(X_{i1}^0, X_{i2}^0)\} \), for \( i = 1, 2, \ldots, n \), be independent pairs of true failure times with joint survival distribution \( S(s, t) = \Pr(X_{i1}^0 > s, X_{i2}^0 > t) \). Let \( \{(W_{i1}, W_{i2})\} \), for \( i = 1, 2, \ldots, n \), be independent pairs of censoring variables from the joint distribution \( C(s, t) = \Pr(W_{i1} > s, W_{i2} > t) \). The variables \( X_{i1}, X_{i2}, \delta_{i1}, \delta_{i2} \) are observed, where

\[
X_{i1} = \min(X_{i1}^0, W_{i1}), \quad X_{i2} = \min(X_{i2}^0, W_{i2}),
\]

and

\[
\delta_{ij} = \begin{cases} 
1 & \text{if } X_{ij} = X_{ij}^0 \\
0 & \text{if } X_{ij} = W_{ij}, \text{ where } j = 1, 2.
\end{cases}
\]

Let \( G(s, t) = \Pr(X_{i1} > s, X_{i2} > t) \) be the joint distribution of \((X_{i1}, X_{i2})\). We will assume that \( \{(X_{i1}^0, X_{i2}^0)\} \) is independent of \((W_{i1}, W_{i2})\). Most procedures for life-table analysis are based on this assumption.

This bivariate model was previously considered by Campbell (1981), Dabrowska (1988), and Pruitt (1991). Campbell studied non-parametric maximum likelihood estimation for the situation where there is a clear distinction between the first member and second member of each pair. Dabrowska introduced a bivariate analogue of the Kaplan-Meier estimator, but Pruitt describes conditions under which this estimator does not yield a proper survival function. Clayton (1978) and Clayton and Cuzick
(1985) used bivariate life-tables merely to estimate the association parameter in a bivariate proportional hazards model for disease incidence in ordered pairs of individuals. Their bivariate model is a frailty model derived from certain assumptions about conditional hazards.

2.1.2 Life-table quantities

Consider a study consisting of $2n$ subjects, where failure times are subject to right censoring. Furthermore, consider the situation where each subject is inspected at a finite set of $m$ times $0 < t_1 < t_2 < \cdots < t_m < \infty$, so exact response times are not observed. Instead, the responses are interval censored, i.e., it is only known whether a subject failed or was censored between two adjacent inspection times. Denote the $m$ time intervals by $A_h = (t_{h-1}, t_h], h = 1, 2, \cdots, m$, where $t_0 = 0$. An individual is said to be "at risk" at time $t_h$ if the event has not yet occurred by time $t_h$ and the individual was not censored before time $t_h$. The quantities used to construct a life-table are:

\begin{align*}
N_h &= \text{Number of observations "at risk" at time } t_{h-1} \\
&= \sum_{i=1}^{n} \{I(X_{i1} > t_{h-1}) + I(X_{i2} > t_{h-1})\} \\
D_h &= \text{Number of failures in } A_h = (t_{h-1}, t_h] \\
&= \sum_{i=1}^{n} \{I(X_{i1} \in A_h, \delta_{i1} = 1) + I(X_{i2} \in A_h, \delta_{i2} = 1)\} \\
C_h &= \text{Number of withdrawals in } A_h = (t_{h-1}, t_h] \\
&= \sum_{i=1}^{n} \{I(X_{i1} \in A_h, \delta_{i1} = 0) + I(X_{i2} \in A_h, \delta_{i2} = 0)\},
\end{align*}

where $I(x)$ is an indicator function, which is 1 if $x$ is true, and zero otherwise.
2.2 PL Estimate

2.2.1 Definition

The conditional failure probability in time interval \( A_h \) is

\[ q_h = \text{Pr( an individual dies in } A_h \mid \text{ he survives beyond } A_{h-1} ) \]

the unconditional survival probability is

\[ P_h = \text{Pr( an individual survives beyond } A_h ) \].

Kaplan and Meier (1958) first studied the properties of the PL(product limit) estimator and the AL(actuarial) estimator of \( P_h \) in the univariate case. The fundamental papers of Kaplan and Meier (1958) and Chiang (1961), Gilbert (1962), Efron (1967), Breslow (1969, 1970), Thomas (1972), Breslow and Crowley (1974) contributed to the development of the theoretical properties of these estimators. Breslow and Crowley (1974) outlined a general theory for the large sample properties in the univariate case where any subject is assumed to respond independently of any other subject.

We will consider properties of the PL and AL estimators for \( P_h \) when the study consists of \( n \) pairs of subjects where each subject responds independently of any other subject from any other pair, but responses for subjects in the same pair can be correlated. The PL estimator ignoring pairs is defined as

\[ \hat{P}_h = (1 - \hat{q}_1)(1 - \hat{q}_2) \cdots (1 - \hat{q}_h), \]

where \( \hat{q}_h = \frac{D_h}{N_h} \). This will be called the "PL estimator ignoring pairs or groups" or just the "PL estimator". In the next section the large sample properties of \( \hat{P}_h \) will be established.
2.2.2 Large sample properties of $\hat{P}_h$

Let $X_{i1}, i = 1, 2, \cdots, n$ be $n$ independent and identically distributed life times with survivor function $S(t)$, and let $X_{i2}, i = 1, 2, \cdots, n$ be iid life times with the same survival function $S(t)$. However $X_{i1}$ is not necessary independent of $X_{i2}$. Properties of $\hat{P}_h$ will be examined for both the no censoring case and the censoring case.

2.2.2.1 The case with no censored observations. Establishing the consistency of $\hat{P}_h$ is straightforward when there is no censoring. Define

$$f = (f_{11}, f_{12}, \cdots, f_{1m}, f_{21}, \cdots, f_{2m}, f_{31}, \cdots, f_{3m}, \cdots, f_{m1}, \cdots, f_{mm}),$$

$$\vartheta = (\vartheta_{11}, \vartheta_{12}, \cdots, \vartheta_{1m}, \vartheta_{21}, \cdots, \vartheta_{2m}, \vartheta_{31}, \cdots, \vartheta_{3m}, \cdots, \vartheta_{m1}, \cdots, \vartheta_{mm}),$$

where $f_{hk}$ is the number of pairs where the first unit is observed to fail during time interval $(t_{h-1}, t_h]$ and the second unit is observed to fail during time interval $(t_{h-1}, t_h]$, and $\vartheta_{hk}$ is the joint failure probability corresponding to $f_{hk}$. Then, $f$ has a multinomial distribution with parameters $n$ and $\vartheta$. It follows that as $n \to \infty$, the distribution of $\sqrt{n^{-1}(f - n\vartheta)}$ converges to a multivariate normal distribution with mean 0 and covariance matrix, $\text{diag}(\vartheta) - \vartheta\vartheta'$.

Now define the following quantities:

$$f_1 = (u_1, u_2, u_3, \cdots, u_m),$$

where $u_h = \sum_{j=1}^{m} f_{hj}$ is the number of failures during time interval $(t_{h-1}, t_h]$ experienced by the subjects listed first in the pairs, and

$$f_2 = (v_1, v_2, v_3, \cdots, v_m),$$
where \( v_h = \sum_{i=1}^{m} f_{ih} \) is the number of failures experienced during time interval \((t_{h-1}, t_h]\) for the subjects listed second in the pairs. Also define a vector of marginal failure probabilities as

\[ \pi = (\pi_1, \pi_2, \cdots, \pi_m), \]

where

\[ \pi_h = \int_{t_{h-1}}^{t_h} \frac{dS(x)}{dS(x) \mid = \Pr (\text{an individual dies in } (t_{h-1}, t_h]) \]

is the failure probability for time interval \((t_{h-1}, t_h]\). These quantities can be expressed as linear functions of \( \beta \) as follows:

\[
\begin{align*}
\beta_1 &= I_{m \times m} \otimes 1^T \beta = A \beta, \\
\beta_2 &= 1^T \otimes I_{m \times m} \beta = B \beta,
\end{align*}
\]

where

\[
A_{m \times m^2} = \begin{pmatrix}
11 \cdots 1 & 00 \cdots 0 & 00 \cdots 0 \\
00 \cdots 0 & 11 \cdots 1 & 00 \cdots 0 \\
\vdots & \vdots & \vdots \\
00 \cdots 0 & 00 \cdots 0 & 00 \cdots 0 \\
\end{pmatrix},
\]

\[
B_{m \times m^2} = \begin{pmatrix}
100 \cdots 0 & 100 \cdots 0 & 100 \cdots 0 \\
010 \cdots 0 & 010 \cdots 0 & 010 \cdots 0 \\
\vdots & \vdots & \vdots \\
000 \cdots 1 & 000 \cdots 1 & 000 \cdots 1
\end{pmatrix}.
\]

Since either member of the pair could be arbitrarily designated as the first member, the restriction, \( \tau = A \theta = B \theta \) is imposed. Let \((\beta_1^*, \beta_2^*)\) be \( f^* \) and \((\tau^*, \pi^*)\) be
\( \pi^* \). It follows that as \( n \to \infty \), the distribution of \( \sqrt{n}^{-1}(f^* - n\pi^*) \) converges to a multivariate normal distribution with mean 0 and covariance matrix,

\[
\Sigma = \begin{bmatrix} A \\ B \end{bmatrix} (\text{diag}(\vartheta) - \vartheta \vartheta)' \begin{bmatrix} A' & B' \end{bmatrix}
\]

\[
= \begin{bmatrix} \text{diag}(\pi) - \pi \pi' \\ \text{diag}(\vartheta) - \pi \pi' \end{bmatrix} \begin{bmatrix} \text{diag}(\pi) - \pi \pi' \\ \text{diag}(\vartheta) - \pi \pi' \end{bmatrix}
\]

(2.1)

Now consider \( \hat{q}_h = \frac{u_h + v_h}{N_h} \), where \( N_h = \sum_{i=h}^{m} u_i + \sum_{i=h}^{m} v_i \) is the number of individuals “at risk”. Each \( \hat{q}_j \) is a smooth function of \( f^* \). Thus \( \sqrt{n}(\hat{q} - q^*) \sim N(0, n \Sigma_{q^*}) \) as \( n \to \infty \) and \( \Sigma_{q^*} \) can be obtained from the Delta method.

It immediately follows that \( \hat{q}_h \) is a consistent estimate of

\[
q_h^* = \frac{\pi_h/2 + \pi_i/2}{\sum_{i=h}^{m} \pi_i/2 + \sum_{i=h}^{m} \pi_i/2} = \frac{\pi_h}{\sum_{i=h}^{m} \pi_i} = \frac{S(t_{h-1}) - S(t_h)}{S(t_{h-1})},
\]

where \( S(t_h) \) is the true unknown value of the underlying marginal survivor function. Finally, \( \hat{P}_h \) is also a consistent estimate of \( P_h \) because \( \hat{P}_h \) is a continuous function of \( q_h^* \). Let us now turn to the censored case, which is more complicated than the no censoring case.

2.2.2.2 The case with censored observations. Let \( X_{i1}, i = 1, 2, \ldots, n \) be \( n \) independent and identically distributed (iid) failure times with survivor function \( S(t) \) and let \( X_{i2}, i = 1, 2, \ldots, n \) be iid failure times with the same survivor function \( S(t) \). \( X_{i1} \) and \( X_{i2} \) need not be independent. Let \( W_{i1}, i = 1, 2, \ldots, n \) be \( n \) independent and identically distributed censoring times with survivor function \( C(t) \), and let \( W_{i2}, i = 1, 2, \ldots, n \) be iid censoring times with the same survivor function \( C(t) \). Independence
of \( W_{i1} \) and \( W_{i2} \) is not necessary but it is assumed that the censoring mechanism does not affect the true life times. Define

\[
f = (f_{11}, f_{12}, \ldots, f_{1m}, f_{21}, \ldots, f_{2m}, f_{31}, \ldots, f_{3m}, \ldots, f_{m1}, \ldots, f_{mm})
\]

and

\[
\vartheta = (\vartheta_{11}, \vartheta_{12}, \ldots, \vartheta_{1m}, \vartheta_{21}, \ldots, \vartheta_{2m}, \vartheta_{31}, \ldots, \vartheta_{3m}, \ldots, \vartheta_{m1}, \ldots, \vartheta_{mm}),
\]

where \( f_{hh} \) is the number of pairs such that the first unit is observed to fail during time interval \((t_{h-1}, t_h]\) and the second unit is observed to fail during time interval \((t_{h-1}, t_h]\), and \( \vartheta_{hh} \) is the joint failure probability corresponding to \( f_{hh} \). Define

\[
d = (d_{11}, d_{12}, \ldots, d_{1m}, d_{21}, \ldots, d_{2m}, d_{31}, \ldots, d_{3m}, \ldots, d_{m1}, \ldots, d_{mm})
\]

and

\[
\nu = (\nu_{11}, \nu_{12}, \ldots, \nu_{1m}, \nu_{21}, \ldots, \nu_{2m}, \nu_{31}, \ldots, \nu_{3m}, \ldots, \nu_{m1}, \ldots, \nu_{mm}),
\]

where \( d_{hh} \) is the number of pairs such that the first unit is observed to fail during time interval \((t_{h-1}, t_h]\) and the second unit is censored during time interval \((t_{h-1}, t_h]\), and \( \nu_{hh} \) is the joint probability corresponding to \( d_{hh} \). Define

\[
c = (c_{11}, c_{12}, \ldots, c_{1m}, c_{21}, \ldots, c_{2m}, c_{31}, \ldots, c_{3m}, \ldots, c_{m1}, \ldots, c_{mm})
\]

and

\[
e = (e_{11}, e_{12}, \ldots, e_{1m}, e_{21}, \ldots, e_{2m}, e_{31}, \ldots, e_{3m}, \ldots, e_{m1}, \ldots, e_{mm}),
\]

where \( c_{hh} \) is the number of pairs such that the first unit is censored during time interval \((t_{h-1}, t_h]\) and the second unit is observed to fail during time interval \((t_{h-1}, t_h]\), and \( e_{hh} \) is the joint probability corresponding to \( c_{hh} \). Finally, define

\[
g = (g_{11}, g_{12}, \ldots, g_{1m}, g_{21}, \ldots, g_{2m}, g_{31}, \ldots, g_{3m}, \ldots, g_{m1}, \ldots, g_{mm})
\]
and

$$\varphi = (\varphi_{11}, \varphi_{12}, \cdots, \varphi_{1m}, \varphi_{21}, \cdots, \varphi_{2m}, \varphi_{31}, \cdots, \varphi_{3m}, \cdots, \varphi_{m1}, \cdots, \varphi_{mm}),$$

where $g_{hh'}$ is the number of pairs such that the first unit is censored during time interval $(t_{h-1}, t_h]$ and the second unit is censored during time interval $(t_{h-1}, t_h]$, and $\varphi_{hh'}$ is the joint probability corresponding to $g_{hh'}$. These vectors are combined into large vectors

$$V = (f, \delta, \epsilon, g),$$

and

$$\Theta = (\vartheta, \nu, \varphi', \varphi).$$

Then, $V$ has a multinomial distribution with parameters $n$ and $\Theta$. It follows from the multivariate central limit that as $n \to \infty$, the distribution of $\sqrt{n}^{-1}(V - n\Theta)$ converges to multivariate normal distribution with mean 0 and covariance matrix,

$$diag(\Theta) - \Theta\Theta'.$$

Now define the following quantities:

$$f_1 = (u_1, u_2, \cdots, u_m, w_1, w_2, \cdots, w_m),$$

where $u_h$ is the number of failures and $w_h$ is the number of censored units during time interval $(t_{h-1}, t_h]$ experienced by the subjects listed first in the pairs, and

$$f_2 = (v_1, v_2, \cdots, v_m, z_1, z_2, \cdots, z_m),$$

where $v_h$ is the number of failures and $z_h$ is the number of censored units experienced during time interval $(t_{h-1}, t_h]$ for the subjects listed second in the pairs.

$$\phi = (\pi_1, \pi_2, \cdots, \pi_m, r_1, r_2, \cdots, r_m).$$
where $\pi_h$ and $\tau_h$ are the failure probability and censoring probability, respectively such that

$$ \pi_h = \Pr(t_{h-1} < X_{ij} \leq t_h, X_{ij} \leq W_{ij}) = \int_{t_{h-1}}^{t_h} C(x) \mid dS(x) \mid $$

and

$$ \tau_h = \int_{t_{h-1}}^{t_h} S(x) \mid dC(x) \mid . $$

These quantities can be expressed as linear functions of $V$ or $\Theta$ as follows:

$$ f_1 = QV, $$
$$ f_2 = RV, $$
$$ \phi = Q\Theta = R\Theta, $$

where

$$ Q_{2m \times 4m^2} = \begin{pmatrix} A & A & 0 & 0 \\ 0 & 0 & A & A \end{pmatrix}, $$

$$ R_{2m \times 4m^2} = \begin{pmatrix} B & 0 & B & 0 \\ 0 & B & 0 & B \end{pmatrix}, $$

$$ A_{m \times m^2} = \begin{pmatrix} 11 \cdots 1 & 00 \cdots 0 & 00 \cdots 0 & \cdots & 00 \cdots 0 \\ 00 \cdots 0 & 11 \cdots 1 & 00 \cdots 0 & \cdots & 00 \cdots 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 00 \cdots 0 & 00 \cdots 0 & 00 \cdots 0 & \cdots & 11 \cdots 1 \end{pmatrix}. $$
\[ B_{m \times m^2} = \begin{pmatrix} 100 \cdots 0 & 100 \cdots 0 & 100 \cdots 0 & \cdots & 100 \cdots 0 \\ 010 \cdots 0 & 010 \cdots 0 & 010 \cdots 0 & \cdots & 010 \cdots 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 000 \cdots 1 & 000 \cdots 1 & 000 \cdots 1 & \cdots & 000 \cdots 1 \end{pmatrix}. \]

Let \((f_1, f_2)\)' be \(f^*\) and \((\phi, \phi)\)' be \(\phi^*\). It follows that as \(n \to \infty\), the distribution of \(\sqrt{n^{-1}}(f^* - n\phi^*)\) converges to a multivariate normal distribution with mean 0 and covariance matrix,

\[ \Sigma = \begin{bmatrix} Q & R \\ \end{bmatrix} (\text{diag}(\Theta) - \Theta \Theta)' [Q' \quad R']. \] (2.2)

As in the no censoring case, \(\hat{q}_h = \frac{u_h + v_h}{N_h}\), where \(N_h = \sum_{i=h}^m (u_i + v_i + w_i + z_i)\), is a \(\hat{q}_j\)'s are smooth function of \(f^*\). Thus, \(\sqrt{n}(\hat{q} - q^*) \sim N(0, n\Sigma q^*)\) as \(n \to \infty\) and \(\Sigma q^*\) can be obtained from (2.2) by the Delta method based.

It follows that \(\hat{q}_h\) is a consistent estimate of

\[ q_h^* = \frac{\pi_h/2 + \pi_k/2}{\sum_{i=h}^m (\pi_i + \pi_k + \tau_i + \tau_k)/2} = \frac{\pi_h}{\sum_{i=h}^m (\pi_i + \tau_i)}. \]

Since

\[ \sum_{i=h}^m (\pi_i + \tau_i) = S(t_{h-1})C(t_{h-1}), \]

then

\[ q_h^* = \frac{\int_{t_{h-1}}^{t_h} C(x) \, | \, dS(x) |}{S(t_{h-1})C(t_{h-1})}. \] (2.3)

In general \(q_h^*\) may not be equal to \(\frac{S(t_{h-1}) - S(t_h)}{S(t_{h-1})}\), so \(\hat{q}_h\) is not necessarily a consistent estimate of \(q_h\). Furthermore, if we have \(2n\) independent observations that are
subject to random censorship as described in section 2.1, \( \mathbb{E}(\hat{q}_h) \) has exactly the same formula as equation (2.3).

The large sample properties of the AL estimate can be derived in a similar way. For the AL estimator, \( \hat{q}_h \) is equal to \( \frac{u_h+v_h}{N_h-(w_h+c_h)/2} \) and \( \hat{q}^* \) is

\[
\frac{\int_{t_{h-1}}^{t_h} C(x) \mid dS(x) \mid}{S(t_{h-1})C(t_{h-1}) - 0.5 \int_{t_{h-1}}^{t_h} S(x) \mid dC(x) \mid}.
\]

Thus, consistency properties of the PL and AL estimators in the univariate case are the same as those for the PL and AL estimators when pairs are ignored. Variances of the PL and AL estimators are affected by correlation within pair and this is discussed in the next section.

Breslow and Crowley (1974) proposed a necessary and sufficient condition, relating the survival and censoring distributions, for the consistency of the AL estimator for the univariate case. They examined finite sample bias of the AL estimates for estimating the survival probability through simulation study and concluded that the bias will not be serious unless the number of intervals is fewer than ten.

2.2.3 Variance of the PL estimator

2.2.3.1 Notation of conditional binary responses. In this section we derive a modification to the Greenwood formula for the variance of the PL estimator computed from independent groups of correlated responses. We start by formulating interval and right censored survival data as conditional binary responses. The data consist of \( n \) independent pairs, or groups, and the groups may consist of individual life times, which are not independent each other.
Define the random variable $Y_{hij}$:

$$Y_{hij} = \begin{cases} 
1 & \text{if observational unit } j \text{ in group } i \text{ fails during interval } h \text{ given success up to } t_{h-1} \\
0 & \text{if observational unit } j \text{ in group } i \text{ succeeds or censored during interval } h \text{ given success up to } t_{h-1}
\end{cases}$$

where $h = 1, 2, \ldots, m$ intervals, $i = 1, 2, \ldots, n$ groups and $j = 1, 2, \ldots, n_{hi}$ observational units. $Y_{hij}$ is not defined for any observational unit that has failed or been censored prior to $t_{h-1}$. Also, $Y_{hij}$ and $Y_{hij}'$ correspond to the same unit if $j = j'$, and $Y_{hij}$ and $Y_{hij}'$ are independent if $i \neq i'$.

### 2.2.3.2 Formula for $\text{Cov}_{hh}(Y_{hij}, Y_{hij}')$. Since $Y_{hij}$ is a "Bernoulli" random variable, the mean and variance for $Y_{hij}$ can be constructed easily as

$$E_h(Y_{hij}) = \frac{S(t_{h-1}) - S(t_h)}{S(t_{h-1})} = q_h.$$ 

and

$$\text{Var}_h(Y_{hij}) = E_h(Y_{hij}^2) - [E_h(Y_{hij})]^2 = E_h(Y_{hij}) - [E_h(Y_{hij})]^2.$$ 

The subscript $h$ of $E_h$ and $\text{Var}_h$ indicates the conditional mean and variance of a unit surviving to time $t_{h-1}$ and the subscript $hh$ of $\text{Cov}_{hh}$ and $E_{hh}$ indicates conditional covariance on both units surviving to time $t_{h-1}$.

For two units in the same group and at risk in the same interval,

$$\text{Cov}_{hh}(Y_{hij}, Y_{hij}') = E_{hh}(Y_{hij}Y_{hij}') - E_{hh}(Y_{hij})E_{hh}(Y_{hij}).$$

Since
20

\[
Y_{hij}Y_{hij'} = \begin{cases} 
1 & \text{if observational units } j \text{ and } j' \text{ in group } i \text{ both fail} \\
\text{during time interval } h \text{ given success up to } t_{h-1} \\
0 & \text{otherwise},
\end{cases}
\]

\[
E_{hh}(Y_{hij}Y_{hij'}) = \Pr(X_{ij} \in (t_{h-1}, t_h], X_{ij'} \in (t_{h-1}, t_h] \mid X_{ij} > t_{h-1}, X_{ij'} > t_{h-1}) \\
= \frac{S(t_{h-1}, t_{h-1}) - 2S(t_h, t_{h-1}) + S(t_h, t_h)}{S(t_{h-1}, t_{h-1})}
\]

assuming \( S(t_h, t_{h-1}) = S(t_{h-1}, t_h) \), and

\[
E_{hh}(Y_{hij}) = \Pr \left[ X_{ij} \in (t_{h-1}, t_h] \mid X_{ij} > t_{h-1}, X_{ij'} > t_{h-1} \right] \\
= \frac{S(t_{h-1}, t_{h-1}) - S(t_h, t_{h-1})}{S(t_{h-1}, t_{h-1})} \\
= \frac{S(t_{h-1}, t_{h-1}) - S(t_h, t_{h-1})}{S(t_{h-1}, t_{h-1})}
\]

where \( S(\cdot, \cdot) \) is the joint survival function of \( X_{ij} \) and \( X_{ij'} \), and assume that \( X_{ij} \) and \( X_{ij'} \) are exchangeable random variables.

2.2.3.3 Formula for \( \text{Cov}_{hh}(Y_{hij}, Y_{hij'}) \). For results on two units in the same group during different time intervals,

\[
\text{Cov}_{hh}(Y_{hij}, Y_{hij'}) = E_{hh}(Y_{hij}Y_{hij'}) - E_{hh}(Y_{hij})E_{hh}(Y_{hij'})
\]

Since

\[
Y_{hij}Y_{hij'} = \begin{cases} 
1 & \text{if observational unit } j \text{ fails during } h \text{ and } j' \text{ fails during } h' \text{ in group } i \text{ given success for unit } j \text{ up to } t_{h-1} \text{ and success for unit } j' \text{ up to } t_{h-1} \\
0 & \text{otherwise},
\end{cases}
\]

\[
E_{hh}(Y_{hij}Y_{hij'}) = \Pr \left[ X_{ij} \in (t_{h-1}, t_h], X_{ij'} \in (t_{h-1}, t_h] \mid X_{ij} > t_{h-1}, X_{ij'} > t_{h-1} \right] \\
= \frac{S(t_{h-1}, t_{k-1}) - S(t_h, t_{k-1}) - S(t_{h-1}, t_k) + S(t_h, t_k)}{S(t_{h-1}, t_{k-1})}
\]
where $S( , )$ is a joint survival function of $X_{ij}$ and $X_{ij'}$, and assume that $h < k$.

Now $E_{ihh}(Y_{hiij})$ and $E_{ihh}(Y_{h'ij'})$ are

$$E_{ihh}(Y_{hiij}) = \Pr [X_{ij} \in (t_{h-1}, t_h) \mid X_{ij} > t_{h-1}, X_{ij'} > t_{h'-1}]$$
$$= \frac{S(t_{h-1}, t_h') - S(t_h, t_{h'-1})}{S(t_{h-1}, t_{h'-1})}$$
$$= \frac{S(t_{h-1}, t_h') - S(t_h, t_{h'-1})}{S(t_{h-1}, t_{h'-1})}.$$

$$E_{ihh}(Y_{h'ij'}) = \Pr [X_{ij'} \in (t_{h'-1}, t_h) \mid X_{ij} > t_{h-1}, X_{ij'} > t_{h-1}]$$
$$= \frac{S(t_{h-1}, t_{h'-1}) - S(t_{h'}, t_{h'-1})}{S(t_{h'-1}, t_{h'-1})}.$$

However $\text{Cov}_{ihh}(Y_{hiij}, Y_{h'ij'})$ should be zero because $\Pr [X_{ij} \in (t_{h-1}, t_h) \mid X_{ij} > t_{h-1}] = 0$ when $h < h'$.

Moments for the conditional binary responses $Y_{hiij}$ are needed to obtain the variances of the PL estimates. Next, consider the covariance matrix of the estimates of the $p_h$'s, where $p_h = 1 - q_h$ and the covariance matrix of the estimates of the $P_{h'}$'s.

2.2.3.4 Covariance matrix of $P_h$'s. The conditional survival probability, $p_h$ is $1 - q_h$ and $p_h = \Pr (T > t_h \mid T > t_{h-1}) = \frac{P_h}{P_{h-1}}$, where $P_h = \Pr (X_{ij} > t_h)$.

The PL estimator for $p_h$ is $\hat{p}_h = 1 - \frac{D_h}{N_h}$.

Now $D_h$ can be defined as nice function of $Y_{hiij}$.

$$D_h = \text{Number of units to fail in } (t_{h-1}, t_h] = \left( \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} Y_{hiij} \right),$$
where $n_{hi}$ is the
number of units at risk in \( i^{th} \) group in \( h^{th} \) interval, and \( N_h \) is the number of units at risk at time \( t_{h-1} \) such that

\[
N_h = \sum_{i=1}^{n} n_{hi}. 
\]

Define the vector \( D \) as

\[
D' = (D_1, D_2, D_3, \ldots, D_m) 
\]

\[
= \sum_{i=1}^{n} \left( \sum_{j=1}^{n_{1i}} Y_{1ij}, \sum_{j=1}^{n_{2i}} Y_{2ij}, \sum_{j=1}^{n_{3i}} Y_{3ij}, \ldots, \sum_{j=1}^{n_{mi}} Y_{mij} \right). 
\]

Thus \( \text{Var}(D_h) \) can be expressed as the function of covariances of \( Y_{hij} \), \( \text{Var}(D_h) = \text{Var} \left( \sum_{i=1}^{n} \sum_{j=1}^{n_{hi}} Y_{hij} \right) \). Since \( \text{Cov}_{hh}(Y_{hij}, Y_{h'i'j}) = 0 \),

\[
\text{Var}(D_h) = \sum_{i=1}^{n} \sum_{j=1}^{n_{hi}} \text{Var}_h(Y_{hij}) + 2 \sum_{i=1}^{n} K_{hi} \text{Cov}_{hh}(Y_{hij}, Y_{h'i'j}), 
\]

where \( K_{hi} = \binom{n_{hi}}{2} \), and \( K_{hi} = 0 \) if \( n_{hi} \leq 1 \). If we consider just paired life time data, \( \sum_{i=1}^{n} K_{hi} \) is exactly the same as the number of pairs at risk at time \( t_{h-1} \) (i.e., the number of pairs for which both units survive through time \( t_{h-1} \)).

The covariance of \( D_h \) and \( D_{h'} \) is equal to \( \text{Cov} \left( \sum_{i=1}^{n} \sum_{j=1}^{n_{hi}} Y_{hij}, \sum_{i=1}^{n} \sum_{j=1}^{n_{h'i}} Y_{h'i'j} \right) \).

Since \( \text{Cov}(Y_{hij}, Y_{h'i'j}) = 0 \),

\[
\text{Cov}(D_h, D_{h'}) = \sum_{i=1}^{n} \text{Cov} \left( \sum_{j=1}^{n_{hi}} Y_{hij}, \sum_{j=1}^{n_{h'i}} Y_{h'i'j} \right) 
\]

\[
= \sum_{i=1}^{n} (n_{hi}(n_{hi} - 1)) \text{Cov}_{hh}(Y_{hij}, Y_{h'i'j}) 
\]

\[
= K_{hh} \text{Cov}(Y_{hij}, Y_{h'i'j}), 
\]

(2.5)
where $K_{hh'} = \sum_{i=1}^{n} n_{h_i}(n_{h_i} - 1)$, and it is assumed that $\text{Cov}(Y_{h_i j}, Y_{h_i j'})$ are same for all $i$'s.

The most commonly used formula for the variance of the survival probability estimate was suggested by Greenwood (1926) assuming $\text{Cov}(\hat{p}_i, \hat{p}_j) = 0$, if $i \neq j$, and $\text{Var}(\hat{p}_h) = \frac{p_h(1-p_h)}{N_h}$ in the univariate case. However the assumptions should be changed for the multivariate life time data. A modified Greenwood formula can be obtained as follows.

Since $\hat{p}_h = \hat{p}_1 \hat{p}_2 \cdots \hat{p}_h$, then $\text{Cov}(\hat{P}_h) \approx dVd$ by delta method, where

$$
\begin{pmatrix}
\begin{array}{cccc}
\frac{\partial \hat{P}_1}{\partial \hat{p}_1} & 0 & 0 & \cdots \cdots & 0 \\
\frac{\partial \hat{P}_2}{\partial \hat{p}_1} & \frac{\partial \hat{P}_2}{\partial \hat{p}_2} & 0 & \cdots \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{\partial \hat{P}_h}{\partial \hat{p}_1} & \frac{\partial \hat{P}_h}{\partial \hat{p}_2} & \frac{\partial \hat{P}_h}{\partial \hat{p}_3} & \cdots \cdots & \frac{\partial \hat{P}_h}{\partial \hat{p}_h}
\end{array}
\end{pmatrix},
$$

and

$$V = \text{Cov}(\hat{p}_1, \hat{p}_2, \hat{p}_3, \cdots, \hat{p}_h).$$

Since $\text{Var}(\hat{P}_h)$ is the diagonal vector of $\text{Cov}(\hat{P}_h)$,

$$\text{Var}(\hat{P}_h) = \hat{p}_h^2 \sum_{i=1}^{h} \sum_{j=1}^{h} \frac{\text{Cov}(\hat{p}_i, \hat{p}_j)}{\hat{p}_i \hat{p}_j} = \hat{p}_h^2 \sum_{i=1}^{h} \frac{\text{Var}(\hat{p}_i)}{\hat{p}_i^2} + \hat{p}_h^2 \sum_{i=1}^{h} \sum_{j \neq i}^{h} \frac{\text{Cov}(\hat{p}_i, \hat{p}_j)}{\hat{p}_i \hat{p}_j},$$

which looks like Greenwood formula with an additional part. Evaluation of (2.6)
requires numerical values for \( \text{Var}(\hat{p}_h) \) and \( \text{Cov}(\hat{p}_h, \hat{p}_{h'}) \) and their estimates. Here,

\[
\text{Var}(\hat{p}_h) = \frac{\text{Var}(D_h)}{(N_h)^2}, \quad \text{and} \quad \text{Cov}(\hat{p}_h, \hat{p}_{h'}) = \frac{\text{Cov}(D_h, D_{h'})}{N_h N_{h'}}.
\]

Thus, using (2.4) and (2.5), estimates of \( \text{Var}(\hat{p}_h) \) and \( \text{Cov}(\hat{p}_h, \hat{p}_{h'}) \) are

\[
\text{Var}(\hat{p}_h) = \frac{\hat{\text{Var}}(D_h)}{(N_h)^2} = \frac{\hat{q}_h(1 - \hat{q}_h)}{N_h} + \frac{2 \sum_{i=1}^{n} K_{hi} (\hat{q}_{hh} - (\hat{\epsilon}_{hh})^2)}{(N_h)^2}, 
\]

\[
\text{Cov}(\hat{p}_h, \hat{p}_{h'}) = \frac{\text{Cov}(D_h, D_{h'})}{N_h N_{h'}} = \frac{K_{hh'}(\hat{q}_{hh'} - \hat{\epsilon}_{hh'}^1 \hat{\epsilon}_{hh'}^2)}{N_h N_{h'}},
\]

where

\( \hat{q}_h \) is an estimate of \( \mathbb{E}_h(Y_{hij}) \),
\( \hat{q}_{hh} \) is an estimate of \( \mathbb{E}_{hh}(Y_{hij} Y_{hij}) \),
\( \hat{\epsilon}_{hh} \) is an estimate of \( \mathbb{E}_{hh}(Y_{hij}) \),
\( \hat{q}_{hh'} \) is an estimate of \( \mathbb{E}_{hh'}(Y_{hij} Y_{hij'}) \),
\( \hat{\epsilon}_{hh'}^1 \) is an estimate of \( \mathbb{E}_{hh'}(Y_{hij}) \), and
\( \hat{\epsilon}_{hh'}^2 \) is an estimate of \( \mathbb{E}_{hh'}(Y_{hij}) \).

These estimators are described in section 4.2.1 of Chapter 4.

2.3 AL Estimate

If the number of censored observations, \( C_h \), is not zero in \( (t_{h-1}, t_h] \), \( D_h/N_h \) might be expected to underestimate \( q_h \), the conditional failure probability, because
some of the censored observations at \((t_{h-1}, t_h]\) could actually fail in that time interval. Therefore, an adjustment of \(D_h/N_h\) is desirable. The most commonly used formula for the estimation of \(q_h\) is

\[ \hat{q}_h = \frac{D_h}{N_h - C_h/2}. \]  

The denominator \(N_h - C_h/2\) is called the "adjusted number at risk" and its notation is \(N'_h\). The \(\hat{q}_h\) using \(N'_h\) is the so-called standard life-table estimate or actuarial estimate. In this paper, it is called the AL estimate.

The variance of the AL estimate ignoring pairs is not much different from that of the PL estimate ignoring pairs because there are corresponding quantities in numerator and denominator of (2.6) that are adjusted in a similar way. Formulas for these additional adjustments to (2.7) and (2.8) are presented in section 4.2.2 of Chapter 4.

2.4 Flexibility for General Multivariate Response Times

The previous methods are easily extended to situations involving correlated response times occurring in groups of different sizes, i.e., singletons, some pairs, some triplets, some quadruples, and so on. Point estimates of marginal survival probabilities are simply obtained from a life-table method where the grouping is ignored.

To evaluate modified Greenwood formula for variances and covariances, the quantities with double subscripts \(h, h'\), or \(hi\) in (2.8) and (2.7) are computed from information provided by the pairs, triplets, quadruples ..., but the singletons are not used to estimate these quantities. The singletons are used along with the pairs, triplets, quadruples ... to estimate the quantities with single subscript in (2.8) and (2.7). To estimate the quantities with double subscripts, we need to make all possible
pairs for each triplet, each quadruple, etc .... It is assumed that the any original pair
or any pair obtained from a triplet, or a quadruple, etc, has the same bivariate joint
distribution. Therefore, the data contain 3 pairs for each triplet, 6 pairs for each
quadruple, etc ... that can be used to estimate the quantities with double subscripts
$h_h$, $h_i$, or $h_i$. These are simply inserted into the formulas given in section 4.2 of
Chapter 4.

Hence, the modified Greenwood formula is easily applied to situations where
there are independent cohorts of correlated response times of varying sizes, including
singletons. This extension is used in the following example.

2.5 Example

2.5.1 Description of the angioplasty study

The data were obtained from 301 patients treated by doctors at the Iowa Heart
Center in Des Moines, Iowa, between 1987 and 1991. Each patient underwent at least
one angioplasty procedure to remove or reduce obstructions in leg blood vessels. Some
patients underwent more than one angioplasty procedure involving blood vessels in
different legs or different parts of the same leg. There were 484 procedures performed
on the 301 patients.

Only procedures that satisfied the criterion for initial clinical success were in­
cluded in the life-table analysis of failure times. A procedure was defined as an initial
clinical success if the ankle-arm index (AAI) recorded after completion of the proce­
dure was at least 0.1 larger than the AAI recorded immediately before procedures.
The AAI is the blood pressure reading taken near the ankle divided by a correspond­
ing blood pressure reading taken in the arm on the same side of the body. This is
a measure of relative blood flow in the leg. The initial success criterion reduced the
data set to 159 procedures performed on 115 patients. There were 79 patients with a
single procedure, 30 with 2 procedures, 4 with 3 procedures, and 2 with 4 procedures.
The angioplasty data set is presented in Appendix D.

2.5.2 Life-table analysis

The procedures were inspected at 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 42, and
54 months after the operation. A procedure was considered to have failed if the
AAI was no longer 0.1 larger than the pre-operation level. There are two types
of censored observations. The first type corresponds to procedures that is lost to
following up and no information is recorded after some inspection time. The other
type corresponds to right censored observations for subject still available for followup
that are caused by allowing subjects to continually enter the study between 1987 and
1991 but terminating observation on all subjects at the end of 1991. However these
two types are considered identically in life-table analysis as withdrawals. The results
for the PL estimator are presented in Table 2.1 and the results for the AL estimator
are shown in Table 2.2. The estimated survival curves obtained from the PL and AL
methods are presented at Figure 2.1. Variance from the modified Greenwood formula
is slightly larger than those provided by the unmodified Greenwood formula. The
differences between the modified and unmodified Greenwood formula are not large
in this case because the data contain many independent singletons.
Table 2.1: Life-table for PL computed from angioplasty data

<table>
<thead>
<tr>
<th>interval in months $A_h$</th>
<th>$#$ of failures $D_h$</th>
<th>$#$ of withdrawals $C_h$</th>
<th>$#$ at risk $N_h$</th>
<th>$1 - \hat{q}_h$</th>
<th>PL</th>
<th>usd(PL)</th>
<th>msd(PL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,3]</td>
<td>18</td>
<td>33</td>
<td>159</td>
<td>0.887</td>
<td>0.887</td>
<td>0.0251</td>
<td>0.0256</td>
</tr>
<tr>
<td>(3,6]</td>
<td>10</td>
<td>22</td>
<td>108</td>
<td>0.907</td>
<td>0.805</td>
<td>0.0336</td>
<td>0.0356</td>
</tr>
<tr>
<td>(6,9]</td>
<td>11</td>
<td>14</td>
<td>76</td>
<td>0.855</td>
<td>0.688</td>
<td>0.0434</td>
<td>0.0457</td>
</tr>
<tr>
<td>(9,12]</td>
<td>2</td>
<td>11</td>
<td>51</td>
<td>0.961</td>
<td>0.661</td>
<td>0.0457</td>
<td>0.0476</td>
</tr>
<tr>
<td>(12,15]</td>
<td>1</td>
<td>5</td>
<td>38</td>
<td>0.974</td>
<td>0.644</td>
<td>0.0477</td>
<td>0.0494</td>
</tr>
<tr>
<td>(15,18]</td>
<td>1</td>
<td>7</td>
<td>32</td>
<td>0.969</td>
<td>0.624</td>
<td>0.0503</td>
<td>0.0568</td>
</tr>
<tr>
<td>(18,21]</td>
<td>0</td>
<td>2</td>
<td>24</td>
<td>1.000</td>
<td>0.624</td>
<td>0.0503</td>
<td>0.0568</td>
</tr>
<tr>
<td>(21,24]</td>
<td>1</td>
<td>5</td>
<td>22</td>
<td>0.955</td>
<td>0.595</td>
<td>0.0554</td>
<td>0.0609</td>
</tr>
<tr>
<td>(24,30]</td>
<td>3</td>
<td>5</td>
<td>16</td>
<td>0.813</td>
<td>0.484</td>
<td>0.0735</td>
<td>0.0753</td>
</tr>
<tr>
<td>(30,36]</td>
<td>0</td>
<td>2</td>
<td>8</td>
<td>1.000</td>
<td>0.484</td>
<td>0.0735</td>
<td>0.0754</td>
</tr>
<tr>
<td>(36,42]</td>
<td>0</td>
<td>4</td>
<td>6</td>
<td>1.000</td>
<td>0.484</td>
<td>0.0735</td>
<td>0.0755</td>
</tr>
<tr>
<td>(42,54]</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1.000</td>
<td>0.484</td>
<td>0.0735</td>
<td>0.0755</td>
</tr>
<tr>
<td>(54,\infty)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

$\#: Number$

usd(PL): Standard deviation of PL estimate from unmodified Greenwood formula

msd(PL): Standard deviation of PL estimate from modified Greenwood formula
Table 2.2: Life-table for AL computed from angioplasty data

<table>
<thead>
<tr>
<th>interval in months</th>
<th># of failures $A_h$</th>
<th># of withdrawals $D_h$</th>
<th># of withdrawals $C_h$</th>
<th>Adj. # at risk $\hat{q}_h$</th>
<th>$1 - \hat{q}_h$</th>
<th>AL</th>
<th>usd(AL)</th>
<th>msd(AL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,3]</td>
<td>18</td>
<td>33</td>
<td>142.5</td>
<td>0.874</td>
<td>0.874</td>
<td>0.0278</td>
<td>0.0283</td>
<td></td>
</tr>
<tr>
<td>(3,6]</td>
<td>10</td>
<td>22</td>
<td>97.0</td>
<td>0.897</td>
<td>0.784</td>
<td>0.0368</td>
<td>0.0384</td>
<td></td>
</tr>
<tr>
<td>(6,9]</td>
<td>11</td>
<td>14</td>
<td>69.0</td>
<td>0.841</td>
<td>0.659</td>
<td>0.0463</td>
<td>0.0481</td>
<td></td>
</tr>
<tr>
<td>(9,12]</td>
<td>2</td>
<td>11</td>
<td>45.5</td>
<td>0.956</td>
<td>0.630</td>
<td>0.0486</td>
<td>0.0500</td>
<td></td>
</tr>
<tr>
<td>(12,15]</td>
<td>1</td>
<td>5</td>
<td>35.5</td>
<td>0.972</td>
<td>0.612</td>
<td>0.0504</td>
<td>0.0516</td>
<td></td>
</tr>
<tr>
<td>(15,18]</td>
<td>1</td>
<td>7</td>
<td>28.5</td>
<td>0.965</td>
<td>0.591</td>
<td>0.0530</td>
<td>0.0594</td>
<td></td>
</tr>
<tr>
<td>(18,21]</td>
<td>0</td>
<td>2</td>
<td>23.0</td>
<td>1.000</td>
<td>0.591</td>
<td>0.0530</td>
<td>0.0594</td>
<td></td>
</tr>
<tr>
<td>(21,24]</td>
<td>1</td>
<td>5</td>
<td>19.5</td>
<td>0.949</td>
<td>0.560</td>
<td>0.0583</td>
<td>0.0636</td>
<td></td>
</tr>
<tr>
<td>(24,30]</td>
<td>3</td>
<td>5</td>
<td>13.5</td>
<td>0.778</td>
<td>0.436</td>
<td>0.0779</td>
<td>0.0792</td>
<td></td>
</tr>
<tr>
<td>(30,36]</td>
<td>0</td>
<td>2</td>
<td>7.0</td>
<td>1.000</td>
<td>0.436</td>
<td>0.0779</td>
<td>0.0792</td>
<td></td>
</tr>
<tr>
<td>(36,42]</td>
<td>0</td>
<td>4</td>
<td>4.0</td>
<td>1.000</td>
<td>0.436</td>
<td>0.0779</td>
<td>0.0792</td>
<td></td>
</tr>
<tr>
<td>(42,54]</td>
<td>0</td>
<td>1</td>
<td>1.5</td>
<td>1.000</td>
<td>0.436</td>
<td>0.0779</td>
<td>0.0792</td>
<td></td>
</tr>
<tr>
<td>(54,\infty)</td>
<td>1</td>
<td>0</td>
<td>1.0</td>
<td>0.000</td>
<td>0.000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
</tbody>
</table>

\#: Number
usd(AL): Standard deviation of AL estimate from unmodified Greenwood formula
msd(AL): Standard deviation of AL estimate from modified Greenwood formula
Figure 2.1: Comparison of PL and AL survival curves for angioplasty data
3.1 Likelihood Function

3.1.1 Univariate case

Kaplan and Meier (1958) derived the product limit estimator of the survivor function as a maximum likelihood estimator under the case of fixed censorship. Turnbull (1974) derived the likelihood function for interval censored data that are subject to both right and left random censoring. This is commonly called doubly censored data, and Turnbull referred to right censored observations as "losses" and left censored observations as "late entries". Turnbull (1974) established sufficient conditions for the existence and uniqueness of the consistent maximum likelihood estimates. He also gave explicit expressions for the variances and covariances of the asymptotic normal distribution in the case where the number of time intervals remains fixed as the number of respondents becomes large.

Suppose we observe an exact failure or censoring time $X_i$ for each subject. $X_i$ can be defined as $X_i = \min(X_i^0, W_i)$, where $X_i^0$ is the true survival time, with survival probability $P_h = \Pr(X_i^0 > t_h)$ and $W_i$ is the censoring time with $G_h = \Pr(W_i > t_h)$. Now suppose that subjects are inspected only at a finite set of time
points $0 < t_1 < t_2 < \cdots < t_m < \infty$, so it is only known whether a failure or censoring event occurred between two inspection times. Exact times are not observed. These are called interval censored data. Turnbull (1974) assumed that the withdrawals, $C_h$ can only occur at the end of a time interval $(t_{h-1}, t_h]$. This assumption was also made by Kaplan and Meier (1958) in their previous development of the "product limit estimator". Turnbull also assumed that censoring times are independent of failure times. Under those assumptions, $W_i$ has a discrete distribution and the likelihood function is proportional to

$$L = \prod_{h=1}^{m} (P_{h-1} - P_h)^{\alpha_h}(P_h)^{c_h} G_h^{\alpha_h} g_h^{c_h},$$  \hspace{1cm} (3.1)

where $\alpha_h$ is the number of failures observed in $(t_{h-1}, t_h]$, and $c_h$ is the number of withdrawals at $t_h$, and $g_h = \Pr(W_i = t_h)$. Since we are only interested in estimating the survival probabilities $\{P_h\}$, we only need to maximize

$$\prod_{h=1}^{m} (P_{h-1} - P_h)^{\alpha_h}(P_h)^{c_h}$$  \hspace{1cm} (3.2)

subject to the constraints that $1 \geq P_1 \geq P_2 \geq \cdots \geq P_m \geq 0$. We note that (3.2) can also be derived under an assumption of fixed censorship.

3.1.2 Bivariate case

In this Chapter, Turnbull's procedure will be extended to the bivariate case where both members of each pair have the same marginal survival distribution. Kaplan and Meier (1958) and Turnbull (1974) expressed the non-parametric likelihood function in terms of survival probabilities, but it is more convenient to express the likelihood in terms of the joint failure probabilities $\{\pi_{ij}\}$ for the time intervals to avoid constrained
maximization of the likelihood function. Following Turnbull (1974), we will assume
that censored observations can only occur at the end of a time interval.

There are four cases for bivariate life time data:

(i) Both units are observed to fail.
(ii) The first unit is observed to fail and second unit is censored.
(iii) The first unit is censored and the second unit is observed to fail.
(iv) Both units are censored.

In the following development each member of a pair is assumed to have the same
marginal distribution, so no distinction is made with respect to order. Hence case
(ii) can not be distinguished from case (iii). Consequently, both the likelihood and the
estimates obtained from maximizing the likelihood should be invariant to ordering
within pairs. This is achieved by including appropriate symmetry constraints on the
parameters in the likelihood function.

Define four tables ($F$, $C1$, $C2$, and $CC$) corresponding to the four cases listed
above, and denote the elements of those tables as follows: $F_{ij}$ is the number of pairs
where the first unit of the pair fails in $(t_{i-1}, t_i]$ and the second unit fails in $(t_{j-1}, t_j]$, $C1_{ij}$ is the number of pairs where the first unit of the pair fails in $(t_{i-1}, t_i]$ and the
second unit is censored at $t_j$. $C2_{ij}$ is the number of pairs where the first unit is
censored at $t_i$ and the second unit fails in $(t_{j-1}, t_j]$, and $CC_{ij}$ is the number of
pairs where the first unit of the pair is censored at $t_i$ and the second unit is censored
at $t_j$.

Thus the total number of pairs is

$$n = \sum_{i=1}^{m} \sum_{j=1}^{m} (F_{ij} + C1_{ij} + C2_{ij} + CC_{ij}).$$
We need to know the probability function for each case to establish the likelihood function. The notation for survival data established in Chapter 2 is also used here.

Define the following quantities:

\[
\begin{align*}
\pi_{ij} &= \Pr(t_{i-1} < X_1^0 \leq t_i, t_{j-1} < X_2^0 \leq t_j) \\
T_{ij} &= \Pr(W_1 > t_i, W_2 > t_j) \\
\Delta_{ij} &= \Pr(W_1 > t_i, W_2 = t_j) \\
\Xi_{ij} &= \Pr(W_1 = t_i, W_2 > t_j) \\
\Gamma_{ij} &= \Pr(W_1 = t_i, W_2 = t_j).
\end{align*}
\]

Also, \( \delta_i \) is 1 if the \( i^{th} \) unit is not censored, otherwise \( \delta_i \) is zero.

Define \( a = m + 1, \ t_0 = 0, \) and \( t_{(m+1)} = t_a = \infty. \) Then, the probability function for each case is as follows: For the first case,

\[
\Pr(t_{i-1} < X_1 \leq t_i, t_{j-1} < X_2 \leq t_j, \delta_1 = 1, \delta_2 = 1) = \Pr(t_{i-1} < X_1^0 \leq t_i, t_{j-1} < X_2^0 \leq t_j, W_1 > t_i, W_2 > t_j) = \pi_{ij} T_{ij},
\]

since the distribution for failure times and censoring times are independent. For the second case,

\[
\Pr(t_{i-1} < X_1 \leq t_i, t_{j-1} < X_2 \leq t_j, \delta_1 = 0, \delta_2 = 1) = \Pr(X_1^0 > t_i, t_{j-1} < X_2^0 \leq t_j, W_1 = t_i, W_2 > t_j) = \gamma_{ij} \Delta_{ij},
\]

where \( \gamma_{ij} = \sum_{k=i+1}^{a} \pi_{ik}. \) Similarly, for the third case,

\[
\Pr(t_{i-1} < X_1 \leq t_i, t_{j-1} < X_2 \leq t_j, \delta_1 = 1, \delta_2 = 0) = \Pr(X_1^0 > t_i, t_{j-1} < X_2^0 \leq t_j, W_1 = t_i, W_2 = t_j) = \eta_{ij} \Xi_{ij},
\]
where \( \eta_{ij} = \sum_{l=i+1}^{q} \pi_{lj} \). For the fourth case, where times are right censored for both members of the pair,

\[
Pr(t_{i-1} < X_1 \leq t_i, t_{j-1} < X_2 \leq t_j, \delta_1 = 0, \delta_2 = 0) = Pr(X_1^o > t_i, X_2^o > t_j, W_1 = t_i, W_2 = t_j) = \zeta_{ij} \Gamma_{ij},
\]

where \( \zeta_{ij} = \sum_{l=i+1}^{q} \sum_{k=j+1}^{q} \pi_{lk} \). Consequently, the non-parametric likelihood is proportional to

\[
L = \prod_{i=1}^{m} \prod_{j=1}^{m} \left[ (\pi_{ij})^{Fi_j} (\gamma_{ij})^{G1_{ij}} (\eta_{ij})^{G2_{ij}} (\zeta_{ij})^{G3_{ij}} \right], \quad (3.3)
\]

Since we are only interested in estimating the joint failure probabilities \( \{ \pi_{ij} \} \), we need only maximize

\[
L^* = \prod_{i=1}^{m} \prod_{j=1}^{m} \left[ (\pi_{ij})^{Fi_j} (\sum_{k=j+1}^{a} \pi_{ik})^{G1_{ij}} (\sum_{l=i+1}^{a} \pi_{lj})^{G2_{ij}} (\sum_{l=i+1}^{a} \sum_{k=j+1}^{a} \pi_{lk})^{G3_{ij}} \right], \quad (3.4)
\]

with respect to \( \pi_{ij} \).

### 3.2 Likelihood Equations

The restrictions on the \( \{ \pi_{ij} \} \) must be considered when (3.4) is maximized. The assumptions that the marginal survival distributions for two life time variables in each pair are homogeneous and the two life time variables are exchangeable random variables imply that \( \pi_{ij} \) is equal to \( \pi_{ji} \) for all \((i,j)\). Since \( \pi_{ij} \) is a joint failure probability, \( \pi_{ij} \) is not less than zero for any \((i,j)\), and the sum of the \( \pi_{ij} \)'s is 1. For these restrictions, the number of parameters to estimate is \( a(a+1)/2 - 1 \).
From (3.4), \( \log L^* \) is

\[
\sum_{i,j}^{m} \left[ F_{ij} \log \pi_{ij} + C_{1ij}(\log \sum_{k=j+1}^{a} \pi_{ik}) + C_{2ij}(\log \sum_{l=i+1}^{a} \pi_{lj}) + CC_{ij}(\log \sum_{l=i+1}^{a} \sum_{k=j+1}^{a} \pi_{lk}) \right].
\]

(3.5)

To eliminate the need to check the restrictions, \( 0 \leq \pi_{ij} \leq 1 \) and \( \sum \sum \pi_{ij} = 1 \), in iterative numerical algorithms for maximizing \( \log L^* \), make the transformation \( \theta_{ij} = \log (\pi_{ij}/\pi_{aa}) \). Since \( \pi_{ij} = \pi_{aa} \exp \theta_{ij} \), the log likelihood function (3.5) becomes

\[
\Lambda = \sum_{i,j}^{m} \left\{ F_{ij}(\theta_{ij} + \log \pi_{aa}) + C_{1ij}[\log(\pi_{aa} \sum_{k=j+1}^{a} e^{\theta_{ik}})] + C_{2ij}[\log(\sum_{l=i+1}^{a} \pi_{aa} \sum_{k=j+1}^{a} e^{\theta_{lk}})] 
+ CC_{ij}[\log(\pi_{aa}) + C_{1ij}[\log(\sum_{l=i+1}^{a} \sum_{k=j+1}^{a} e^{\theta_{lk}})]
+ C_{2ij}[\log(\sum_{l=i+1}^{a} e^{\theta_{lj}})] + CC_{ij}[\log(\sum_{l=i+1}^{a} \sum_{k=j+1}^{a} e^{\theta_{lk}})] \right\},
\]

(3.6)

where \( \pi_{aa} \) is

\[
\pi_{aa} = \frac{1}{1 + \sum_{i,j \neq a} e^{\theta_{ij}}},
\]

(3.7)

This must be maximized subject to the constraints that \( \theta_{i,j} = \theta_{j,i} \) for all \((i,j)\).

Now consider the first partial derivative of \( \Lambda \) with respect to \( \theta_{uv} \). Since we assume \( \pi_{ij} \) is equal to \( \pi_{ji} \), \( \theta_{ij} \) is equal to \( \theta_{ji} \). Thus there are two cases for the first partial derivative of \( \Lambda \) with respect to \( \theta_{uv} \) corresponding to \( u \neq v \) and \( u = v \). If \( u \neq v \), then

\[
\frac{d\Lambda}{d\theta_{uv}} = -2n\pi_{aa}e^{\theta_{uv}} + (F_{uv} + F_{vu})
+ \sum_{i=1}^{u-1} \left[ (C_{1vi} + C_{2iv}) \frac{e^{\theta_{uv}}}{\sum_{l=i+1}^{a} e^{\theta_{lv}}} \right]
\]
If \( u = v \), then

\[
\frac{d\Lambda}{d\theta_{uu}} = -n\pi_{aa}e^{\theta_{uu}} + F_{uu}
\]

\[
+ \sum_{i=1}^{u-1} \left[ (C_{1ui} + C_{2iu}) \frac{e^{\theta_{uu}}}{\sum_{l=i+1}^{u} e^{\theta_{lu}}} \right]
\]

\[
+ \sum_{i=1}^{u-1} \sum_{j=1}^{v-1} \left[ (CC_{ij} + CC_{ji}) \frac{e^{\theta_{uu}}}{\sum_{l=i+1}^{u} \sum_{k=j+1}^{v} e^{\theta_{lk}}} \right].
\] (3.8)

Equations (3.8) and (3.9), the likelihood equations for \( \{\theta_{ij}\} \), must be solved numerically using some iterative algorithm. If there is no censoring, equations (3.8) and (3.9) have explicit solutions

\[
\hat{\theta}_{uv} = \log(F_{uv} + F_{vu}) - \log(2F_{aa})
\]

\[
\hat{\theta}_{uu} = \log F_{uu} - \log(F_{aa}).
\] (3.10)

Then a set of starting values for the iterative algorithm is obtained by solving (3.10) using only the \( \{F_{ij}\} \) counts and ignoring the counts in \( C1, C2, \) and \( CC \). We used the Davidson-Fletcher-Powell method, which is originally due to Davidson (1959) and justified by Fletcher and Powell (1963), to get close to the solution and then applied a few iterations of a Newton-Raphson algorithm to get the final answer. The Newton-Raphson algorithm requires second partial derivatives, and the inverse of the matrix of second partial derivatives evaluated at the final solution provides an estimate of
the covariance matrix for the parameter estimates. The fourteen different formulas used to compute the second partial derivative matrix are shown in Appendix A. The S-plus programs used to evaluate the loglikelihood (3.6), the first partial derivatives of the loglikelihood (3.6), and the second partial derivatives of the loglikelihood (3.6) with respect to $\theta_{ij}$ are listed on Appendix B. The procedure for evaluating MLE's is further discussed in Chapter 4.

We can get the MLE's of the $\pi_{ij}$'s directly from the MLE's for the $\theta_{ij}$'s. It follows that the MLE's of the marginal survival probabilities are

$$\hat{P}_h = 1 - \sum_{i=1}^{a} \sum_{j=1}^{b} \hat{\pi}_{ij}, \quad h = 1, 2, \ldots, m. \quad (3.11)$$

The MLE's for the joint survival probabilities are given by

$$\hat{c}_{ij} = \sum_{l=i+1}^{a} \sum_{k=j+1}^{b} \hat{\pi}_{lk}. \quad (3.12)$$

From the large sample properties of maximum likelihood estimates (Lehmann, 1983), $\hat{\theta} = (\hat{\theta}_{11}, \hat{\theta}_{12}, \ldots, \hat{\theta}_{1a}, \hat{\theta}_{22}, \ldots, \hat{\theta}_{2a}, \hat{\theta}_{33}, \ldots, \hat{\theta}_{ma})$ is asymptotically normal with asymptotic mean $\theta$ and covariance matrix equal to the inverse of the expectation of the negative of the second partial derivative matrix of (3.6). Since $\pi$, $P$, and $\zeta$ are functions of $\theta$, their MLE's also have the asymptotic normality and other properties such as "efficiency" and "consistency" by the invariance property of the MLE, and the covariance matrices of $\hat{P}$ and $\hat{\zeta}$ can be evaluated by Delta method, where $\pi$, $P$, and $\zeta$ are corresponding vectors of $\{\pi_{ij}\}$, $\{P_i\}$, and $\{\zeta_{ij}\}$.

It is not difficult to derive the likelihood function for general multivariate response times, but it is more tedious to obtain the likelihood equations and the second partial derivative matrix because the number of parameters increases as you
include triplets, quadruples, etc. Including triplets and quadruples can also increase the chance that the MLE’s are not unique. Uniqueness and existence of MLE’s are discussed in the next section for the case dealing only with pairs.

3.3 Existence and Uniqueness of MLE

Campbell (1981) studied bivariate survival estimators using the non-parametric likelihood (3.4) parameterized in terms of the survival probabilities \( \{\zeta_{ij}\} \), and discussed existence and uniqueness of the maximum likelihood estimates. However Campbell (1981) did not include either the symmetry restrictions or the weaker restrictions corresponding to homogeneous margins. The matrix of second partial derivatives with respect to \( \{\zeta_{ij}\} \) is less tedious to derive than the second partial matrix with respect to \( \{\pi_{ij}\} \) under the symmetry restrictions, but maximizing the loglikelihood function with respect to \( \{\zeta_{ij}\} \) requires an algorithm for constrained optimization to ensure that resulting parameter estimates satisfy the restriction \( \zeta_{ij} \geq \zeta_{kl} \) for \( i, j \) such as \( i \leq j \) and \( k \leq l \). Consequently, we previously expressed the loglikelihood as a function of \( \{\pi_{ij}\} \), but in this section it is more convenient to express the loglikelihood as a function of \( \{\zeta_{ij}\} \).

Campbell (1981) showed that a bivariate self-consistent estimate is a MLE. He proved existence of the MLE and showed that the likelihood function is convex in the \( \zeta_{ij} \)'s. This implies that the likelihood is unimodal and the MLE is unique up to possible flat spots. The arguments in his paper can be modified to show that the bivariate self-consistent estimate is a MLE and discuss the existence and uniqueness of the maximum likelihood estimate for \( \zeta_{ij} \) under the restrictions of homogeneous margins and symmetry for \( \pi_{ij} \).
Recall that the four data tables are denoted by
\[ F = \frac{F + F'}{2}, \]
\[ C1 = \frac{C1 + C2}{2}, \]
\[ C2 = \frac{C1 + C2}{2}, \]
\[ CC = \frac{CC + CC'}{2}. \]

We will show that the likelihood function is convex function of the parameters under the symmetry restrictions, \( \zeta_{ij} = \zeta_{ji} \) by showing the matrix of second partial derivatives of \(- \log L^*\), denoted by \( M \), with respect to the \( \zeta_{ij} \)'s is nonnegative definite by decomposing this matrix in a manner similar to that used by Campbell.

Define the matrix \( K \) by multiplying each element the \( IJ \times IJ \) matrix \( M \) by the corresponding element of \( Q = \left( q(ij)(kl) \right) \), where \( q(ij)(kl) = (-1)^{i+j}(-1)^{k+l} \times X_{ijkl} \), and
\[
X_{ijkl} = \begin{cases} 
2 & \text{if } i = j \text{ and } k = l, \\
1 & \text{if } i = j \text{ and } k \neq l, \\
\text{or } i \neq j \text{ and } k = l, \\
0.5 & \text{elsewhere.} 
\end{cases}
\]

Then, the matrix \( K \) has all nonnegative entries.

Next define the following matrices:

- \( W \) is diagonal \( IJ \times IJ \) matrix with entries \( \frac{2CC_{ij}}{\zeta_{ij}^2} \) at \((ii)(ii)\) and \( \frac{CC_{ij}}{2\zeta_{ij}^2} \) at \((ij)(ij)\), for \( i \neq j \).

- \( A_{ij} \) is the \( IJ \times IJ \) matrix with ones at \((ij)(ij)\), \((ij)(i,j-1)\), \((i,j-1)(ij)\), and \((i,j-1)(i,j-1)\) with zeros elsewhere.
• $A_{2ij}$ is the $IJ \times IJ$ matrix with ones at $(ij)(ji)$, $(ij)(j-1,i)$, $(i,j-1)(ji)$, and $(i,j-1)(j-1,i)$ with zeros elsewhere.

• $B_{1ij}$ is the $IJ \times IJ$ matrix with ones at $(ij)(ij)$, $(i-1,j)(ij)$, $(ij)(i-1,j)$, and $(i-1,j)(i-1,j)$ with zeros elsewhere.

• $B_{2ij}$ is the $IJ \times IJ$ matrix with ones at $(ij)(ji)$, $(i-1,j)(ji)$, $(ij)(j,i-1)$, and $(i-1,j)(j,i-1)$ with zeros elsewhere.

• $D_{1ij}$ be the $IJ \times IJ$ matrix with ones at $(kl)(kl)$ with zeros elsewhere, where $k,k'=i-1$ or $i$, and $l,l'=j-1$ or $j$.

• $D_{2ij}$ be the $IJ \times IJ$ matrix with ones at $(kl)(lk)$ with zeros elsewhere, where $k,k'=i-1$ or $i$, and $l,l'=j-1$ or $j$.

Then, the $K$ matrix is decomposed as

$$K = \sum_{i,j} a_{ij} A_{1ij} + \sum_{i,j} b_{ij} B_{1ij} + \sum_{i,j} d_{ij} D_{1ij} + \sum_{i,j} a_{ij} A_{2ij} + \sum_{i,j} b_{ij} B_{2ij} + \sum_{i,j} d_{ij} D_{2ij},$$

(3.12)

where $a_{ij} = \frac{C_{2ij}}{\eta_{ij}^2}$, $b_{ij} = \frac{C_{1ij}}{\gamma_{ij}^2}$, and $d_{ij} = \frac{F_{ij}}{\pi_{ij}^2}$ are all nonnegative. Since $W$, $A_{mij}$, $B_{mij}$, and $D_{mij}$, for $m = 1,2$, are all nonnegative definite, $K$ is a nonnegative definite matrix, and it is obviously symmetric.

Next let $T$ be the first column of $\sqrt{2} E$, where $E = e_{ij}(kl)$, $e_{ij}(kl) = (-1)^{i+j}(-1)^l \times X_{ij}^{-1}$, where
Note that $E \ast K = M$, where "$\ast$" indicates elementwise multiplication. Then, by defining $z$ as

$$z = T \ast x,$$  

(3.13)

it follows that

$$x'Mx = z'Kz,$$  

(3.14)

for any $x$. Hence, $x'Mx \geq 0$ for any $x$ because $K$ is a nonnegative matrix for. It follows that $M$ is a nonnegative definite matrix, too. For example, suppose we have a 4 by 4 nonnegative definite matrix

$$K_{4 \times 4} = \begin{pmatrix}
a & b & c & d \\
b & e & f & g \\
c & f & h & i \\
d & g & i & j
\end{pmatrix}.$$  

Then, $M$ is

$$M_{4 \times 4} = \begin{pmatrix}
.5a & -b & -c & .5d \\
-b & 2e & 2f & -g \\
-c & 2f & 2h & -i \\
.5d & -g & -i & j
\end{pmatrix}.$$
and elements of $z$ are

$$z_1 = \sqrt{5} x_1 \quad z_2 = -\sqrt{2} x_2$$

$$z_3 = -\sqrt{2} x_3 \quad z_4 = \sqrt{5} x_4$$

in (3.13).

$M$ need not be positive definite. This implies that the loglikelihood function can have flat spots when there are censored observations. Numerical procedures for maximizing the loglikelihood allow for additional searching if a flat spot is encountered to determine if further improvement in maximizing the likelihood can be achieved by leaving the flat spot. If not, an estimate that maximizes the likelihood has been found, but it is not unique. On the other hand, the MLE exists and is unique if the algorithm converges to a point where $M$ is positive definite.
CHAPTER 4. SIMULATION STUDY

In this Chapter we present results of a simulation study for comparing finite sample properties of the product limit (PL) estimator, the actuarial life-table (AL) estimator, and the maximum likelihood estimator (MLE) for marginal survival probabilities developed in Chapter 2 and Chapter 3. Attention is restricted to bivariate survival data. We examine the effects of strength of correlation between responses within pairs, the level of censoring, the number of time intervals, and sample size on the bias and relative efficiency of the estimators and the biases of the modified and unmodified Greenwood formulas for estimating the variance of the PL and AL estimators.

4.1 Simulation of Bivariate Survival Data

Bivariate distributions with identical Weibull marginal distributions were used to generate samples of paired failure times in the simulation study. Among the popular parametric distributions in the analysis of lifetime data, the Weibull distribution is the most widely used model in industrial and biomedical studies. Biomedical applications include analyses of animal carcinogenesis experiments (Peto and Lee, 1973), lung cancer incidence in cigarette smokers (Whittemore and Altschuler, 1976), a heart transplant study (Beck, 1979), and a leukaemia study (Aitkin and Clayton,
The hazard function of the two parameter Weibull distribution is
\[ h(t; \theta, \beta) = \frac{\beta}{\theta} \left( \frac{t}{\theta} \right)^{\beta - 1}, \quad t > 0, \theta > 0, \beta > 0. \]

When the shape parameter \( \beta \) is larger than 1, the hazard function is monotone increasing. The hazard function is monotone decreasing if \( \beta < 1 \), and it is constant for \( \beta = 1 \), the exponential case. In this simulation study, pairs of failure times were generated from a bivariate Weibull distribution where the survivor function for each univariate margin has the form
\[ S(t; \theta, \beta) = \exp \left( -\frac{t}{\theta} \right)^\beta, \quad t > 0, \theta > 0, \beta > 0. \]

Lawless (1982) notes that \( \beta \) values vary across different lifetime situations, but in many applications \( \beta \) is in the range from 1 to 3. Applications of Weibull survival functions in recent studies related to cancer, for example Hashimoto et al. (1992), Horio et al. (1993), and Yamaguchi et al. (1992), yielded estimates of survival curves with shape parameters close to 3. Peto and Lee (1973) obtained estimates of \( \beta \) between 2 and 4 in their analyses of a series of carcinogenesis experiments. Beck (1979), Aitkin and Clayton (1980), Aitkin et al. (1983) and Peto and Lee (1973) considered the exponential distribution (\( \beta = 1 \)) in addition to the Weibull distribution. Consequently, \( \beta \) values of 1 and 3 were selected for our simulation study.

The value of the scale parameter \( \theta \) will depend on the time units of the measured lifetimes. Peto and Lee (1973) obtained estimates of \( \theta \) between 50 and 150 weeks when time was recorded in weeks. Thus, Weibull distributions with \( \beta = 1 \) or 3, and \( \theta = 100 \), were used as marginal distributions in the simulation of bivariate survival times.
Breslow and Crowley (1974) suggest that a uniform distribution can provide a good approximation to the censoring distribution in many cases. They used uniform censoring time distributions in a simulation study of the effect of the number of time intervals on the bias of the AL estimate. Since it often provides unbiased estimates of survival probabilities under uniform withdrawals, the AL estimator is generally preferred over the PL estimator in such situations. Thus, uniform marginal distributions were used to simulate censoring times in this Chapter.

4.1.1 Generating bivariate failure times with Weibull marginal distribution

Pairs of correlated Weibull failure times were simulated as follows. First use the Moran (1967) scheme for generating a pair of correlated exponential random variables, \((Y_1, Y_2)\). Here

\[
Y_1 = \frac{(U_1^2 + U_2^2)}{2}
\]

and

\[
Y_2 = \frac{(U_3^2 + U_4^2)}{2},
\]

(4.1)

where \(U_1, U_2, U_3,\) and \(U_4\) are generated from the Normal distribution with mean 0 and variance 1 so that \((U_1, U_3)\) and \((U_2, U_4)\) are mutually independent, but each pair has a bivariate normal distribution with correlation \(w\). Then, \(Y_1\) and \(Y_2\) are both standard exponential random variables with

\[
E(Y_j) = 1, \quad \text{and} \quad \text{Var}(Y_j) = 1, \quad j = 1, 2
\]

and

\[
\text{Corr}(Y_1, Y_2) = \text{Cov}(Y_1, Y_2) = 1/2 \text{Cov}(U_1^2, U_3^2) = w^2.
\]
Table 4.1: Comparing $\text{Corr}(X_1, X_2)$ and $\text{Corr}(Y_1, Y_2) = w^2$

<table>
<thead>
<tr>
<th>$w^2$</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Corr}(X_1, X_2)$</td>
<td>0.084</td>
<td>0.276</td>
<td>0.451</td>
<td>0.648</td>
<td>0.878</td>
</tr>
</tbody>
</table>

Next, two correlated Weibull random variables, $X_1$ and $X_2$ are obtained as

$$X_1 = \theta Y_1^{1/\beta}$$

and

$$X_2 = \theta Y_2^{1/\beta},$$

where $\theta$ is the scale parameter and $\beta$ is the shape parameter for each of the Weibull margins.

The relationship between $w^2$ and the correlation between $X_1$ and $X_2$ was approximated with an additional set of simulations. For each $w^2$ value, $w^2 = 0.0, 0.1, 0.2, \ldots, 0.9$, we generated 10,000 pairs of $(X_1, X_2)$ values given $\theta = 100$ and $\beta = 3$, to approximate the correlation between $X_1$ and $X_2$. Estimated values of $\text{Corr}(X_1, X_2)$ are plotted against $w^2$ in Figure 4.1 and the estimates of $\text{Corr}(X_1, X_2)$ are presented in Table 4.1.

Using least squares estimation to fit a curve to the plot in Figure 4.1, we get the interpolation formula

$$\text{Corr}(X_1, X_2) = 0.8152w^2 + 0.1848w^4. \quad (4.3)$$

This curve is displayed in Figure 4.1, and it begins at (0,0) and ends at (1,1). Figure 4.1 also contains a 45 degree reference line to show how much Cor$(X_1, X_2)$ deviates from Cor$(Y_1, Y_2) = w^2$ for different values of $w^2$. Only $w^2$ values are reported in
the results for the simulation study and the corresponding values of $\text{Corr}(X_1, X_2)$ can be determined from (4.3).

4.1.2 Generating bivariate censoring times with uniform marginal distribution

Suppose the censoring time variable $W$ has the uniform distribution on the interval $(0, \varpi)$, and the lifetime variable $X$ has the Weibull distribution, $\text{Wei}(\theta, \beta)$. Given $P_c$, the probability that a subject is censored, and values for $\theta$ and $\beta$, we can derive the value of $\varpi$ that provides a specific level of censoring. Since $X$ and $W$ are independent,

$$P_c = \Pr(W - X \leq 0)$$
where \( f(x) \) and \( f(w) \) are pdfs for \( X \) and \( W \), respectively. The approximate value of \( P_c \) (4.5) can be obtained since \( \int_0^\infty f(x) \, dx \) and \( \int_0^\infty f(w) \, dw \) are small for large value of \( \omega \). The accuracy of this approximation is presented in Figure 4.2. Then, the approximate value of \( \omega \) using (4.5) is

\[
\omega = \frac{\theta \Gamma(1 + \beta^{-1})}{P_c}. \tag{4.6}
\]

Pairs of correlated censoring times, \( (W_1, W_2) \) with identical marginal distributions uniformly distributed on \( (0, \omega) \), were also generated using Moran's bivariate distribution with exponential margins. First, generate bivariate exponential random variables \( (Y_1, Y_2) \) as in (4.1). Then, compute

\[
W_1 = \omega(1 - \exp(-Y_1)),
\]

\[
W_2 = \omega(1 - \exp(-Y_2)). \tag{4.7}
\]

Then, \( W_j \) has an uniform distribution on \( (0, \omega) \) for \( j = 1, 2 \). The correlation between \( W_1 \) and \( W_2 \) was approximated by computing the sample correlation of 10,000 simulated pairs for each \( \omega^2 = 0.0, 0.1, 0.2, \ldots, 0.9 \) given \( \theta = 100, \beta = 3 \), and \( P_c = 0.2 \). These are displayed in Figure 4.3 as points on a curve lying below the 45 degree reference line. Estimates of correlations between \( W_1 \) and \( W_2 \), are presented in Table 4.2.
Figure 4.2: Accuracy of approximation (4.5) for the censoring probability when $\beta = 1$ and $\theta = 100$.

Table 4.2: Comparison of $Corr(W_1, W_2)$ and $Corr(Y_1, Y_2) = w^2$

<table>
<thead>
<tr>
<th>$w^2$</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Corr(W_1, W_2)$</td>
<td>0.074</td>
<td>0.258</td>
<td>0.426</td>
<td>0.639</td>
<td>0.876</td>
</tr>
</tbody>
</table>
Figure 4.3: Comparison of $\text{Corr}(W_1, W_2)$ and $\text{Corr}(Y_1, Y_2) = w^2$

This formula is also constrained to force the curve through (0,0) and (1,1). For any $w^2$, the value of $\text{Corr}(W_1, W_2)$ can be determined from (4.8). In the simulation study, we only used the case where $w^2 = 0.9$ to simulate correlated pairs of censoring times.

4.1.3 Simulated data tables

We used the S-plus (1991) function “rnorm” to generate vectors of standard normal random variables. These are used to compute values for $(X_1, X_2)$ in (4.2), and values of $(W_1, W_2)$ in (4.7). Each row of the resulting data matrix has the form

$$\text{Corr}(W_1, W_2) = 0.7344w^2 + 0.2656w^4.$$  (4.8)
\[(\min\{X_1, W_1\}, \delta_1, \min\{X_2, W_2\}, \delta_2), \text{ where } \delta_i = 1 \text{ if } \min\{X_i, W_i\} = X_i, \text{ otherwise } \delta_i = 0.\]

Time was partitioned into intervals using either \(m = 4\) or \(m = 9\) inspection times. If \(m = 4\), the intervals are \((0, 30], (30, 60], (60, 90], (90, 120], (120, \infty)\). If \(m = 9\), the intervals are \((0, 15], (15, 30], (30, 45], \ldots, (105, 120], (120, 135], (135, \infty)\). The survival curves when \(\beta = 1\) and \(\beta = 3\) are shown on Figure 4.4 and 4.5.

Given \(m\), the data are now categorized into the four tables \(F, C_1, C_2, CC\), described in Chapter 3. To simplify the formulas these tables are slightly modified in this Chapter as follows: \(F_{ij}\) is the number of pairs where one unit of the pair fails in \((t_{i-1}, t_i]\) and the other unit fails in \((t_{j-1}, t_j]\), \(C_{1ij}\) is the number of pairs where the first unit of the pairs fails in \((t_{i-1}, t_i]\) and the other unit is censored in \((t_{j-1}, t_j]\). \(C_{2ij}\) is the number of pairs where the first unit is censored in \((t_{j-1}, t_j]\)
and the second unit fails in \((t_{j-1}, t_j]\), \(CC_{ij}\) is the number of pairs where one unit of the pair is censored in \((t_{i-1}, t_i]\) and the other unit is censored in \((t_{j-1}, t_j]\), for \(i, j = 1, 2, \ldots, (m + 1)\) and \(t_0 = 0, t_{(m+1)} = \infty\). The right censored units belonging to the last interval \((t_m, \infty)\) are not counted in either \(C1_{ij}, C2_{ij}\), or \(CC_{ij}\) but they are included in \(F_{ij}\). Thus \(C1_i(m+1) = 0\) for any \(i\), \(C2_j(m+1) = 0\) for any \(j\), \(CC_i(m+1) = 0\) for any \(i\), and \(CC_j(m+1) = 0\) for any \(j\). Thus the total number of pairs is

\[
n = \sum_{i=1}^{m+1} \sum_{j=1}^{m+1} (F_{ij} + C1_{ij} + C2_{ij} + CC_{ij}).
\]

Then, we can classify the simulated times for each pair into one of the four data tables: \(F, C1, C2, \) and \(CC\), as described above. When no distinction is made between

Figure 4.5: Survival probability when \(\beta = 3\) and \(\theta = 100\).
the two members of any pair, the tables are modified as follows:

\[
\begin{align*}
F &= \frac{F + F'}{2} \\
C1 &= \frac{C1 + C2'}{2} \\
C2' &= \frac{C1 + C2'}{2} \\
CC &= \frac{CC + CC'}{2},
\end{align*}
\]

where \( F' \) is the transpose of \( F \). Note that \( C1 \) is identical to \( C2' \) after this modification.

### 4.2 Procedure

In this section, we describe the procedures used to evaluate the PL estimate, AL estimate, and the MLE, and also their variances and covariances, from the four data tables: \( F, C1, C2, \) and \( CC \). Let \( a = m + 1 \). Thus, the time interval boundaries are \( t_0 < t_1 < t_2 < \cdots < t_a \), where \( t_0 = 0 \) and \( t_a = t_{(m+1)} = \infty \). Each table can be considered as an \( a \) by \( a \) matrix, where \( a \) is the number of time intervals. The elements of four tables are expressed as follows: \( F_{ij} = F[i, j], C1_{ij} = C1[i, j], C2_{ij} = C2[i, j], \) and \( CC_{ij} = CC[i, j] \).

#### 4.2.1 PL estimate

Define the elements of two vectors \( D \) and \( C \) as

\[
\begin{align*}
D_i &= \sum_{j=1}^{a} (F[i,j] + F[j,i] + C1[i,j] + C2[i,j]) \\
C_i &= \sum_{j=1}^{a} (C1[i,j] + C2[i,j] + CC[i,j] + CC[j,i]),
\end{align*}
\]
where \( i = 1, 2, 3, \ldots, a \), and \( a \) is the number of intervals. Then \( D_i \) is the number of individuals observed to fail within the \( i^{th} \) interval and \( C_i \) is the number of individuals censored within the \( i^{th} \) interval.

Let \( N_i \) be the number of individuals at risk during the \( i^{th} \) interval, then \( N_i \) is

\[
N_i = N_{i-1} - D_{i-1} - C_{i-1},
\]

where \( n \) is the number of pairs and \( N_1 = 2n \).

Let \( q_i \) be the conditional failure probability and \( P_i \) be the survival probability defined at section 2.2, and denote estimators by

\[
\hat{q}_i = \frac{D_i}{N_i}
\]

and

\[
\hat{P}_i = \prod_{h=1}^{i} (1 - \hat{q}_h),
\]

respectively. \( \hat{P}_i \) is called the PL estimate in this paper.

The following quantities are used to obtain the estimate of the covariance matrix of the \( P_h \)'s described in section 2.2. Define

\[
N_{hh} = \sum_{i=h}^{a} \sum_{j=h}^{a} (F[i, j] + C1[i, j] + C2[i, j] + CC[i, j])
\]

and

\[
N_{hh}' = \sum_{i=h}^{a} \sum_{j=h}^{a} (F[i, j] + C1[i, j] + C2[i, j] + CC[i, j]).
\]

Then,

\[
\hat{q}_h = \frac{D_h}{N_h},
\]

\[
\hat{q}_{hh} = \frac{F[h, h]}{N_{hh}}.
\]
\[
\hat{\theta}_{hh} = \frac{\sum_{i=h}^{a} (F[i, h] + F[h, i] + C1[h, i] + C2[i, h])}{2N_{hh}}
\]

\[
\hat{q}_{hh'} = \frac{F[h, h]}{N_{hh'}} \sum_{j=h'}^{a} (F[j, h] + F[h, j] + C1[h, j] + C2[j, h])
\]

\[
\hat{e}_{1hh'} = \frac{\sum_{i=h}^{a} (F[i, h] + F[h, i] + C1[h, i] + C2[i, h])}{2N_{hh'}}
\]

\[
\hat{e}_{2hh'} = \frac{\sum_{i=h}^{a} (F[i, h] + F[h', i] + C1[h', i] + C2[i, h])}{2N_{hh'}}. \tag{4.16}
\]

In bivariate case, \( \sum_{i=1}^{n} K_{hi} \) of equation (2.5) and \( K_{hh'} \) of equation (2.6) are simplified as follows:

\[
\sum_{i=1}^{n} K_{hi} = \sum_{i=h}^{a} \sum_{j=h}^{a} (F[i, j] + C1[i, j] + C2[i, j] + CC[i, j])
\]

\[
K_{hh'} = \sum_{i=h}^{a} \sum_{j=h'}^{a} (F[i, j] + C1[i, j] + C2[i, j] + CC[i, j])
\]

\[
+ \sum_{i=h}^{a} \sum_{j=h'}^{a} (F[j, i] + C1[j, i] + C2[j, i] + CC[j, i]). \tag{4.17}
\]

Note that \( \sum_{i=1}^{n} K_{hi} \) is equal to \( N_{hh} \) and \( K_{hh'} \) is equal to 2 times of \( N_{hh'} \).

4.2.2 AL estimate

For the AL estimate, we need to adjust \( N_{h}, N_{hh}, N_{hh'}, \sum_{i=1}^{n} K_{hi}, \) and \( K_{hh'} \) for uniform censoring within time intervals. Essentially this considers each censored observation as being at risk for half of the time intervals. The adjusted quantities
These adjusted quantities are substituted into the formulas for the PL estimate and its variances to get the AL estimate and its variance.

4.2.3 Maximum likelihood estimation

The Davidson-Fletcher-Powell (DFP) method can be applied when only the first partial derivative is available, and Newton-Raphson method requires the second partial derivative matrix. These two algorithms were used to find the estimates of the $\theta_{ij}$’s in the log likelihood function (3.6). For 10 time intervals, the time required to compute the second partial derivative matrix makes it more efficient to first use the Davidson-Fletcher-Powell method to get close to the solution and then use a few iteration of the Newton-Raphson method to get the final answer. However for 5 time intervals, it took just a few minutes to get a solution when only the Newton-Raphson method was used.

Equation (3.10) was used to obtain a starting values for the DFP algorithm. They were also used to obtain starting values for the Newton-Raphson method in the case with 5 time intervals. If $F[u,v] = 0$, then $\hat{\theta}_{uv} = -\infty$ in the equation (3.10). In
this case, $-10$ is used for $\hat{\theta}_{uv}$ as a starting value of $\theta_{uv}$.

The maximum of the absolute values of the differences between the $\theta_{ij}$ values at successive iterations was used for the converge criterion of the Newton-Raphson algorithm. For the DFP algorithm, the Euclidean distance between vectors of estimates from successive iterations was used for a stopping rule. The critical value selected for the convergence criterion will affect the accuracy of the final solutions and the computing time. In my simulation program, 0.0001 was selected as the critical value of the convergence criterion for the DFP algorithm and 0.00001 was chosen for the Newton-Raphson method.

4.3 Results

The following results were obtained from 500 simulated data sets for each case. The PL, AL, and maximum likelihood estimates of survival probabilities were derived for each data set, and the means and variances of those estimates were computed. Now define the following quantities:
\[ TP_h = \text{True marginal survival probability at time } t_h \]
\[ MA_h = \text{Average of } 500 \text{ AL estimates of } P_h \]
\[ MP_h = \text{Average of } 500 \text{ PL estimates of } P_h \]
\[ MM_h = \text{Average of } 500 \text{ MLE of } P_h \]
\[ VA_h = \text{Variance of } 500 \text{ AL estimates of } P_h \]
\[ VP_h = \text{Variance of } 500 \text{ PL estimates of } P_h \]
\[ VM_h = \text{Variance of } 500 \text{ MLE of } P_h \]
\[ AvuA_h = \text{Average of } 500 \text{ estimates of } \text{Var(AL estimate)} \]
\[ \text{from the unmodified Greenwood formula} \]
\[ AvuP_h = \text{Average of } 500 \text{ estimates of } \text{Var(PL estimate)} \]
\[ \text{from the unmodified Greenwood formula} \]
\[ AvmA_h = \text{Average of } 500 \text{ estimates of } \text{Var(AL estimate)} \]
\[ \text{from the modified Greenwood formula} \]
\[ AvmP_h = \text{Average of } 500 \text{ estimates of } \text{Var(PL estimate)} \]
\[ \text{from the modified Greenwood formula} \]
\[ AvM_h = \text{Average of } 500 \text{ estimates of } \text{Var(MLE)} \]
\[ \text{from the second partial derivative matrix,} \]

where the unmodified Greenwood formula is based on standard life-table analysis that assumes the responses are all independent of each other.

4.3.1 Bias for the true survival probabilities

The means of the PL estimates, AL estimates, and MLE’s computed from the 500 simulated data sets were compared to the true survival probabilities to check the
bias. Define the average of the bias estimates across the time intervals as

$$mB(x) = \frac{\sum_{h=1}^{m} (Mx_h - TP_h)}{m},$$

(4.19)

where $Mx_h$ can be $MA_h$, $MP_h$, or $MM_h$. A second measure that does not allow positive bias in some time intervals to cancel negative bias from other intervals is

$$mB^2(x) = \frac{\sum_{h=1}^{m} (Mx_h - TP_h)^2}{m}.$$  

(4.20)

Values of $mB(x)$ are shown in Table 4.3 and values of $mB^2(x)$ are shown in Table 4.4 for the MLE, PL, and AL estimates. Table 4.3 and Table 4.4 yield similar conclusions. The biases are most affected by the proportion of censored observations, the number of time intervals, and the value of $\beta$. If the number of subjects is increased, the biases tend to be slightly reduced. The MLE and PL estimates are overestimated under uniform censoring scheme for any case. The AL estimates are slightly overestimated if $\beta = 3$ but slightly underestimated if $\beta = 1$. The biases decrease as the proportion of censored observations decreases. As expected, biases for the MLE, PL, and AL estimates are less for the 10 time interval cases than 5 time interval cases. The biases for the cases with $\beta = 1$ are less than those with $\beta = 3$ for the MLE and PL estimates.

For any case, the AL estimate has the smallest bias, and the biases are relatively close to zero because the simulated censoring observations come from uniform distribution.

The MLE and PL estimates, derived under the same censoring assumptions, exhibit similar behavior. This is seen by comparing the values of $mB(MLE)$ and
mB(PL) in Table 4.3 and comparing the values of \( mB^2(\text{MLE}) \) and \( mB^2(\text{PL}) \) in Table 4.4.

### 4.3.2 Compare \( VA_h \) to \( AvuA_h \) and \( VP_h \) to \( AvuP_h \)

When we ignore the correlations within groups, we may apply the univariate life-table analysis to get estimate of variances of survival probability estimates from Greenwood formula. Averages of ratios of estimates of standard deviations are used to make comparisons. Values of

\[
RU_{sd}(\text{PL}) = \frac{\sum_{h=1}^{m} \sqrt{VP_h}}{m}
\]

are presented in Table 4.5 for each of the 27 cases considered in the simulation study. Values larger than one indicate that both \( AvuP_h \) and \( AvuA_h \) are underestimated by as much as 26\% where \( w^2 = .8 \) and as much as 7\% when \( w^2 = .3 \). Both \( RU_{sd}(\text{PL}) \) and \( RU_{sd}(\text{AL}) \) increase when either \( w^2 \) or the censoring probability are increased, but these ratios are not strongly affected by changes in the sample size, the number of time intervals, or the value of \( \beta \).
4.3.3 Compare $VA_h$ to $AvmA_h$ and $VP_h$ to $AvmP_h$

Simulated estimates of the true standard deviations are compared to the estimates of the standard deviations of the PL and AL estimates obtained from the modified Greenwood formula by presenting values of

$$RMsd(PL) = \frac{\sum_{h=1}^{m} \sqrt{VP_h}}{m}$$

$$RMsd(AL) = \frac{\sum_{h=1}^{m} \sqrt{VA_h}}{m}$$

in Table 4.6 for each of the 27 cases considered in the simulation study.

We did not find any significant effects of sample size, time intervals, censoring level, $w^2$ or $\beta$ on either $RMsd(PL)$ or $RMsd(AL)$. The modified variance estimate tends to be slightly larger than the true variance for moderate sample sizes. Lawless (1982) showed that Greenwood formula tends to slightly overestimate the true variance in the univariate case, as well.

Values of $RMsd(PL)$ and $RMsd(AL)$ are close to one. Therefore, the estimates of variances for the PL and AL estimates provided by the modified Greenwood formula seem to be nearly unbiased.

4.3.4 Relative efficiency of the MLE and PL estimators

Since the MLE is asymptotically efficient and the MLE and the PL estimator use the same assumptions about the censoring distribution, we compare the simulated
variance of the MLE to the simulated variance of the PL estimator to check the efficiency of PL estimator. Define

$$REMP = \frac{\sum_{h=1}^{m} V P_h}{V M_h}.$$  \hspace{1cm} (4.25)

Values of $REMP$ are listed in Table 4.7. None of the factors appear to have a significant effect on $REMP$. Values of $REMP$ are close to one. Thus, it appears that the PL estimator is a quite efficient even for small sample sizes.

4.4 Summary

MLE, PL, and AL estimators were considered to estimate marginal survival probabilities under a uniform censoring scheme for the 27 different cases. The number of time intervals, the censoring probability, and the distribution of failures ($\beta$) affect the biases of all of these estimates. The AL estimator is less biased than either the MLE or the PL estimator. The bias for the MLE and the PL estimator can be reduced by increasing the number of intervals, increasing the number of subjects or reducing the censoring probability, but the bias for the AL estimator tends to be very slightly reduced by those factors.

The true variance of the PL estimator is essentially the same as the true variance of the MLE for each of the 27 cases considered in simulation study. The true variance of AL estimator tends to be slightly larger than the variance of the PL estimator (See Table 4.8). The differences become smaller when either the number of time intervals or the number of pairs is increased. Consequently, we recommend the AL estimator for estimating marginal survival probabilities because it has smaller bias than either the MLE or the PL estimator and it is much easier to compute than the MLE.
The unmodified Greenwood formula substantially underestimates standard deviations in the presence of positive within pair correlations. The estimates of variances of AL and PL estimates provided by the modified Greenwood formula are nearly unbiased. There are no significant factors that affect the estimates from the modified Greenwood formula.
Table 4.3: Comparing $mB(x) \times 10^3$ for 3 estimates

<table>
<thead>
<tr>
<th>MLE</th>
<th>PL</th>
<th>AL</th>
<th>$m + 1$</th>
<th>Cen</th>
<th>$w^2$</th>
<th>$n$</th>
<th>$\beta$</th>
<th>id</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.155</td>
<td>5.156</td>
<td>1.127</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>2.697</td>
<td>2.839</td>
<td>-1.807</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7.103</td>
<td>6.747</td>
<td>2.695</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>50</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>5.615</td>
<td>5.542</td>
<td>0.906</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>5.852</td>
<td>5.842</td>
<td>1.820</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>100</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>1.252</td>
<td>2.025</td>
<td>-2.458</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>100</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>6.088</td>
<td>5.746</td>
<td>1.714</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>50</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>3.692</td>
<td>4.444</td>
<td>-0.058</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>50</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>19.024</td>
<td>18.18</td>
<td>4.377</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>13.342</td>
<td>13.031</td>
<td>-2.303</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>100</td>
<td>1</td>
<td>a</td>
</tr>
<tr>
<td>21.833</td>
<td>20.603</td>
<td>6.829</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>50</td>
<td>3</td>
<td>b</td>
</tr>
<tr>
<td>15.097</td>
<td>14.529</td>
<td>-0.629</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>50</td>
<td>1</td>
<td>c</td>
</tr>
<tr>
<td>19.336</td>
<td>17.492</td>
<td>3.892</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>100</td>
<td>3</td>
<td>d</td>
</tr>
<tr>
<td>11.217</td>
<td>11.493</td>
<td>-3.528</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>100</td>
<td>1</td>
<td>e</td>
</tr>
<tr>
<td>19.868</td>
<td>17.286</td>
<td>3.632</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>50</td>
<td>3</td>
<td>f</td>
</tr>
<tr>
<td>13.139</td>
<td>12.804</td>
<td>-2.177</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>50</td>
<td>1</td>
<td>g</td>
</tr>
<tr>
<td>0.637</td>
<td>0.637</td>
<td>0.637</td>
<td>5</td>
<td>0.0</td>
<td>0.0</td>
<td>100</td>
<td>1</td>
<td>h</td>
</tr>
<tr>
<td>0.706</td>
<td>0.706</td>
<td>0.706</td>
<td>5</td>
<td>0.0</td>
<td>0.0</td>
<td>100</td>
<td>3</td>
<td>i</td>
</tr>
<tr>
<td>5.306</td>
<td>5.302</td>
<td>1.270</td>
<td>5</td>
<td>0.1</td>
<td>0.2</td>
<td>100</td>
<td>3</td>
<td>j</td>
</tr>
<tr>
<td>5.452</td>
<td>5.364</td>
<td>1.337</td>
<td>5</td>
<td>0.1</td>
<td>0.5</td>
<td>100</td>
<td>3</td>
<td>k</td>
</tr>
<tr>
<td>4.709</td>
<td>4.531</td>
<td>0.490</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>200</td>
<td>3</td>
<td>l</td>
</tr>
<tr>
<td>4.284</td>
<td>4.13</td>
<td>0.113</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>200</td>
<td>3</td>
<td>m</td>
</tr>
<tr>
<td>4.543</td>
<td>4.311</td>
<td>0.278</td>
<td>5</td>
<td>0.1</td>
<td>0.5</td>
<td>200</td>
<td>3</td>
<td>n</td>
</tr>
<tr>
<td>1.719</td>
<td>1.620</td>
<td>-0.227</td>
<td>10</td>
<td>0.1</td>
<td>0.3</td>
<td>200</td>
<td>3</td>
<td>o</td>
</tr>
<tr>
<td>1.669</td>
<td>1.606</td>
<td>-0.239</td>
<td>10</td>
<td>0.1</td>
<td>0.8</td>
<td>200</td>
<td>3</td>
<td>p</td>
</tr>
<tr>
<td>2.163</td>
<td>2.295</td>
<td>0.454</td>
<td>10</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>q</td>
</tr>
<tr>
<td>1.775</td>
<td>1.633</td>
<td>-0.214</td>
<td>10</td>
<td>0.1</td>
<td>0.5</td>
<td>200</td>
<td>3</td>
<td>r</td>
</tr>
</tbody>
</table>

$m + 1$: Number of intervals including $\infty$

Cen: Censoring probability

$w^2$: Correlation between two failure times

$n$: Number of pairs

$\beta$: Shape parameter of Weibull distribution

id: Id for each case.
Table 4.4: Comparing $mB^2(x) \times 10^6$ for 3 estimates

<table>
<thead>
<tr>
<th></th>
<th>MLE</th>
<th>PL</th>
<th>AL</th>
<th>$m + 1$</th>
<th>Cen</th>
<th>$w^2$</th>
<th>$n$</th>
<th>$\beta$</th>
<th>id</th>
</tr>
</thead>
<tbody>
<tr>
<td>36.758</td>
<td>36.949</td>
<td>1.906</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>7.542</td>
<td>8.318</td>
<td>4.512</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>70.782</td>
<td>63.462</td>
<td>10.790</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>50</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>33.433</td>
<td>32.468</td>
<td>4.386</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>50</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>44.543</td>
<td>44.281</td>
<td>3.924</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>100</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>3.454</td>
<td>4.821</td>
<td>8.746</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>100</td>
<td>1</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>55.340</td>
<td>49.303</td>
<td>5.537</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>50</td>
<td>3</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>14.155</td>
<td>20.442</td>
<td>0.389</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>50</td>
<td>1</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>517.517</td>
<td>469.921</td>
<td>28.409</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>180.415</td>
<td>171.944</td>
<td>14.336</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>100</td>
<td>1</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>660.981</td>
<td>582.554</td>
<td>60.499</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>50</td>
<td>3</td>
<td>b</td>
<td></td>
</tr>
<tr>
<td>234.723</td>
<td>217.575</td>
<td>15.269</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>50</td>
<td>1</td>
<td>c</td>
<td></td>
</tr>
<tr>
<td>516.422</td>
<td>420.200</td>
<td>19.272</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>100</td>
<td>3</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>135.527</td>
<td>139.783</td>
<td>34.605</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>100</td>
<td>1</td>
<td>e</td>
<td></td>
</tr>
<tr>
<td>568.735</td>
<td>425.696</td>
<td>20.077</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>50</td>
<td>3</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>177.540</td>
<td>169.234</td>
<td>17.517</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>50</td>
<td>1</td>
<td>g</td>
<td></td>
</tr>
<tr>
<td>1.415</td>
<td>1.415</td>
<td>1.415</td>
<td>5</td>
<td>0.0</td>
<td>0.0</td>
<td>100</td>
<td>1</td>
<td>h</td>
<td></td>
</tr>
<tr>
<td>0.562</td>
<td>0.562</td>
<td>0.562</td>
<td>5</td>
<td>0.0</td>
<td>0.0</td>
<td>100</td>
<td>3</td>
<td>i</td>
<td></td>
</tr>
<tr>
<td>39.555</td>
<td>39.577</td>
<td>2.478</td>
<td>5</td>
<td>0.1</td>
<td>0.2</td>
<td>100</td>
<td>3</td>
<td>j</td>
<td></td>
</tr>
<tr>
<td>40.199</td>
<td>39.094</td>
<td>2.593</td>
<td>5</td>
<td>0.1</td>
<td>0.5</td>
<td>100</td>
<td>3</td>
<td>k</td>
<td></td>
</tr>
<tr>
<td>33.256</td>
<td>30.804</td>
<td>0.893</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>200</td>
<td>3</td>
<td>l</td>
<td></td>
</tr>
<tr>
<td>28.720</td>
<td>27.138</td>
<td>0.707</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>200</td>
<td>3</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>30.531</td>
<td>27.580</td>
<td>0.503</td>
<td>5</td>
<td>0.1</td>
<td>0.5</td>
<td>200</td>
<td>3</td>
<td>n</td>
<td></td>
</tr>
<tr>
<td>4.763</td>
<td>4.246</td>
<td>0.151</td>
<td>10</td>
<td>0.1</td>
<td>0.3</td>
<td>200</td>
<td>3</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>4.293</td>
<td>4.021</td>
<td>0.279</td>
<td>10</td>
<td>0.1</td>
<td>0.8</td>
<td>200</td>
<td>3</td>
<td>p</td>
<td></td>
</tr>
<tr>
<td>6.564</td>
<td>7.432</td>
<td>0.347</td>
<td>10</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>q</td>
<td></td>
</tr>
<tr>
<td>4.707</td>
<td>4.018</td>
<td>0.265</td>
<td>10</td>
<td>0.1</td>
<td>0.5</td>
<td>200</td>
<td>3</td>
<td>r</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.5: Comparison of $VA_h$ to $AvuA_h$ and $VP_h$ to $AvuP_h$

<table>
<thead>
<tr>
<th>$RU_{sd}(PL)$</th>
<th>$RU_{sd}(AL)$</th>
<th>$m + 1$</th>
<th>Cen</th>
<th>$w^2$</th>
<th>$n$</th>
<th>$\beta$</th>
<th>id</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0410</td>
<td>1.0416</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>1.0696</td>
<td>1.0693</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1.0754</td>
<td>1.0728</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>50</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1.0387</td>
<td>1.0413</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>50</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>1.1918</td>
<td>1.1929</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>100</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>1.2637</td>
<td>1.2632</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>100</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>1.2063</td>
<td>1.2050</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>50</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>1.2295</td>
<td>1.2300</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>50</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>1.0530</td>
<td>1.0456</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>1.0472</td>
<td>1.0444</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>100</td>
<td>1</td>
<td>a</td>
</tr>
<tr>
<td>1.0793</td>
<td>1.0706</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>50</td>
<td>3</td>
<td>b</td>
</tr>
<tr>
<td>1.0411</td>
<td>1.0400</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>50</td>
<td>1</td>
<td>c</td>
</tr>
<tr>
<td>1.1997</td>
<td>1.1970</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>100</td>
<td>3</td>
<td>d</td>
</tr>
<tr>
<td>1.2320</td>
<td>1.2355</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>100</td>
<td>1</td>
<td>e</td>
</tr>
<tr>
<td>1.1812</td>
<td>1.1751</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>50</td>
<td>3</td>
<td>f</td>
</tr>
<tr>
<td>1.2120</td>
<td>1.2145</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>50</td>
<td>1</td>
<td>g</td>
</tr>
<tr>
<td>0.9937</td>
<td>0.9937</td>
<td>5</td>
<td>0.0</td>
<td>0.0</td>
<td>100</td>
<td>1</td>
<td>h</td>
</tr>
<tr>
<td>1.0204</td>
<td>1.0204</td>
<td>5</td>
<td>0.0</td>
<td>0.0</td>
<td>100</td>
<td>3</td>
<td>i</td>
</tr>
<tr>
<td>1.0283</td>
<td>1.0283</td>
<td>5</td>
<td>0.1</td>
<td>0.2</td>
<td>100</td>
<td>3</td>
<td>j</td>
</tr>
<tr>
<td>1.0878</td>
<td>1.0892</td>
<td>5</td>
<td>0.1</td>
<td>0.5</td>
<td>100</td>
<td>3</td>
<td>k</td>
</tr>
<tr>
<td>1.0272</td>
<td>1.0255</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>200</td>
<td>3</td>
<td>l</td>
</tr>
<tr>
<td>1.1455</td>
<td>1.1441</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>200</td>
<td>3</td>
<td>m</td>
</tr>
<tr>
<td>1.0675</td>
<td>1.0659</td>
<td>5</td>
<td>0.1</td>
<td>0.5</td>
<td>200</td>
<td>3</td>
<td>n</td>
</tr>
<tr>
<td>1.0477</td>
<td>1.0470</td>
<td>10</td>
<td>0.1</td>
<td>0.3</td>
<td>200</td>
<td>3</td>
<td>o</td>
</tr>
<tr>
<td>1.1734</td>
<td>1.1729</td>
<td>10</td>
<td>0.1</td>
<td>0.8</td>
<td>200</td>
<td>3</td>
<td>p</td>
</tr>
<tr>
<td>1.0888</td>
<td>1.0896</td>
<td>10</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>q</td>
</tr>
<tr>
<td>1.0984</td>
<td>1.0979</td>
<td>10</td>
<td>0.1</td>
<td>0.5</td>
<td>200</td>
<td>3</td>
<td>r</td>
</tr>
</tbody>
</table>
Table 4.6: Comparison of $V A_{h}$ to $AvmA_{h}$ and $VP_{h}$ to $AvmP_{h}$

<table>
<thead>
<tr>
<th>$RMsd(PL)$</th>
<th>$RMsd(AL)$</th>
<th>$m + 1$</th>
<th>Cen</th>
<th>$w^2$</th>
<th>$n$</th>
<th>$\beta$</th>
<th>id</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9940</td>
<td>0.9976</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>1.0105</td>
<td>1.0126</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1.0236</td>
<td>1.0239</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>50</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0.9805</td>
<td>0.9852</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>50</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>1.0152</td>
<td>1.0173</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>100</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>1.0363</td>
<td>1.0365</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>100</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>1.0247</td>
<td>1.0245</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>50</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>1.0093</td>
<td>1.0106</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>50</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>1.0076</td>
<td>1.0085</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>0.9988</td>
<td>1.0036</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>100</td>
<td>1</td>
<td>a</td>
</tr>
<tr>
<td>1.0306</td>
<td>1.0300</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>50</td>
<td>3</td>
<td>b</td>
</tr>
<tr>
<td>0.9933</td>
<td>0.9996</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>50</td>
<td>1</td>
<td>c</td>
</tr>
<tr>
<td>1.0345</td>
<td>1.0363</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>100</td>
<td>3</td>
<td>d</td>
</tr>
<tr>
<td>1.0264</td>
<td>1.0322</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>100</td>
<td>1</td>
<td>e</td>
</tr>
<tr>
<td>1.0141</td>
<td>1.0108</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>50</td>
<td>3</td>
<td>f</td>
</tr>
<tr>
<td>1.0113</td>
<td>1.0165</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>50</td>
<td>1</td>
<td>g</td>
</tr>
<tr>
<td>0.9939</td>
<td>0.9973</td>
<td>5</td>
<td>0.1</td>
<td>0.2</td>
<td>100</td>
<td>3</td>
<td>j</td>
</tr>
<tr>
<td>0.9966</td>
<td>1.0000</td>
<td>5</td>
<td>0.1</td>
<td>0.5</td>
<td>100</td>
<td>3</td>
<td>k</td>
</tr>
<tr>
<td>0.9798</td>
<td>0.9803</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>200</td>
<td>3</td>
<td>l</td>
</tr>
<tr>
<td>0.9732</td>
<td>0.9730</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>200</td>
<td>3</td>
<td>m</td>
</tr>
<tr>
<td>0.9781</td>
<td>0.9784</td>
<td>5</td>
<td>0.1</td>
<td>0.5</td>
<td>200</td>
<td>3</td>
<td>n</td>
</tr>
<tr>
<td>0.9882</td>
<td>0.9885</td>
<td>10</td>
<td>0.1</td>
<td>0.3</td>
<td>200</td>
<td>3</td>
<td>o</td>
</tr>
<tr>
<td>0.9887</td>
<td>0.9886</td>
<td>10</td>
<td>0.1</td>
<td>0.8</td>
<td>200</td>
<td>3</td>
<td>p</td>
</tr>
<tr>
<td>1.0077</td>
<td>1.0132</td>
<td>10</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>q</td>
</tr>
<tr>
<td>0.9904</td>
<td>0.9908</td>
<td>10</td>
<td>0.1</td>
<td>0.5</td>
<td>200</td>
<td>3</td>
<td>r</td>
</tr>
<tr>
<td>0.9683</td>
<td>0.9714</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>500</td>
<td>3</td>
<td>s</td>
</tr>
<tr>
<td>0.9892</td>
<td>0.9881</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>1000</td>
<td>3</td>
<td>t</td>
</tr>
<tr>
<td>0.9850</td>
<td>0.9828</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>500</td>
<td>3</td>
<td>u</td>
</tr>
<tr>
<td>0.9883</td>
<td>0.9867</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>1000</td>
<td>3</td>
<td>v</td>
</tr>
</tbody>
</table>
Table 4.7: Relative efficiency of the MLE to the PL estimator

<table>
<thead>
<tr>
<th>RE</th>
<th>m + 1</th>
<th>Cen</th>
<th>w²</th>
<th>n</th>
<th>β</th>
<th>id</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9966</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>0.9981</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0.9991</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>50</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1.0023</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>50</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>1.0069</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>100</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>0.9979</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>100</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>0.9975</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>50</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>0.9984</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>50</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>0.9969</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>0.9938</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>100</td>
<td>1</td>
<td>a</td>
</tr>
<tr>
<td>0.9873</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>50</td>
<td>3</td>
<td>b</td>
</tr>
<tr>
<td>1.0035</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>50</td>
<td>1</td>
<td>c</td>
</tr>
<tr>
<td>1.0061</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>100</td>
<td>3</td>
<td>d</td>
</tr>
<tr>
<td>1.0053</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>100</td>
<td>1</td>
<td>e</td>
</tr>
<tr>
<td>0.9863</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>50</td>
<td>3</td>
<td>f</td>
</tr>
<tr>
<td>1.0015</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>50</td>
<td>1</td>
<td>g</td>
</tr>
<tr>
<td>1.0000</td>
<td>5</td>
<td>0.0</td>
<td>0.0</td>
<td>100</td>
<td>1</td>
<td>h</td>
</tr>
<tr>
<td>1.0000</td>
<td>5</td>
<td>0.0</td>
<td>0.0</td>
<td>100</td>
<td>3</td>
<td>i</td>
</tr>
<tr>
<td>0.9959</td>
<td>5</td>
<td>0.1</td>
<td>0.2</td>
<td>100</td>
<td>3</td>
<td>j</td>
</tr>
<tr>
<td>1.0033</td>
<td>5</td>
<td>0.1</td>
<td>0.5</td>
<td>100</td>
<td>3</td>
<td>k</td>
</tr>
<tr>
<td>0.9994</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>200</td>
<td>3</td>
<td>l</td>
</tr>
<tr>
<td>1.0006</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>200</td>
<td>3</td>
<td>m</td>
</tr>
<tr>
<td>0.9985</td>
<td>5</td>
<td>0.1</td>
<td>0.5</td>
<td>200</td>
<td>3</td>
<td>n</td>
</tr>
<tr>
<td>1.0007</td>
<td>10</td>
<td>0.1</td>
<td>0.3</td>
<td>200</td>
<td>3</td>
<td>o</td>
</tr>
<tr>
<td>1.0027</td>
<td>10</td>
<td>0.1</td>
<td>0.8</td>
<td>200</td>
<td>3</td>
<td>p</td>
</tr>
<tr>
<td>0.9972</td>
<td>10</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>q</td>
</tr>
<tr>
<td>0.9974</td>
<td>10</td>
<td>0.1</td>
<td>0.5</td>
<td>200</td>
<td>3</td>
<td>r</td>
</tr>
</tbody>
</table>
Table 4.8: Comparison of true standard deviation of PL to AL

<table>
<thead>
<tr>
<th>( \sqrt{V P_h} )</th>
<th>( \sqrt{V A_h} )</th>
<th>( \frac{\sqrt{V P_h}}{\sqrt{V A_h}} )</th>
<th>( m + 1 )</th>
<th>Cen</th>
<th>( w^2 )</th>
<th>( n )</th>
<th>( \beta )</th>
<th>( \text{id} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0273</td>
<td>0.0274</td>
<td>0.9951</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>0.0361</td>
<td>0.0362</td>
<td>0.9968</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0.0393</td>
<td>0.0394</td>
<td>0.9988</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>50</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0.0495</td>
<td>0.0498</td>
<td>0.9940</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>50</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>0.0315</td>
<td>0.0317</td>
<td>0.9950</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>100</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>0.0426</td>
<td>0.0427</td>
<td>0.9971</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>100</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>0.0447</td>
<td>0.0448</td>
<td>0.9976</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>50</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>0.0585</td>
<td>0.0587</td>
<td>0.9961</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>50</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>0.0297</td>
<td>0.0298</td>
<td>0.9952</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>0.0368</td>
<td>0.0371</td>
<td>0.9906</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>100</td>
<td>1</td>
<td>a</td>
</tr>
<tr>
<td>0.0423</td>
<td>0.0424</td>
<td>0.9971</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>50</td>
<td>3</td>
<td>b</td>
</tr>
<tr>
<td>0.0516</td>
<td>0.0521</td>
<td>0.9895</td>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>50</td>
<td>1</td>
<td>c</td>
</tr>
<tr>
<td>0.0341</td>
<td>0.0345</td>
<td>0.9893</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>100</td>
<td>3</td>
<td>d</td>
</tr>
<tr>
<td>0.0431</td>
<td>0.0437</td>
<td>0.9855</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>100</td>
<td>1</td>
<td>e</td>
</tr>
<tr>
<td>0.0470</td>
<td>0.0472</td>
<td>0.9946</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>50</td>
<td>3</td>
<td>f</td>
</tr>
<tr>
<td>0.0598</td>
<td>0.0606</td>
<td>0.9864</td>
<td>5</td>
<td>0.3</td>
<td>0.8</td>
<td>50</td>
<td>1</td>
<td>g</td>
</tr>
<tr>
<td>0.0267</td>
<td>0.0269</td>
<td>0.9956</td>
<td>5</td>
<td>0.1</td>
<td>0.2</td>
<td>100</td>
<td>3</td>
<td>j</td>
</tr>
<tr>
<td>0.0285</td>
<td>0.0287</td>
<td>0.9946</td>
<td>5</td>
<td>0.1</td>
<td>0.5</td>
<td>100</td>
<td>3</td>
<td>k</td>
</tr>
<tr>
<td>0.0192</td>
<td>0.0192</td>
<td>0.9980</td>
<td>5</td>
<td>0.1</td>
<td>0.3</td>
<td>200</td>
<td>3</td>
<td>l</td>
</tr>
<tr>
<td>0.0216</td>
<td>0.0216</td>
<td>0.9972</td>
<td>5</td>
<td>0.1</td>
<td>0.8</td>
<td>200</td>
<td>3</td>
<td>m</td>
</tr>
<tr>
<td>0.0200</td>
<td>0.0200</td>
<td>0.9977</td>
<td>5</td>
<td>0.1</td>
<td>0.5</td>
<td>200</td>
<td>3</td>
<td>n</td>
</tr>
<tr>
<td>0.0180</td>
<td>0.0180</td>
<td>0.9998</td>
<td>10</td>
<td>0.1</td>
<td>0.3</td>
<td>200</td>
<td>3</td>
<td>o</td>
</tr>
<tr>
<td>0.0203</td>
<td>0.0203</td>
<td>0.9995</td>
<td>10</td>
<td>0.1</td>
<td>0.8</td>
<td>200</td>
<td>3</td>
<td>p</td>
</tr>
<tr>
<td>0.0256</td>
<td>0.0257</td>
<td>0.9980</td>
<td>10</td>
<td>0.1</td>
<td>0.3</td>
<td>100</td>
<td>3</td>
<td>q</td>
</tr>
<tr>
<td>0.0187</td>
<td>0.0187</td>
<td>0.9994</td>
<td>10</td>
<td>0.1</td>
<td>0.5</td>
<td>200</td>
<td>3</td>
<td>r</td>
</tr>
</tbody>
</table>
CHAPTER 5. TESTS FOR EQUALITY OF TWO DISTRIBUTIONS WITH LIFE-TABLE MODEL

5.1 Introduction

There are several ways to test the equality of two survival distributions, which is often an important issue in the analysis of medical data and in industrial reliability studies. Gehan (1965) adapted the Wilcoxon test to censored data. Cox (1972) considered a two-sample problem with the proportional hazard model and derived an asymptotic two sample test statistic based on a score function. It is similar to test statistics (Mantel and Haenzel, 1959; Mantel, 1963) obtained by placing each failed unit in a $2 \times 2$ contingency table. These statistics are applicable to ranked continuous survival time data assuming no ties in data.

Kalbfleisch and Prentice (1973) developed a discrete model for interval censored data to solve the problem with ties in the Cox model. Thompson (1977) also studied methodology for grouping observations and a logistic model was introduced using Cox's (1970) binary data methods. The logistic model leads back to the Cox (1972) model as the grouping interval lengths approach zero.

Lininger et al. (1979) considered situations where the experimental units are randomly assigned to one of two treatments within strata which are formed by values of covariates. They compared four test statistics, Gehan, Mantel-Haenzel, continu-
ity corrected Mantel-Haenzel, and Cox's test through a Monte Carlo study. The continuity corrected Mantel-Haenzel statistic was recommended by Mantel (1966). Note that the Mantel-Haenzel statistic was also called the log rank statistic by Peto and Peto (1972). The Gehan and Mantel-Haenzel statistics tended to have type I error levels closest to the nominal level $\alpha = 0.05$, the continuity corrected Mantel-Haenzel statistic was consistently conservative, with $\alpha$ often less than 0.035. The Gehan statistic was only slightly less powerful than the other statistics. The power of these statistics mainly depended on the total number of failures observed and was otherwise little affected by the degree of censoring or the number of strata. Lininger et al. (1979) suggest that the Cox statistic may be preferable for large data sets with many regression variables.

In small samples, Lininger et al. (1979) concluded that the Cox statistic is not so appropriate as a Mantel-Haenzel statistic because the type I error level for the Cox statistic is not as close to the nominal level $\alpha = 0.05$ as the type I error levels for the Mantel-Haenzel statistics. Farewell and Dahlberg (1983) showed that this result is due to the need to estimate a large number of nuisance parameters in the approach based on the Cox model. Thus, Farewell and Dahlberg (1983) suggest another method that does not require such estimation of nuisance parameters and they show that the modified model perform similarly to a Mantel-Haenzel statistic in small samples. All of these exhibit similar performance in large samples. Little work has been done on comparison of survivor functions using non-parametric estimates involving correlated response times.

Methods for comparing non-parametric estimates of two survivor functions obtained from interval censored data are developed in this Chapter. In section 5.2, we
review methods for the case where all response times are independent. Two different situations for the correlated response times can be considered: First, there are two independent sets of subjects, each set has independent groups and there are correlated individuals within groups. In this situation, each set of subjects is given a different treatment or each set of subjects is distinguished by a different level of a covariate such as sex or a certain medical condition. This case is discussed in section 5.3 and an example is given in section 5.4. The other case is where individuals within groups are randomly assigned to two different treatments. This case is briefly considered in section 5.5.

5.2 Univariate Independent Response Times

The simplest way to test the equality of two survivor functions uses the asymptotic normality for the estimates of the survival probabilities to obtain a Wald test as shown in section 5.3. Other methods presented by Lawless (1982) are reviewed in this section. The methods we consider here are extensions of univariate life-table methods to include a dummy regression variable. We define a dummy regression variable $x$ that has the values 0 or 1 according to whether an individual comes from the first or second population. We now define the following quantities:
\( N_{ih} = \) Number of observations "at risk" at time \( t_{h-1} \) from the \( i^{th} \) population,

\( D_{ih} = \) Number of failures in \((t_{h-1}, t_h]\) from the \( i^{th} \) population,

\( C_{ih} = \) Number of withdrawals in \((t_{h-1}, t_h]\) from the \( i^{th} \) population,

\( P_{ih} = \) Pr(an individual survives past \( t_h \) | the individual is from the \( i^{th} \) population),

\[ P_{ih} = \frac{P_{ih}}{P_{i(h-1)}} \]

\[ = \text{Pr(an individual survives past } t_h \text{ | the individual survives past } t_{h-1} \text{ and the individual is from the } i^{th} \text{ population).} \]

For the time being we assume that the censoring events \( C_{ih} \) can only occur at the end of the time interval \((t_{h-1}, t_h]\). The observed number of failures in \((t_{h-1}, t_h]\), \( D_{ih} \) are binomially distributed random variables with parameters \((N_{ih}, 1 - p_{ih})\). Under these assumptions, the likelihood function is proportional to

\[
\prod_{h=1}^{m} \left\{ [1 - p_{1h}]^{D_{1h}} p_{1h} N_{1h} [1 - p_{2h}]^{D_{2h}} p_{2h} N_{2h} \right\}. \quad (5.1)
\]

Let \( N_h \) be the risk set at time \( t_{h-1} \), \( D_h \) be the set of individuals observed to fail in \((t_{h-1}, t_h]\), and \( C_h \) be the set of individuals censored in \((t_{h-1}, t_h]\). Then (5.1) can be expressed as

\[
\prod_{h=1}^{m} \left\{ \prod_{l \in D_h} [1 - p_h(x_l)] \prod_{l \in N_h - D_h} p_h(x_l) \right\}, \quad (5.2)
\]

where \( p_h(0) = p_{1h} \) and \( p_h(1) = p_{2h} \). Brown (1983) also used the binomial distribution of \( D_{ih} \) to derive a likelihood function.
Cox (1972) suggested the proportional hazards regression model with conditional hazard function

$$h(t \mid x) = h_0(t)\exp(x\beta), \quad (5.3)$$

and conditional survivor function

$$S(t \mid x) = [S_0(t)]\exp(x\beta), \quad (5.4)$$

where $x$ is a vector of regression variables, $\beta$ is a vector of regression coefficients, and the baseline survivor function is $S_0(t) = \exp\left\{-\int_0^t h_0(u)\,du\right\}$. Then (5.2) can be obtained from (5.3) by assuming that the lifetime of an individual with regression variable $x$ comes from a proportional hazard model. From (5.4),

$$P_h(x) = \Pr(\text{an individual survives past } t_h \mid x)$$

$$= S(t_h \mid x)$$

$$= S_0(t_h)\exp(x\beta) = P_h^\exp(x\beta). \quad (5.5)$$

Then,

$$p_h(x) = \frac{P_h(x)}{P_{h-1}(x)} = \frac{\exp(x\beta)}{P_h}, \quad (5.6)$$

where $p_h = p_h(0)$. Then the likelihood function (5.2) becomes

$$\prod_{h=1}^m \left\{ \prod_{l \in D_h} [1 - p_h \exp(x\beta)] \frac{\prod_{l \in N_h - D_h} p_h \exp(x\beta)} {1 + r_h \exp(x\beta)} \right\}. \quad (5.7)$$

This approach was introduced by Kalbfleisch and Prentice (1973). Let $r_h$ be $\frac{1-p_h}{p_h}$, then

$$p_h(x) = [1 + r_h \exp(x\beta)]^{-1}. \quad (5.8)$$
Note that this is a logistic model. Lawless (1982) notes that this model is flexible and convenient, as logistic models are in many other discrete data situations. Prentice and Gloeckler (1978) used a \( \log(-\log p_h) \) reparameterization.

A test that survival probabilities are the same for the two populations can be based on a score function. From the logistic model,

\[
\begin{align*}
    p_{1h} &= p_h(0) = (1 + r_h)^{-1}, \\
    p_{2h} &= p_h(1) = (1 + r_h e^\beta)^{-1}.
\end{align*}
\]

Therefore, the null hypothesis \( H_0 : p_{1h} = p_{2h} \) is equivalent to \( H_0 : \beta = 0 \). Under \( H_0 : \beta = 0 \), the test statistic derived from the score function is

\[
Z = \frac{U(0)}{[I(0)]^{0.5}},
\]

where \( U(0) \) is the first partial derivative of the loglikelihood function with respect to \( \beta \), evaluated at \( \beta = 0 \), and \( I(0) \) is the second partial derivative evaluated at \( \beta = 0 \). Under \( H_0 \), \( Z^2 \) is approximately distributed as a \( \chi^2(1) \) random variable for large sample size.

An alternative method is to replace (5.2) by a partial likelihood suggested by Cox (1972). The partial likelihood is

\[
\prod_{h=1}^{m} \left\{ \frac{\exp(s_h \beta)}{\left( \sum_{l \in N_h} \exp(x_l \beta) \right)^{d_h}} \right\},
\]

where \( d_h \) is the number of individuals observed to fail in \((t_{h-1}, t_h]\) and \( s_h = \sum_{l \in D_h} x_l \). Using the score function based on (5.12), we can test the equality of
two distributions. Lawless (1982) said that the two models are nearly equivalent and the tests are virtually identical when intervals are short.

In the previous procedures, we have assumed that all censoring takes place at the ends of the time intervals. If censoring times are uniformly distributed within time intervals, Thompson (1977) suggests the following procedure. Define $G_h = N_h - D_h - C_h$. Then the likelihood (5.2) is replaced by

$$\prod_{h=1}^{m} \left\{ \prod_{l \in D_h} [1 - p_h(x_l)] \prod_{l \in G_h} p_h(x_l) \prod_{l \in C_h} p_h(x_l)^{0.5} \right\}, \quad (5.13)$$

and the loglikelihood of (5.13) is

$$\sum_{h=1}^{m} \left\{ \sum_{l \in D_h} [1 - p_h(x_l)] + \sum_{l \in G_h} p_h(x_l) + 0.5 \sum_{l \in C_h} p_h(x_l) \right\}. \quad (5.14)$$

This implies that the contribution to the loglikelihood of censored individuals in $(t_{h-1}, t_h]$ is halved. Under $H_0 : p_h(0) = p_h(1)$, the maximization of (5.14) gives the estimates $\hat{p}_h = 1 - \frac{d_h}{n_h - 0.5c_h}$, which is the AL estimate, where $d_h$, $n_h$, and $c_h$ are the number of individuals in $D_h$, $N_h$, and $C_h$, respectively.

### 5.3 Correlated Response Times where Treatments are applied to Different Independent Sets of Cohorts

Suppose that there are $n$ independent groups, and some of the groups are randomly assigned to one treatment, while the others are assigned to a second treatment. Within each group, the individuals may give correlated responses. Instead of exact response times, we have interval censored data. In this situation, it is often required to compare the effects of treatments. For this reason, we sometimes need to test the equality of two survivor functions.
The methods introduced in section 5.1 cannot be applied in this situation because they are based on the independent responses. However, in Chapter 2 we showed the asymptotic normality of PL and AL estimates for this situation. Using this property we can compare two survival functions. Let \( P_1 \) be the vector of true survival probabilities under the first treatment and let \( P_2 \) be the corresponding vector of survival probabilities under the second treatment. \( \hat{P}_1 \) and \( \hat{P}_2 \) denote vectors of either the PL or AL estimates from Chapter 2. Covariance matrices of the \( \hat{P}_i \)'s, \( i = 1, 2 \), are derived in section 2.2.3 of Chapter 2. Some quantities used to estimate covariances of the PL estimates are introduced in section 4.2.1 and those for the AL estimates are introduced in section 4.2.2 of Chapter 4.

Let \( V(\hat{P}_i) \) denote the covariance matrix of \( \hat{P}_i \) derived in section 2.2.3.4 of Chapter 2. Then a test \( H_0 : P_1 - P_2 = 0 \) is derived from a Wald statistic as follows. Since \( \hat{P}_1 - \hat{P}_2 \) has an asymptotic normal distribution with mean 0 and covariance matrix \( \Sigma = V(\hat{P}_1) + V(\hat{P}_2) \) under \( H_0 \), the test statistic

\[
Z^2 = (\hat{P}_1 - \hat{P}_2)' \hat{\Sigma}^{-1} (\hat{P}_1 - \hat{P}_2),
\]

(5.15)

where \( \hat{\Sigma} \) is the estimator for \( \Sigma \) computed from the estimates given in section 4.2.1 for the PL estimates of \( P_1 \) and \( P_2 \), and it is computed from estimates given in section 4.2.2 for the AL estimates of \( P_1 \) and \( P_2 \). \( \hat{\Sigma} \) will be a consistent estimator if the appropriate assumptions about the censoring distribution are satisfied.

Then, when \( H_0 \) is true, \( Z^2 \) has a \( \chi^2 \) distribution with \( m \) degree of freedom, where \( m \) is the rank of \( \hat{\Sigma} \).
5.4 Example for Section 5.3

The angioplasty data analyzed in section 2.4 can be separated into two independent sets by any one of four variables: gender (male or female), presence or absence of diabetes mellitus (DM), occurrence or lack of occurrence of a previous myocardial infarction (MI), or presence or absence of hypertension (HTN). We will make separate comparisons of survivor functions for each of these four factors.

In Figures 5.1 - 5.4, PL estimates of two survival curves are shown for each of the gender, diabetes mellitus, myocardial infarction, and hypertension factors. Figures 5.5 - 5.8 show corresponding curves by AL estimates of survival curves. Although the Figures suggest different survival curves for two levels of factor, tests based on (5.15) shown in Tables 5.1 and 5.2 reveal no significant difference for any of the four factors. This is a result of the high failure rates in the first 9 months and the high levels of censoring. Consequently, there are relatively few individuals at risk after months where the survivor curves appear to be most different, and variances of the estimated survival probabilities are large.

As expected, values of the test statistics in Tables 5.1 and 5.2 are smaller when variances and covariances are estimated with modified Greenwood formulas than the unmodified formulas. The differences between using modified and unmodified Greenwood estimates are not large in this case because of the large proportions of singletons in this study.
Table 5.1: $\chi^2$ test for equality of two survivor functions using PL estimates

<table>
<thead>
<tr>
<th>Test</th>
<th>Gender</th>
<th>DM</th>
<th>MI</th>
<th>HTN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmodified</td>
<td>$\chi^2$</td>
<td>12.94</td>
<td>10.92</td>
<td>11.43</td>
</tr>
<tr>
<td>p-value</td>
<td></td>
<td>(.114)</td>
<td>(.206)</td>
<td>(.179)</td>
</tr>
<tr>
<td>Modified</td>
<td>$\chi^2$</td>
<td>11.39</td>
<td>8.92</td>
<td>11.21</td>
</tr>
<tr>
<td>p-value</td>
<td></td>
<td>(.181)</td>
<td>(.349)</td>
<td>(.190)</td>
</tr>
</tbody>
</table>

Table 5.2: $\chi^2$ test for equality of two survivor functions using AL estimates

<table>
<thead>
<tr>
<th>Test</th>
<th>Gender</th>
<th>DM</th>
<th>MI</th>
<th>HTN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmodified</td>
<td>$\chi^2$</td>
<td>12.96</td>
<td>10.93</td>
<td>11.47</td>
</tr>
<tr>
<td>p-value</td>
<td></td>
<td>(.113)</td>
<td>(.206)</td>
<td>(.176)</td>
</tr>
<tr>
<td>Modified</td>
<td>$\chi^2$</td>
<td>11.67</td>
<td>9.06</td>
<td>11.24</td>
</tr>
<tr>
<td>p-value</td>
<td></td>
<td>(.166)</td>
<td>(.337)</td>
<td>(.188)</td>
</tr>
</tbody>
</table>
Suppose we can distinguish between response times within groups, for example, son or father, male or female litter mates, husband or wife, and so on. In these cases, positive within group correlations can increase the power of tests for comparing survival probabilities.

Fujii (1989) tried to test for the equality of marginal distributions on positively dependent bivariate survival data for a special parametric model. He considered a bivariate survival model where given the value of a common variable, the conditional survival times $X$ and $Y$ independently follow Weibull distributions with hazard functions $\beta_1 t^r$ and $\beta_2 t^r$, respectively, where $\beta_1$, $\beta_2$, and $r$ are unknown parameters. Then he proposed a test for $\beta_1$ equal to $\beta_2$.

Clayton (1978) derived the constant odds-ratio model from the constant hazard ratio model for discrete response times given an example of failure times from father-son pair. He suggested that estimating the association parameter $\theta$ in the constant odds-ratio model is asymptotically identical to the Mantel-Haenzel (1959) pooled odds-ratio statistic assuming each pair of trials contributes independently to the information concerning $\theta$. A $\chi^2$ test statistic with 1 degree of freedom for the hypothesis $\theta = 1$, given by Mantel(1963) provides a test of the equality of the survival distributions for fathers and sons.

The development of a Wald test to compare life-table estimators of survival probabilities in this case, is a subject for future research. Let $P_1$ is a vector of survival probabilities from the first treatment and let $P_2$ be the corresponding vector of survival probabilities from the second treatment. In this case, the estimates $\hat{P}_1$
Figure 5.1: Comparison of PL survival curves for males and females

Figure 5.2: Comparison of PL survival curves for presence and absence of diabetes mellitus
Figure 5.3: Comparison of PL survival curves for presence and absence of myocardial infarction

Figure 5.4: Comparison of PL survival curves for presence and absence of hypertension
Figure 5.5: Comparison of AL survival curves for males and females

Figure 5.6: Comparison of AL survival curves for presence and absence of diabetes mellitus
Figure 5.7: Comparison of AL survival curves for presence and absence of myocardial infarction

Figure 5.8: Comparison of PL survival curves for presence and absence of hypertension
and $\hat{P}_2$ are not independent. Consequently, the covariance matrix of $\hat{P}_1 - \hat{P}_2$ used in 5.15 is generally not appropriate. Conditional binary responses similar to those introduced in Chapter 2 may be used to derive an estimate of the covariance matrix of $\hat{P}_1 - \hat{P}_2$.

A non-parametric likelihood ratio test for paired data is obtained by testing the fit of the model considered in Chapter 3 against a quasi-symmetry model that does not restrict the margins to be homogeneous. For large sample sizes, this would provide a chi-square test with $m$ degrees of freedom when there are $m + 1$ time intervals. Evaluation of this test requires development of additional S-plus functions for maximizing the loglikelihood for the quasi-symmetry model, alternatively, one could compare a model that requires homogeneous margins, but not symmetry, against Campbell's (1981) model.
CHAPTER 6. CONCLUSION

The naive PL and AL estimators perform very well as point estimates for marginal survival probabilities in the presence of correlated pairs, triplets, quadruples, and so on. However, the usual Greenwood variance formula, based on the assumption that all response times are independent, leads to underestimation of variances and inflated test statistics in the presence of positive correlations within groups. The modified Greenwood formula derived in Chapter 2 provides accurate estimates of variances.

The maximum likelihood estimator obtained from a non-parametric likelihood function based on a multinomial distribution for the observed counts is also a good estimator. The MLE is a self-consistent estimate and unique up to possible flat spots. The maximum likelihood estimation method provides estimates of joint bivariate survival probabilities. Estimates of marginal survival probabilities are obtained by summing the joint survival probabilities. For the general multivariate case, the number of parameters in the loglikelihood function increases as group sizes increase, which increases the difficulty of evaluating MLE’s, but the life-table estimates can be obtained easily and the variances and covariances are computed from the modified Greenwood formula without any iterative procedure.

The MLE and the PL estimator exhibit similar performance with respect to bias
and efficiency. The life-table method with modified Greenwood formula is very useful and easy to compute in real situations for the point estimators, their variances, and testing equality of two or more survival curves.


APPENDIX A. THE FOURTEEN FORMULAS FOR THE SECOND PARTIAL DERIVATIVES OF THE LOGLIKELYHOOD FUNCTION,
\[
\frac{\partial^2 \Lambda}{\partial \theta_{ux} \partial \theta_{yz}}
\]

Note that a corresponds to $\infty$ here, (i.e., $t_a = \infty$).

1. If $w = x = y = z$,
\[
\frac{\partial^2 \Lambda}{\partial \theta_{uu}^2} = -n\pi_{aa}e^{\theta_{uu}}(1 - \pi_{aa}e^{\theta_{uu}})
\]
\[
+ \sum_{j=1}^{u-1} (C1_{uj} + C2_{ju}) \sum_{k=j+1}^a e^{\theta_{uk}} \left(1 - \frac{e^{\theta_{uu}}}{\sum_{k=j+1}^a e^{\theta_{uk}}} \right)
\]
\[
+ \sum_{i=1}^{u-1} \sum_{j=1}^{u-1} CC_{ij} \sum_{l=i+1}^a \sum_{k=j+1}^a e^{\theta_{lk}} \left(1 - \frac{e^{\theta_{uu}}}{\sum_{l=i+1}^a \sum_{k=j+1}^a e^{\theta_{lk}}} \right).
\]

2. If $w < x$, $w = y$, and $x = z$,
\[
\frac{\partial^2 \Lambda}{\partial \theta_{uw}^2} = -2n\pi_{aa}e^{\theta_{uw}}(1 - \pi_{aa}2e^{\theta_{uw}})
\]
\[
+ \sum_{j=1}^{u-1} (C1_{uj} + C2_{ju}) \sum_{k=j+1}^a e^{\theta_{uk}} \left(1 - \frac{e^{\theta_{uw}}}{\sum_{k=j+1}^a e^{\theta_{uk}}} \right)
\]
\[
+ \sum_{i=1}^{u-1} \left[ (C2_{iv} + C1_{vi}) \frac{e^{\theta_{uv}}}{\sum_{l=i+1}^{a} e^{\theta_{lv}}} \left( 1 - \frac{e^{\theta_{lu}}}{\sum_{l=i+1}^{a} e^{\theta_{lv}}} \right) \right] \\
+ \sum_{i=1}^{u-1} \sum_{j=1}^{v-1} \frac{(CC_{ij} + CC_{ji})}{\sum_{l=i+1}^{a} \sum_{k=j+1}^{a} e^{\theta_{lk}}} \left( 1 - \frac{e^{\theta_{uv}}}{\sum_{l=i+1}^{a} \sum_{k=j+1}^{a} e^{\theta_{lk}}} \right) \\
- \sum_{i=1}^{u-1} \sum_{j=1}^{v-1} \frac{(CC_{ij} + CC_{ji})}{\sum_{l=i+1}^{a} \sum_{k=j+1}^{a} e^{\theta_{lk}}} \left( \frac{e^{2\theta_{uv}}}{\sum_{l=i+1}^{a} \sum_{k=j+1}^{a} e^{\theta_{lk}}} \right)^{2}.
\]

3. If \( w = x < y = z \),
\[
\frac{\partial^{2} \Lambda}{\partial \theta_{uu} \partial \theta_{uv}} = \eta(\pi_{aa}^{2} e^{\theta_{uu}} e^{\theta_{uv}})
\]
\[
- \sum_{i=1}^{u-1} \sum_{j=1}^{v-1} \frac{CC_{ij}}{\left( \sum_{l=i+1}^{a} \sum_{k=j+1}^{a} e^{\theta_{lk}} \right)^{2}}.
\]

4. If \( w = z, w = y, \) and \( y < z \),
\[
\frac{\partial^{2} \Lambda}{\partial \theta_{uu} \partial \theta_{uv}} = \eta(\pi_{aa}^{2} e^{\theta_{uu}} 2e^{\theta_{uv}})
\]
\[
- \sum_{i=1}^{u-1} \left( C_{1ui} + C_{2iu} \right) \left( \sum_{k=j+1}^{a} \theta_{uk} \right)^2 \\
+ \sum_{i=1}^{u-1} \sum_{j=1}^{u-1} \left[ C_{ij} \left( \sum_{l=i+1}^{a} \sum_{k=j+1}^{a} \theta_{lk} \right) \right]
\]

5. If \( w = x \), and \( w < y < z \),

\[
\frac{\partial^2 \Lambda}{\partial \theta_{uu} \partial \theta_{st}} = n(\pi^2 \theta_{uu} 2e_{st})
\]

6. If \( w < x \), \( w = y \), and \( x < z \),

\[
\frac{\partial^2 \Lambda}{\partial \theta_{uu} \partial \theta_{ut}} = 2n(\pi_{aa} \theta_{uu} 2e_{ut})
\]
\[ -\sum_{i=1}^{u-1} \sum_{j=1}^{u-1} \left( CC_{ij} + CC_{ji} \right) \left( \frac{e^{\theta_{uv}} e^{\theta_{st}}}{\sum_{l=i+1}^{a} \sum_{k=j+1}^{a} e^{\theta_{lk}}} \right)^2 \]

\[ -\sum_{i=1}^{u-1} \sum_{j=1}^{u-1} \left( CC_{ij} + CC_{ji} \right) \left( \frac{e^{\theta_{uv}} e^{\theta_{st}}}{\sum_{l=i+1}^{a} \sum_{k=j+1}^{a} e^{\theta_{lk}}} \right)^2 \cdot \]

7. If \( w < x, \ x < y, \ \text{and} \ y < z, \)

\[ \frac{\partial^2 \Lambda}{\partial \theta_{uv} \partial \theta_{st}} = 2n(x^2 e^{\theta_{uv}} 2e^{\theta_{st}}) \]

\[ -\sum_{i=1}^{u-1} \sum_{j=1}^{u-1} \left( CC_{ij} + CC_{ji} \right) \left( \frac{2e^{\theta_{uv}} e^{\theta_{st}}}{\sum_{l=i+1}^{a} \sum_{k=j+1}^{a} e^{\theta_{lk}}} \right)^2 \cdot \]

8. If \( w < x, \ \text{and} \ x = y = z, \)

\[ \frac{\partial^2 \Lambda}{\partial \theta_{uv} \partial \theta_{ss}} = 2n(x^2 e^{\theta_{uv}} 2e^{\theta_{ss}}) \]

\[ -\sum_{i=1}^{u-1} \left( C_{2iv} + C_{1vi} \right) \left( \frac{e^{\theta_{uv}} e^{\theta_{ss}}}{\sum_{l=i+1}^{a} e^{\theta_{lv}}} \right)^2 \cdot \]
9. If \( w < x, \ x < y, \) and \( y = z, \)
\[
\frac{\partial^2 \Lambda}{\partial \theta_{uv} \partial \theta_{ss}} = 2n(\pi^2 \sigma^2 e_{uv}^2 e_{ss}^2)
- \sum_{i=1}^{u-1} \sum_{j=1}^{v-1} (CC_{ij} + CC_{ji}) \left( \frac{\sum_{i}^{a} \sum_{k=j+1}^{a} e_{ik}}{\sum_{i}^{a} \sum_{k=j+1}^{a} e_{ik}} \right)^2.
\]

10. If \( w < x, \ x = y, \) and \( y < z, \)
\[
\frac{\partial^2 \Lambda}{\partial \theta_{uv} \partial \theta_{vt}} = 2n(\pi^2 \sigma^2 e_{uv}^2 e_{vt}^2)
- \sum_{i=1}^{u-1} (C_{2iv} + C_{1vi}) \left( \frac{\sum_{i}^{a} e_{iv}}{\sum_{i}^{a} e_{iv}} \right)^2
- \sum_{i=1}^{u-1} \sum_{j=1}^{v-1} (CC_{ij} + CC_{ji}) \left( \frac{\sum_{i}^{a} \sum_{k=j+1}^{a} e_{ik}}{\sum_{i}^{a} \sum_{k=j+1}^{a} e_{ik}} \right)^2.
\]
11. If \( w < x \), \( w < y < z \), and \( y = z \),

\[
\frac{\partial^2 \Lambda}{\partial \theta_{uv} \partial \theta_{ss}} = 2n(\pi_{aa}^2 e^{\theta_{uv} e^{\theta_{ss}}})
\]

\[
- \sum_{i=1}^{u-1} \sum_{j=1}^{s-1} \left( CC_{ij} + CC_{ji} \right) \frac{e^{\theta_{uv} e^{\theta_{ss}}}}{\left( \sum_{l=i+1}^{a} \sum_{k=j+1}^{a} e^{\theta_{lk}} \right)^2}
\]

12. If \( w < x \), \( w < y < z \), and \( z = x \),

\[
\frac{\partial^2 \Lambda}{\partial \theta_{uv} \partial \theta_{st}} = 2n(\pi_{aa}^2 e^{\theta_{uv} e^{\theta_{st}}})
\]

\[
- \sum_{i=1}^{u-1} \sum_{j=1}^{v-1} \left( CC_{ij} + CC_{ji} \right) \frac{e^{\theta_{uv} e^{\theta_{st}}}}{\left( \sum_{l=i+1}^{a} \sum_{k=j+1}^{a} e^{\theta_{lk}} \right)^2}
\]

\[
- \sum_{i=1}^{u-1} \sum_{j=1}^{s-1} \left( CC_{ij} + CC_{ji} \right) \frac{e^{\theta_{uv} e^{\theta_{st}}}}{\left( \sum_{l=i+1}^{a} \sum_{k=j+1}^{a} e^{\theta_{lk}} \right)^2}
\]

13. If \( w < x \), \( w < y < z \), and \( z < x \),

\[
\frac{\partial^2 \Lambda}{\partial \theta_{uv} \partial \theta_{st}} = 2n(\pi_{aa}^2 e^{\theta_{uv} e^{\theta_{st}}})
\]
\[
\left( \frac{1+f=\eta}{1+f=\eta} \right)^{\psi_{T_0}+\psi_{C}} \right)_{\left( 1+f=\eta \right)^{\psi_{T_0}+\psi_{C}}} = \frac{180^\theta_{m}\theta}{V_{z_0}}
\]

\[x > x \quad \text{and} \quad \forall x > h > m \quad \forall x > m \quad \forall x \]
APPENDIX B. S-PLUS PROGRAMS FOR MAXIMUM LIKELIHOOD ESTIMATOR

Given S-plus functions for evaluating an objective function, the first partial derivatives, and the matrix of second partial derivatives, the objective function can be maximized with any of several available iterative numerical algorithms, such as the Newton-Raphson algorithm. Consequently, we will only display the S-plus functions to evaluate the loglikelihood (3.6), the first partial derivatives of the loglikelihood (3.6), and the second partial derivatives of the loglikelihood (3.6) with respect to $\theta_{ij}$.

- Description of functions.
  - mle.nrobf ; compute $LL$, the value of the loglikelihood function at the current parameter values stored in $pmat$.
  - mle.nrdl ; compute $ders$, the vector of the first partial derivatives of the loglikelihood function.
  - mle.mps ; make a vector from an upper triangular matrix.
  - mle.suvst ; main program to compute $ddmat$, the matrix of second partial derivatives of the loglikelihood.
  - sub functions of mle.suvst to compute different cases of $\frac{\partial^2 LL}{\partial u \partial v \partial \theta_{st}}$.
    * mle.ddu ; if $u = v = s = t$
* mle.duuduv ; if \( u = v, v < s, s = t \)
* mle.duuduv ; if \( u = v, u = s, s < t \)
* mle.duudst ; if \( u = v, u < s < t \)
* mle.duuvdu ; if \( u < v, u = s, v = t \)
* mle.duvdut ; if \( u < v, u = s, v < t \)
* mle.duvstl ; if \( u < v, v < s, s < t \)
* mle.duvst2 ; if \( u < v, v < s, s = t \)
* mle.duvds1 ; if \( u < v, v = s = t \)
* mle.duvds2 ; if \( u < v, v < s, s = t \)
* mle.duvdts ; if \( u < v, u = s, u < v \), \( u > v \)
* mle.duvdst ; if \( u < v, u < s, v = t \)
* mle.duvdst1 ; if \( u < v, u < s, v < t \)
* mle.duvdst2 ; if \( u < v, u < s, v < t, t < v \)
* mle.duvdst3 ; if \( u < v, u < s < v, t > v \)

- Common arguments and variables.

- \( F, C1, C2, CC \); four data tables described in section 3.1.2.
- \( m \); number of time intervals, \( t_m = \infty \) in these programs.
- \( N \); number of pairs.
- \( pmat \); matrix of values for \( \theta_{ij} = \log (\pi_i / \pi_{mm}) \).
- \( pmm \); \( \pi_{mm} \).
- \( LL \); value of the loglikelihood function at \( pmat \).
- \( dmat \); upper triangular matrix of \( \frac{\partial LL}{\partial \theta_{ij}} \).
ders; vectorization of $dmat$, size is $\frac{m(m+1)}{2} - 1$

i.e., $ders = (\frac{\partial LL}{\partial \theta_{11}}, \ldots, \frac{\partial LL}{\partial \theta_{1m}}, \frac{\partial LL}{\partial \theta_{22}}, \ldots, \frac{\partial LL}{\partial \theta_{2m}}, \frac{\partial LL}{\partial \theta_{33}}, \ldots, \frac{\partial LL}{\partial \theta_{m(m-1)}})$

$ddmat$; matrix of $\frac{\partial^2 LL}{\partial \theta_{uv} \partial \theta_{st}}$, size is

$$\frac{m(m+1)}{2} - 1 \text{ by } \frac{m(m+1)}{2} - 1.$$

- S-plus functions.

```
mle.nrobf <- function(m, pmat, pmm, F, C1, C2, CC, N)
{
  i <- 0
  LL <- 0
  repeat {
    i <- i + 1
    if(i == m)
      ii <- m
    else ii <- i + 1
    j <- 0
    repeat {
      j <- j + 1
      if(j == m)
        jj <- m
      else jj <- j + 1
      LL <- LL + F[i, j] * log(pmat[i, j]) + C2[i, j] * log((
        sum(pmat[(ii:m), (jj:m)])) + C1[i, j] * log((sum(
        pmat[i, (jj:m)]))) + CC[i, j] * log((sum(pmat[
        ii:m), (jj:m)])))
    };
    if(i == m)
      break
  }
  if(i == m)
    break
  LL <- LL + N * log(pmm)
  LL
}
```
"mle.nrd1"<-
function(m, pmat, pmm, F, C1, C2, CC, N)
{
    dmat <- matrix(0, m, m)
    dmat[1, 1] <- F[1, 1] - N * pmm * pmat[1, 1]
    v <- 1
    repeat {
        v <- v + 1
        j <- 0
        repeat {
            j <- j + 1
            if(j == (v - i))
                break
        }
        if(v == (m - 1))
            break
    }
    for(v in 2:(m - 1))
    u <- 0
    repeat {
        u <- u + 1
        v <- u
        repeat {
            v <- v + 1
        }
    }
}
j <- 0
repeat {
  j <- j + 1
  dmat[u, v] <- dmat[u, v] + (C1[u, j] + C2[j, u])
  /sum(pmat[u, (j + 1):m])
  if(j == (v - 1))
    break
}
i <- 0
if(u != 1) {
  repeat {
    i <- i + 1
    dmat[u, v] <- dmat[u, v] + (C2[i, v] + C1[v, i]) / sum(pmat[(i + 1):m, v])
    if(i == (u - 1))
      break
  }
  if(i == (u - 1))
    break
}
if(v == m)
  break
if(u == (m - 1))
  break

ders <- mle.mps(dmat)
ders
"mle.mps" <-
function(dmat) {
  m <- length(dmat[1, ])
  ders <- 1:m
  jj <- 0
  i <- 0
  repeat {
    i <- i + 1
    j <- i - 1
    repeat {
      j <- j + 1
      jj <- jj + 1
      ders[jj] <- dmat[i, j]
      if(j == m)
        break
    }
    if(i == (m - 1))
      break
  }
  ders
}
"mle.suvst" <-
function(m, pmat, pmm, F, C1, C2, CC, N)
{
  u <- 0
  i <- 0
  k <- (m * (m + 1))/2 - 1
  ddmat <- matrix(0, k, k)
  ii <- rep(0, k^2)
  repeat {
    u <- u + 1
    v <- u - 1
    repeat {
      v <- v + 1
      s <- 0
      repeat {
        s <- s + 1
        t <- s - 1
        repeat {
          t <- t + 1
          i <- i + 1
          if(u == v) {
            if(v == s) {
              ii[i] <- mle.ddu(m, pmat, pmm, F, C1, C2, CC, N, t)
            }
            if(s == t) {
              ii[i] <- mle.duudv(m, pmat, pmm, F, C1, C2, CC, N, u, t)
            }
            if(u < s) {
              if(s < t) {
                ii[i] <- mle.duudst(m, pmat, pmm, F, C1, C2, CC, N, V, s, t)
              }
              if(v == t) {
                ii[i] <- mle.duvduv(m, pmat, pmm, F, C1, C2, CC, N, v, s, t)
              }
            }
            if(s == t) {
              if(v == t) {
                ii[i] <- mle.duduv(m, pmat, pmm, F, C1, C2, CC, N, s, t)
              }
            }
          }
        }
      }
    }
  }
}

if(u == s)
  if(v < t) {
    ii[i] <- mle.duvdat(m, pmat, pmm, F,
    C1, C2, CC, N, u, v, t)
  }
  if(s < t) {
    ii[i] <- mle.duvdut(m, pmat, pmm, F,
    C1, C2, CC, N, u, v, s, t)
  }
if(v < s)
  if(s == t) {
    ii[i] <- mle.duvdssi(m, pmat, pmm, F,
    C1, C2, CC, N, u, v, s, t)
  }
if(v == s)
  if(s == t) {
    ii[i] <- mle.duvdss1(m, pmat, pmm, F,
    C1, C2, CC, N, u, v, s, t)
  }
if(v < s)
  if(s == t) {
    ii[i] <- mle.duvdvc(m, pmat, pmm, F,
    C1, C2, CC, N, u, v, s, t)
  }
  if(v == t) {
    ii[i] <- mle.duvdsv(m, pmat, pmm, F,
    C1, C2, CC, N, u, v, s, t)
  }
if(s > u)
  if(v > s) {
    if(s == t) {
      ii[i] <- mle.duvdss2(m, pmat, pmm, F,
      C1, C2, CC, N, u, v, t)
    }
  }
if(v == s)
  if(s < t) {
    ii[i] <- mle.duvdvs(t, pmat, pmm, F,
    C1, C2, CC, N, u, v, t)
  }
if(v < s)
  if(s == t) {
    ii[i] <- mle.duvdsvs1(m, pmat, pmm, F,
    C1, C2, CC, N, u, v, s, t)
  }
if(s != t)
  if(t < v) {
    ii[i] <- mle.duvdsvd2(m, pmat, pmm, F,
    C1, C2, CC, N, u, v, s, t)
  }
  if(t > v) {
    ii[i] <- mle.duvdsvd3(m, pmat, pmm, F,
    C1, C2, CC, N, u, v, s, t)
  }
}
if(t == m)
break
if(s == (n - 1))
    break
}
if(v == m)
    break
}
if(u == (m - 1))
    break
}
dmat <- matrix(ii, k, k, byrow = T)
dmat
function(m, pmat, pmm, F, C1, C2, CC, N, u)
{
  duu <- 0
  j <- 0
  if(u != 1) {
    repeat {
      j <- j + 1
      sk <- pmat[u, u]/sum(pmat[u, (j + 1):m])
      duu <- duu + (C1[u, j] + C2[j, u]) * sk * (1 - sk)
      if(j == (u - 1))
        break
    }
    i <- 0
    repeat {
      i <- i + 1
      j <- 0
      repeat {
        j <- j + 1
        slk <- pmat[u, u]/sum(pmat[(i + 1):m, (j + 1):m])
        duu <- duu + CC[i, j] * slk * (1 - slk)
        if(j == (u - 1))
          break
      }
      if(i == (u - 1))
        break
    }
  }
  duu <- duu - N * pmm * pmat[u, u] * (1 - pmm * pmat[u, u])
}

"mle.ddu"<-

"mle.duudvv" <-
function(m, pmat, pmm, F, C1, C2, CC, N, u, v)
{
  duudvv <- 0
  if(u != 1) {
    i <- 0
    repeat {
      i <- i + 1
      j <- 0
      repeat {
        j <- j + 1
        slk <- (pmat[u, u] * pmat[v, v]) / (sum(pmat[(i + 1):m, (j + 1):m]))^2
        duudvv <- duudvv + CC[i, j] * slk
        if(j == (u - 1))
          break
      }
      if(i == (u - 1))
        break
    }
  }
  duudvv <- N * (pmm^2) * (pmat[u, u]) * pmat[v, v] - duudvv
  duudvv
}
```
mle.duuduv <-
function(m, pmat, pmm, F, C1, C2, CC, H, u, v)
{
    duuduv <- 0
    j <- 0
    if(u != 1) {
        repeat {
            j <- j + 1
            sk <- (pmat[u, u] * pmat[u, v])/(sum(pmat[u, (j + 1):m])^2)
            duuduv <- duuduv + (C1[u, j] + C2[j, u]) * sk
            if(j == (u - 1))
                break
        }
        i <- 0
        repeat {
            i <- i + 1
            j <- 0
            repeat {
                j <- j + 1
                slk <- (2 * (pmat[u, u] * pmat[u, v]))/((sum(pmat[(i + 1):m, (j + 1):m])^2)
                duuduv <- duuduv + CC[i, j] * slk
                if(j == (u - 1))
                    break
            }
            if(i == (u - 1))
                break
        }
    }
    duuduv <- H * (pmm^2) * pmat[u, u] * 2 * pmat[u, v] - duuduv
    duuduv
}
```

"mle.duudst"<-

function(m, pmat, pmm, F, C1, C2, CC, N, u, s, t)
{
  duudst <- 0
  if(u != 1) {
    i <- 0
    repeat {
      j <- 0
      repeat {
        i <- i + 1
        j <- j + 1
        slk <- (2 * (pmat[u, u] * pmat[s, t]))/((sum(pmat[(i + i):m, (j + i):m]))^2)
        duudst <- duudst + CC[i, j] * slk
        if(j == (u - 1))
          break
      }
      if(i == (u - 1))
        break
    }
  }
  duudst <- N * (pmm^-2) * (pmat[u, u]) * 2 * pmat[s, t] - duudst
duudst
}
\begin{verbatim}
"mle.duvduv"<-
function(m, pmat, pmm, F, C1, C2, CC, N, u, v)
{
duvduv <- 0
aa <- 0
j <- 0
if(u != 1) {
    repeat {
        j <- j + 1
        sk <- pmat[u, v]/sum(pmat[u, (j + 1):m])
        duvduv <- duvduv + (C1[u, j] + C2[j, u]) * sk * (1 - sk )
        if(j == (v - 1))
            break
    }
    i <- 0
    repeat {
        i <- i + 1
        j <- 0
        repeat {
            j <- j + 1
            slk <- pmat[u, v]/sum(pmat[(i + 1):m, (j + 1):m])
            duvduv <- duvduv + (CC[i, j] + CC[j, i]) * slk * (1 - slk)
            if(j == (v - 1))
                break
        }
        if(i == (u - 1))
            break
    }
}
i <- 0
repeat {
    i <- i + 1
    j <- 0
    repeat {
        j <- j + 1
        sss <- ((pmat[u, v])^2)/((sum(pmat[(i + 1):m, (j + 1):m]))^2)
        aa <- aa + (CC[i, j] + CC[j, i]) * sss
        if(j == (u - 1))
            break
    }
}

\end{verbatim}
break

if(i == (u - 1))
  break

}  

duvduv <- duvduv - aa - 2 * H * pmm * pmat[u, v] * (1 - 2 * pmm * pmat[u, v])

duvduv
```
"mle.duvdut"<-
function(m, pmat, pmm, F, C1, C2, CC, N, u, v, t)
{
  duvdut <- 0
  addl <- 0
  j <- 0
  repeat { 
    j <- j + 1
    sl <- (pmat[u, v] * pmat[u, t])/(sum(pmat[u, (j + 1):m]))^2
    duvdut <- duvdut + (C1[u, j] + C2[j, u]) * sl
    if(j == (v - 1))
      break
  }
  if(u != 1) {
    i <- 0
    repeat {
      j <- i + 1
      s1 <- (pmat[u, v] * pmat[u, t])/(sum(pmat[[i + 1]:m], (j + 1):m))^2
      duvdut <- duvdut + (CC[i, j] + CC[j, i]) * s1
      if(j == (v - 1))
        break
    }
    if(i == (u - 1))
      break
  }
  i <- 0
  repeat {
    j <- i + 1
    s1l <- (pmat[u, v] * pmat[u, t])/(sum(pmat[[i + 1]:m], (j + 1):m))^2
    addl <- addl + (CC[i, j] + CC[j, i]) * s1l
    if(j == u - 1)
      break
  }
  if(i == u - 1)
    break
}

duvdut <- N * (pmm^2) * 4 * pmat[u, v] * pmat[u, t] - duvdut - addl
duvdut
```
```
mle.duvdst1<- 
function(m, pmat, pmm, F, C1, C2, CC, N, u, v, s, t) 
{
    duvdst1 <- 0
    if(u != 1) {
        i <- 0
        repeat {
            i <- i + 1
            j <- 0
            repeat {
                j <- j + 1
                slk <- (2 * (pmat[u, v] * pmat[s, t]))/((sum(pmat[(i + 1):m, (j + 1):m]))^2)
                duvdst1 <- duvdst1 + (CC[i, j] + CC[j, i]) * slk
                if(j == (v - 1))
                    break
            }
            if(i == (u - 1))
                break
        }
    }
    duvdst1 <- 2 * N * (pmm^2) * (pmat[u, v]) * 2 * pmat[s, t] - duvdst1
    duvdst1
}
```
"mle.duvdss1"<-
function(m, pmat, pmm, F, C1, C2, CC, N, u, v, s)
{
  duvdss1 <- 0
  i <- 0
  if(u != 1) {
    repeat {
      i <- i + 1
      sl <- (pmat[u, v] * pmat[s, s])/(sum(pmat[(i + 1):m, v]))^2
      duvdss1 <- duvdss1 + (C2[i, v] + C1[v, i]) * sl
      if(i == (u - 1))
        break
    }
  }
  i <- 0
  repeat {
    i <- i + 1
    j <- 0
    repeat {
      j <- j + 1
      slk <- (pmat[u, v] * pmat[s, s])/(sum(pmat[(i + 1):m, (j + 1):m]))^2
      duvdss1 <- duvdss1 + (CC[i, j] + CC[j, i]) * slk
      if(j == (v - 1))
        break
    }
    if(i == (u - 1))
      break
  }
  duvdss1 <- 2 * N * (pmm^2) * pmat[u, v] * pmat[s, s] - duvdss1
  duvdss1
}
"mle.duvdss2"<-
function(m, pmat, pmm, F, C1, C2, CC, N, u, v, s)
{
  duvdss2 <- 0
  i <- 0
  if(u != 1) {
    i <- 0
    repeat {
      i <- i + 1
      j <- 0
      repeat {
        j <- j + 1
        slk <- (pmat[u, v] * pmat[s, s])/((sum(pmat[(i +
          1):m, (j + 1):m]))^2)
        duvdss2 <- duvdss2 + (CC[i, j] + CCC[j, i]) * slk
        il(j == (v - 1))
        break
      }
      il(i == (u - 1))
      break
    }
  }
  duvdss2 <- 2 * N * (pmm^2) * pmat[u, v] * pmat[s, s] - duvdss2
  duvdss2
}
"mle.duvdvt" <-
function(m, pmat, pmm, F, C1, C2, CC, N, u, v, t)
{
  duvdvt <- 0
  i <- 0
  if(u != 1) {
    repeat {
      i <- i + 1
      sl <- (pmat[u, v] * pmat[v, t])/((sum(pmat[(i + 1):m, v ]))^-2)
      duvdvt <- duvdvt + (C2[i, v] + C1[v, i]) * sl
      if(i == (u - 1))
        break
    }
    i <- 0
    repeat {
      i <- i + 1
      j <- 0
      repeat {
        j <- j + 1
        slk <- (2 * (pmat[u, v] * pmat[v, t]))/((sum( pmat[(i + 1):m, (j + 1):m]))^-2)
        duvdvt <- duvdvt + (CC[i, j] + CC[j, i]) * slk
        if(j == (v - 1))
          break
      }
      if(i == (u - 1))
        break
    }
  }
  duvdvt <- 4 * N * (pmm^-2) * pmat[u, v] * pmat[v, t] - duvdvt
  duvdvt
}
```
mle.duvdss3 <- 
function(m, pmat, pmm, F, C1, C2, CC, H, u, v, s) 
{
  duvdss3 <- 0
  i <- 0
  if(u != 1) {
    i <- 0
    repeat {
      i <- i + 1
      j <- 0
      repeat {
        j <- j + 1
        slk <- (pmat[u, v] * pmat[s, s]) / ((sum(pmat[(i + 1):m, (j + 1):m]))^2)
        duvdss3 <- duvdss3 + (CC[i, j] + CC[j, i]) * slk
        if(j == (s - 1))
          break
      }
      if(i == (u - 1))
        break
    }
  }
  duvdss3 <- 2 * H * (pmm^2) * pmat[u, v] * pmat[s, s] - duvdss3
  duvdss3
} 
```
"mle.duvdsV"<-

function(m, pmat, pmm, F, C1, C2, CC, N, u, v, s)
{
    duvdsv <- 0
    i <- 0
    if(u != 1) {
        repeat {
            j <- i + 1
            slk <- ((pmat[u, v] * pmat[s, v])/(sum(pmat[(i + 1):m, (j + 1):m])^2))
            duvdsv <- duvdsv + (C2[i, v] + C1[v, i]) * slk
            if(i == (u - 1))
                break
        }
        i <- 0
        repeat {
            j <- i + 1
            slk <- ((pmat[u, v] * pmat[s, v])/(sum(pmat[(i + 1):m, (j + 1):m])^2))
            duvdsv <- duvdsv + (CC[i, j] + CC[j, i]) * slk
            if(j == (v - 1))
                break
            if(i == (u - 1))
                break
        }
    }
    duvdsv <- 4 * N * (pmm^2) * pmat[u, v] * pmat[s, v] - duvdsv
    duvdsv
}
```r
"mle.duvdst2" <-
function(m, pmat, pmm, F, C1, C2, CC, N, u, v, s, t)
  {
    duvdst2 <- 0
    i <- 0
    if(u != 1) {
      i <- 0
      repeat {
        i <- i + 1
        j <- 0
        repeat {
          j <- j + 1
          slk <- (pmat[u, v] * pmat[s, t]) / ((sum(pmat[(i + 1):m, (j + 1):m]))^2)
          duvdst2 <- duvdst2 + (CC[i, j] + CC[j, i]) * slk
          if(j == (t - 1))
            break
        }
        if(i == (u - 1))
          break
      }
    }
    i <- 0
    repeat {
      i <- i + 1
      j <- 0
      repeat {
        j <- j + 1
        slk <- (pmat[u, v] * pmat[s, t]) / ((sum(pmat[(i + 1):m, (j + 1):m]))^2)
        duvdst2 <- duvdst2 + (CC[i, j] + CC[j, i]) * slk
        if(j == (s - 1))
          break
      }
      if(i == (u - 1))
        break
    }
    duvdst2 <- 2 * N * (pmm^2) * 2 * pmat[s, t] * pmat[u, v] - duvdst2
    duvdst2
  }
```
"mle.duvdst3"

```r
function(m, pmat, pmm, F, Cl, C2, CC, N, u, v, s, t)
{
  duvdst3 <- 0
  i <- 0
  if(u != 1) {
    i <- 0
    repeat {
      i <- i + 1
      j <- 0
      repeat {
        j <- j + 1
        slk <- ((pmat[u, v] * pmat[s, t]) / (sum(pmat[(
          i + l):m, (j + l):m]))^2)
        duvdst3 <- duvdst3 + (CC[i, j] + CC[j, i]) * slk
        if(j == (v - 1))
          break
      }
      if(i == (u - 1))
        break
    }
  }
  i <- 0
  repeat {
    i <- i + 1
    j <- 0
    repeat {
      j <- j + 1
      slk <- ((pmat[u, v] * pmat[s, t]) / (sum(pmat[(
        i + l):m, (j + l):m]))^2)
      duvdst3 <- duvdst3 + (CC[i, j] + CC[j, i]) * slk
      if(j == (s - 1))
        break
    }
    if(i == (u - 1))
      break
  }
  duvdst3 <- 2 * N * (pmm^-2) * 2 * pmat[s, t] * pmat[u, v] - duvdst3
  duvdst3
}
```
APPENDIX C. S-PLUS PROGRAM FOR THE LIFE TABLE
ANALYSIS OF CORRELATED RESPONSES

S-plus functions are displayed to compute the PL, and AL estimates and the variances provided by the unmodified Greenwood formula and the modified Greenwood formula.

- Common arguments and variables.

  - \( \hat{dai} \); list of data tables.
    * \( dat[[1]], dat[[2]], dat[[3]], dat[[4]] \) correspond to \( F, C1, C2, CC \) for just pairs and triplets .... without singletons.
    * \( dat[[5]] \) has two columns for number of failure and number of censored observations within time intervals ignoring groups.
  - note that \( F, C1, C2, CC \) contains 3 pairs for each triplet, 6 pairs for each quadruple, etc ..., when we make \( dat[[1]], dat[[2]], dat[[3]], dat[[4]] \), but not \( dat[[5]] \).
  - \( pP \); vector of \( \hat{p}_h \) for PL.
  - \( PL \); vector of \( \hat{P}_h \) for PL.
  - \( Nh \); vector of number at risk for PL.
  - \( VPL \); vector of \( \text{Var}(PL) \) from usual Greenwood formula.
- $pA$ ; vector of $\hat{p}_h$ for AL.
- $AL$ ; vector of $\hat{P}_h$ for AL.
- $aNh$ ; vector of adjusted number at risk for AL.
- $VAL$ ; vector of $\text{Var}(AL)$ from usual Greenwood formula.
- $covP$ ; covariance matrix of PL from usual Greenwood formula.
- $covA$ ; covariance matrix of AL from usual Greenwood formula.
- $mCovAL$ ; covariance matrix of AL from the modified Greenwood formula.
- $mcova$ ; covariance matrix of $pA$ for the modified Greenwood formula.
- $mCovPL$ ; covariance matrix of PL from the modified Greenwood formula.
- $mcoup$ ; covariance matrix of $pP$ for the modified Greenwood formula.
- $mvpA$ ; vector of $\text{Var}(\hat{p}_h)$ for AL from the modified Greenwood formula.
- $mvpP$ ; vector of $\text{Var}(\hat{p}_h)$ for PL from the modified Greenwood formula.

- Description of functions.

- $main.life$ ; display all life-table estimates including modified ones.

- sub functions of $main.life$.

  * $alpl.life$ ; list the following vectors or matrices from the fundamental life-table method,
    
    $\text{list}(pP, PL, Nh, VPL, pA, AL, aNh, VAL, covP, covA)$.

  * $mgwA.life$ ; compute $mCovAL$.

  * sub function of $mgwA.life$.

    - $mcov.AL$ ; compute $mcova$ as $V$ on $\text{Cov}(\hat{P}_h) \approx dVd$. 

- $dPp$; compute $d$ matrix on $\text{Cov}(\hat{P}_h) \approx dVd$

* $mgwP.life$; compute $mcovPL$.

* sub function of $mgwP.life$.

- $mcov.PL$; compute $mcovp$ as $V$ on $\text{Cov}(\hat{P}_h) \approx dVd$.

- $dPp$; compute $d$ matrix on $\text{Cov}(\hat{P}_h) \approx dVd$.

- S-plus functions.

"main.life"<-
function(dat)
{
  out <- alpl.life(dat)
  out1 <- out[[1]]
  sdPL <- sqrt(out1[, 4])
  sdAL <- sqrt(out1[, 6])
  out <- cbind(out1[, 2], sdPL, out1[, 6], sdAL)
  l11 <- length(out[, 1])
  out <- out[-l11, ]
  cat(" Output from the unmodified ", fill = T)
  cat(" ", fill = T)
  print(out)
  F <- dat[[1]]
  C1 <- dat[[2]]
  C2 <- dat[[3]]
  CC <- dat[[4]]
  km1 <- out1[, 1:3]
  N <- NULL
  out2 <- mgwP.life(N, F, C1, C2, CC, km1)
  km2 <- out1[, 5:7]
  out3 <- mgwA.life(N, F, C1, C2, CC, km2)
  cat(" ", fill = T)
  cat(" Cov(PL) from the unmodified ", fill = T)
  print(out2[[2]])
  cat(" ", fill = T)
  cat(" ", fill = T)
  cat(" Cov(AL) from the unmodified ", fill = T)
  print(out3[[3]])
  cat(" ", fill = T)
cat(" ", fill = T)
cat(" Cov(PL) from the modified ", fill = T)
print(out2)
cat(" ", fill = T)
cat(" s.d(PL) from the modified; ", fill = T)
cat(" i.e. square root of diag(Cov(PL))", fill = T)
print(sqrt(diag(out2)))
cat(" ", fill = T)
cat(" COv(AL) from the modified ", fill = T)
print(out3)
cat(" ", fill = T)
cat(" s.d(AL) from the modified; ", fill = T)
cat(" i.e. square root of diag(Cov(AL))", fill = T)
print(sqrt(diag(out3)))
cat(" ", fill = T)
cat(" , where s.d indicates that standard deviation.", fill = T)
cat(" ", fill = T)
outt <- list(out[, 1], out[, 3], out[2], out[3], out2, out3)
outt
"alpl.life" <-
function(dat)
{
  m <- length(dat[[5]][, 1])
  Nh <- 0 * i:m
  FF <- dat[[5]][, 1]
  W <- dat[[5]][, 2]
  Nh[i] <- sum(dat[[5]])
  for(i in 2:m)
    Nh[i] <- Nh[i - 1] - FF[i - 1] - W[i - 1]
  aNh <- Nh - 0.5 * W
  qP <- FF/Nh
  qA <- FF/aNh
  qP[is.na(qP)] <- 0
  qA[is.na(qA)] <- 0
  pP <- 1 - qP
  pA <- 1 - qA
  PL <- 0 * i:m
  AL <- PL
  for(i in 2:(m - 1)) {
    PL[i] <- PL[i - 1] * pP[i]
    AL[i] <- AL[i - 1] * pA[i]
  }
  abcP <- dpP(pP, PL)
  covP <- (pP * (1 - pP))/Nh
  covP <- covP[1:m - 1]
  covP[is.na(covP)] <- 0
  covP <- abcP %*% diag(covP) %*% t(abcP)
  VPL <- c(diag(covP), 0)
  abcA <- dpP(pA, AL)
  covA <- (pA * (1 - pA))/aNh
  covA <- covA[1:m - 1]
  covA[is.na(covA)] <- 0
  covA <- abcA %*% diag(covA) %*% t(abcA)
  VAL <- c(diag(covA), 0)
  RR <- cbind(pP, PL, Nh, VPL, pA, AL, aNh, VAL)
  RR <- list(RR, covP, covA)
  RR
}
"mgwA.life" <-
function(F, C1, C2, CC, km)
{
  m <- length(km[, 1])
  lh <- m - 1
  pA <- km[, 1]
  AL <- km[, 2]
  aNh <- km[, 3]
  mvpA <- 1:lh
  mCovAL <- 1:lh
  gg <- 0
  F <- (F + t(F))/2
  CI <- (C1 + t(C2))/2
  C2 <- t(C1)
  CC <- (CC + t(CC))/2
  for(h in 1:lh) {
    ggg <- 1 - pA[h]
    mvpA[h] <- ggg * (1 - ggg)
    mvpA[h] <- mvpA[h]/aNh[h]
    aa <- sum(F[h:m, h:m]) + sum(C1[h:m, h:m]) + sum(C2[h:m, h:m]) + sum(CC[h:m, h:m])
    bb <- F[h, h]
    eee <- sum(F[(h:m), h]) + sum(C2[(h:m), h])
    aa.e2 <- aa - (sum(C2[h, (h:m)])/2 - sum(CC[h, (h:m)])/2
    eee <- eee/aa.e2
    xxx <- bb/aa.e2 - (eee)~2
    xxx <- (2 * aa.e2 * xxx)/(aNh[h]~2)
    xxx[is.na(xxx)] <- 0
  }
  mcova <- mcov.AL(F, C1, C2, CC, lh, mvpA, aNh)
  abc <- dPp(pA, AL)
  mcova[is.na(mcova)] <- 0
  mCovAL <- abc %% mcova %% t(abc)
  mCovAL
}
```
mcov.AL <-
function(F, C1, C2, CC, lh, mvpA, aNh)
{
  m <- lh + 1
  mcovA <- matrix(0, lh, lh)
  i <- 0
  repeat {
    i <- i + 1
    j <- i
    repeat {
      j <- j + 1
      aaa1 <- sum(F[i:m, j:m]) + sum(C1[i:m, j:m]) + sum(C2[i:m, j:m]) + sum(CC[i:m, j:m])
      ee1 <- sum(F[i, j:m] + C1[i, j:m])
      ee2 <- sum(F[i:m, j] + C2[i:m, j])
      adj.a <- sum(C1[(j:m), i]) + sum(C2[(j:m), (i:m)]) + sum(CC[(j:m), i]) + sum(CC[(j:m), (i:m)])
      adj.a <- adj.a / 4
      aj.e1 <- aaa1 - adj.a
      aa <- F[i, j]/aj.e1
      e1 <- ee1/aj.e1
      e2 <- ee2/aj.e1
      aa <- aa - (e1 * e2)
      aa[is.na(aa)] <- 0
      kij <- 2 * aj.e1
      mcovA[i, j] <- (kij * aa)/(aNh[i] * aNh[j])
      if(j == lh)
        break
    }
    if(i == lh - 1)
      break
  }
  mcovA <- diag(mvpA) + mcovA
  mcovA
}
```
"dPp" <-
function(pk, PLk)
{
  h <- length(pk) - 1
  dmat <- matrix(0, h, h)
  i <- 0
  repeat {
    i <- i + 1
    j <- 0
    repeat {
      j <- j + 1
      dmat[i, j] <- PLk[i]/pk[j]
      if(i == j)
        break
    }
    if(i == h)
      break
  }
  dmat
}
"mgwp.life" <-
function(F, C1, C2, CC, km)
{
  m <- length(km[, 1])
  lh <- m - 1
  pP <- km[, 1]
  PL <- km[, 2]
  Wh <- km[, 3]
  mvpP <- 1:lh
  mcovPL <- 1:lh
  gg <- 0
  F <- (F + t(F))/2
  C1 <- (C1 + t(C2))/2
  C2 <- t(C1)
  CC <- (CC + t(CC))/2
  for(h in 1:lh) {
    ggg <- 1 - pP[h]
    mvpP[h] <- ggg * (1 - ggg)
    mvpP[h] <- mvpP[h]/Wh[h]
    aa <- sum(F[h:m, h:m]) + sum(C1[h:m, h:m]) + sum(C2[h:m, h:m]) +
         sum(CC[h:m, h:m])
    bb <- F[h, h]
    eee <- sum(F[(h:m), h] + C2[(h:m), h])
    eee <- eee/(aa)
    xxx <- bb/aa - (eee)^2
    xxx <- (2 * aa * xxx)/(Wh[h]^2)
    xxx[is.na(xxx)] <- 0
    mvpP[h] <- mvpP[h] + xxx
  }
  mcovp <- mcov.PL(F, C1, C2, CC, lh, mvpP, Wh)
  mcovp[is.na(mcovp)] <- 0
  abc <- dPp(pP, PL)
  mcovPL <- abc %*% mcovp %*% t(abc)
  mcovPL
}

"mcov.PL"<-
function(F, C1, C2, CC, lh, mvpP, Nh)
{
  m <- lh + 1
  mcovP <- matrix(0, lh, lh)
  i <- 0
  repeat {
    i <- i + 1
    j <- i
    repeat {
      j <- j + 1
      aaa1 <- sum(F[i:m, j:m]) + sum(C1[i:m, j:m]) + sum(C2[i:m, j:m]) + sum(CC[i:m, j:m])
      eel <- sum(F[i:m, j] + C1[i, j:m])
      ee2 <- sum(F[i:m, j] + C2[i:m, j])
      el <- eel/aaa1
      e2 <- ee2/aaa1
      aa <- F[i, j]/aaa1
      aa <- aa - (el * e2)
      aa[is.na(aa)] <- 0
      kij <- 2 * aaa1
      mcovP[i, j] <- (kij * aa)/(Nh[i] * Nh[j])
      mcovP[j, i] <- mcovP[i, j]
      if(j == lh)
        break
      if(i == lh - 1)
        break
    }
    mcovP <- diag(mvpP) + mcovP
  }
  mcovP
}
APPENDIX D. ANGIOPLASTY DATA

ID ; Patient identification number

SEX (1; Male, 2; Female)

DM : Diabetes Mellitus (1; absence, 2; presence)

MI: Myocardial Infarction (1; absence, 2; presence)

HTN: Hypertension (1; absence, 2; presence)

Pair (1; singleton, 2; pair, .......)

Fail (; if censored,
     \[ t_h; \text{the end of time interval if failed} \])

Cen (; if failed,
     \[ t_h; \text{the end of time interval if censored} \])
Table D.1: Angioplasty Data

<table>
<thead>
<tr>
<th>ID</th>
<th>SEX</th>
<th>DM</th>
<th>MI</th>
<th>HTN</th>
<th>Pair</th>
<th>Fail</th>
<th>Cen</th>
</tr>
</thead>
<tbody>
<tr>
<td>C036</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>21</td>
</tr>
<tr>
<td>D074</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>9</td>
</tr>
<tr>
<td>B018</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>D044</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>9</td>
</tr>
<tr>
<td>F053</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>F059</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>9</td>
</tr>
<tr>
<td>E087</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>.</td>
</tr>
<tr>
<td>D067</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>18</td>
</tr>
<tr>
<td>D050</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>6</td>
</tr>
<tr>
<td>E041</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>.</td>
</tr>
<tr>
<td>B015</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>9</td>
</tr>
<tr>
<td>E002</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>6</td>
</tr>
<tr>
<td>E008</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>15</td>
</tr>
<tr>
<td>D014</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>D042</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>18</td>
</tr>
<tr>
<td>D075</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>6</td>
</tr>
<tr>
<td>E046</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>.</td>
</tr>
<tr>
<td>E021</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>24</td>
<td>.</td>
</tr>
<tr>
<td>D053</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>.</td>
</tr>
<tr>
<td>F034</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>6</td>
</tr>
<tr>
<td>D029</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>30</td>
<td>.</td>
</tr>
<tr>
<td>D035</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>42</td>
</tr>
<tr>
<td>D037</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>36</td>
</tr>
<tr>
<td>D078</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>15</td>
</tr>
<tr>
<td>F085</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>6</td>
</tr>
<tr>
<td>B011</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>D034</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>15</td>
<td>.</td>
</tr>
<tr>
<td>D064</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>12</td>
</tr>
<tr>
<td>D060</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>.</td>
</tr>
<tr>
<td>F080</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>6</td>
</tr>
<tr>
<td>D061</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>.</td>
</tr>
<tr>
<td>D084</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>3</td>
</tr>
</tbody>
</table>
Table D.1 (Continued)

<table>
<thead>
<tr>
<th>ID</th>
<th>SEX</th>
<th>DM</th>
<th>MI</th>
<th>HTN</th>
<th>Pair</th>
<th>Fail</th>
<th>Cen</th>
</tr>
</thead>
<tbody>
<tr>
<td>D028</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>D070</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>B046</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>F010</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>F019</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>B008</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>D049</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>B028</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td>54</td>
</tr>
<tr>
<td>C014</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td>42</td>
</tr>
<tr>
<td>D073</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>F077</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>D072</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>E032</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>D082</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>B044</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>B039</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>F087</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>E062</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>D011</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>C003</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>D002</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>E025</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>C022</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>B022</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>E015</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>F014</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>E028</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>C035</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>E051</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>D059</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>E034</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>D051</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>ID</td>
<td>SEX</td>
<td>DM</td>
<td>MI</td>
<td>HTN</td>
<td>Pair</td>
<td>Fail</td>
<td>Cen</td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>----</td>
<td>----</td>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td>B003</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>B024</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>F044</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>9</td>
</tr>
<tr>
<td>C015</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>F035</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>9</td>
</tr>
<tr>
<td>F047</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>.</td>
</tr>
<tr>
<td>D025</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>24</td>
</tr>
<tr>
<td>D063</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>9</td>
</tr>
<tr>
<td>D054</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>30</td>
<td>.</td>
</tr>
<tr>
<td>F030</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>9</td>
</tr>
<tr>
<td>C006</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>D022</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>42</td>
</tr>
<tr>
<td>C012</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>30</td>
</tr>
<tr>
<td>D009</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>C027</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>E066</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>6</td>
</tr>
<tr>
<td>E066</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>.</td>
</tr>
<tr>
<td>B034</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>.</td>
</tr>
<tr>
<td>B034</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>18</td>
</tr>
<tr>
<td>E064</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>E064</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>12</td>
</tr>
<tr>
<td>D006</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>D006</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>15</td>
</tr>
<tr>
<td>E073</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>6</td>
</tr>
<tr>
<td>E073</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>.</td>
</tr>
<tr>
<td>E052</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>21</td>
</tr>
<tr>
<td>E052</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>.</td>
</tr>
<tr>
<td>E014</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>.</td>
</tr>
<tr>
<td>E014</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>.</td>
</tr>
<tr>
<td>F018</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>F018</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>3</td>
</tr>
</tbody>
</table>
Table D.1 (Continued)

<table>
<thead>
<tr>
<th>ID</th>
<th>SEX</th>
<th>DM</th>
<th>MI</th>
<th>HTN</th>
<th>Pair</th>
<th>Fail</th>
<th>Cen</th>
</tr>
</thead>
<tbody>
<tr>
<td>E059</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>E059</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>D069</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>12</td>
<td>.</td>
</tr>
<tr>
<td>D069</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>D038</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>6</td>
</tr>
<tr>
<td>D038</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>.</td>
</tr>
<tr>
<td>E006</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>E006</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>9</td>
</tr>
<tr>
<td>F054</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>12</td>
</tr>
<tr>
<td>F054</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>12</td>
</tr>
<tr>
<td>E009</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>24</td>
</tr>
<tr>
<td>E009</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>24</td>
</tr>
<tr>
<td>E027</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>E027</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>B012</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>.</td>
</tr>
<tr>
<td>B012</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>6</td>
</tr>
<tr>
<td>F061</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>F061</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>12</td>
</tr>
<tr>
<td>C018</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>12</td>
</tr>
<tr>
<td>C018</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>.</td>
</tr>
<tr>
<td>F075</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>6</td>
</tr>
<tr>
<td>F075</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>6</td>
</tr>
<tr>
<td>E023</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>E023</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>E071</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>.</td>
</tr>
<tr>
<td>E071</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>.</td>
</tr>
<tr>
<td>B040</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>B040</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>F074</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>6</td>
</tr>
<tr>
<td>F074</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>.</td>
</tr>
</tbody>
</table>
Table D.1 (Continued)

<table>
<thead>
<tr>
<th>ID</th>
<th>SEX</th>
<th>DM</th>
<th>MI</th>
<th>HTN</th>
<th>Pair</th>
<th>Fail</th>
<th>Cen</th>
</tr>
</thead>
<tbody>
<tr>
<td>E067</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>E067</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>E068</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>E068</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>E022</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>18</td>
</tr>
<tr>
<td>E022</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>18</td>
</tr>
<tr>
<td>E060</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>E060</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>F084</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>9</td>
</tr>
<tr>
<td>F084</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>9</td>
</tr>
<tr>
<td>C034</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>18</td>
</tr>
<tr>
<td>C034</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>18</td>
</tr>
<tr>
<td>C002</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>30</td>
</tr>
<tr>
<td>C002</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>B038</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>B038</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>B038</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>C038</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>C038</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>C038</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>.</td>
<td>9</td>
</tr>
<tr>
<td>F031</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>.</td>
<td>36</td>
</tr>
<tr>
<td>F031</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>F031</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>E058</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>E058</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>E058</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>.</td>
<td>6</td>
</tr>
<tr>
<td>B021</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>B021</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>B021</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>B021</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>D021</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>.</td>
</tr>
<tr>
<td>D021</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>.</td>
</tr>
<tr>
<td>D021</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>18</td>
<td>.</td>
</tr>
<tr>
<td>D021</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>.</td>
</tr>
</tbody>
</table>