NONPARAMETRIC ESTIMATION FROM CENSORED BIVARIATE OBSERVATIONS

BY

ALVARO MUÑOZ

TECHNICAL REPORT NO. 60
AUGUST 1980

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT 2 RO1 GM21215-06

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
NONPARAMETRIC ESTIMATION FROM CENSORED BIVARIATE OBSERVATIONS

by

Alvaro Muñoz

TECHNICAL REPORT NO. 60
August 1980

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT 2 RO1 GM21215-06

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ACKNOWLEDGEMENTS

I would like to thank Professor Byron Wm. Brown, Jr. for his expert advice and strong encouragement during the preparation of this thesis. My thanks also to Professor Rupert G. Miller, Jr. for his enlightening comments during the course of this research and for his invaluable writing guidance. Their direction and extremely good spirits were essential for this endeavor.

I thank Professor Bradley Efron for consenting to be a reader of the dissertation and Dr. Lynn Spitler for letting me use the data in the final chapter.

Sincere thanks to Beatriz who gave me the best encouragement during my graduate work; to my parents for their confidence throughout my entire education, and to Daniel who was born with this thesis (both at the same time and, somehow, with it).

For the very fine job of typing this manuscript, I thank Karola Decleve.
## Table of Contents

1. **Introduction**  
   1.1 The Kaplan-Meier Estimator  
   1.2 The Bivariate Censored Data Problem  
   1.3 Summary of Results  

2. **The Estimator and Its Computation**  
   2.1 The Generalized Maximum Likelihood Estimator, $\hat{S}$  
   2.2 Computation of $\hat{S}$ using the Self-consistency Property  
   2.3 A Third Algorithm to Compute $\hat{S}$  

3. **The Generalized Maximum Likelihood Estimator for the Bivariate Censored Data Problem**  
   3.1 Definition of Generalized Maximum Likelihood Estimators  
   3.2 The GMLE for the Bivariate Censored Data Problem  
   3.3 Comparing the Univariate and Bivariate GMLE  

4. **Self-consistency**  
   4.1 Definition of Self-consistency  
   4.2 Self-consistency and Generalized Maximum Likelihood Estimation
5. EXAMPLE

5.1 The Data

5.2 The Estimator $\hat{S}$

5.3 The Kaplan-Meier Estimator for the Survival vs. the Marginal Estimator from $\hat{S}$

5.4 Redistributing the Line Masses of $\hat{S}$

5.5 Mutual Inconsistency of $\hat{S}$ and $\hat{S}_0$

REFERENCES
1. INTRODUCTION

In a variety of experiments (e.g., medical clinical studies) one is interested in estimating the probability that some failure (e.g., death) will not happen before a specific time. In some of these studies all individuals are not observed to this time of failure. At the time of the analysis some individuals may not have failed, and the length of observation may vary from one individual to the next.

Kaplan and Meier (1958) proposed an estimator for the case of one failure time per individual. This paper focuses on an analogous multivariate problem, where one can observe two failure times per individual, one or both of which might be censored. The goal is to estimate the joint probabilities involving the two failure times. The relevance of the problem is shown by numerous examples. In some cases the data are naturally paired; such as observations on kidneys. Sometimes there are two sequential observations on the same individual (e.g., relapse and death). The censoring on these examples is univariate and it is the type of censoring this paper will concentrate on.

1.1 The Kaplan-Meier Estimator

Let $T_1, \ldots, T_n$ and $L_1, \ldots, L_n$ be i.i.d. sequences of positive random variables with $S_1(t) = P(T > t)$ and $S_0(t) = P(L > t)$. Let the two sequences be independent of each other.
The $T_i$ are the failure times of the observational units and the $L_i$ are the censoring times. The data available to the statistician are the $n$ pairs $(t_1, d_1), \ldots, (t_n, d_n)$ where $t_i = T_i \wedge L_i = \min(T_i, L_i)$ and $d_i = I(t_i = T_i)$. (Throughout, $I(A)$ is the indicator function of the set $A$.)

The Kaplan-Meier estimator (hereafter referred to as KME) is given by

$$\hat{S}_1(t) = \prod_{i=1}^{\left\lfloor t \right\rfloor} \left( \frac{n-i}{n-i+1} \right)^{d_i} \quad \text{for } t \leq t(n),$$

where $t(1) < t(2) < \ldots < t(n)$. For $t > t(n)$, $\hat{S}_1(t) = 0$ if $t(n)$ is an uncensored observation and it is undefined otherwise.

Kaplan and Meier (1958) derived $\hat{S}_1(t)$ from a product of conditional probabilities estimated by observed binomial proportions. They also proved that this estimator is the maximum likelihood estimator in the nonparametric sense or, as it is called by Bailey (1979), the generalized maximum likelihood estimator (hereafter referred to as GMLE). This fact is also proved by Johansen (1978), who gives a more precise definition of the generalized maximum likelihood estimator. Johansen and Kaplan and Meier referred to Kiefer and Wolfowitz (1956) for such a definition.

Efron (1967) introduced the concept of self-consistency for an estimator and then proved that the KME is the only self-consistent
estimator for the univariate censored data problem. He also described an algorithm for calculating the KME which uses what he called the redistribute-to-the-right distribution function. These two ideas were applied by Peterson (1975) in the competing risks problem.

The asymptotic properties of the KME have been studied by different authors. Strong consistency is discussed by Meier (1975) using the product form of the estimator; by Peterson (1977) who expressed the estimator as a function of empirical subsurvival functions, and by Langberg et al (1979) using competing risks theory. The uniformity of the strong consistency has been proved recently by Földes and Rejtö (1980). Asymptotic normality is discussed by Breslow and Crowley (1974), and Aalen (1976).

1.2 The Bivariate Censored Data Problem

The problem arises when there are two failure times for each observational unit. Cox (1972) suggested that the problem could be obviated by the use of conditional hazard rates and time dependent covariates. However, as Kalbfleisch and Prentice (1979) noted, the joint analysis of data with several failure times would be of interest in itself. This paper discusses some aspects of such a joint analysis.

Let $T_{11} = (T_{11}, T_{21}), \ldots, T_{n} = (T_{1n}, T_{2n})$ be i.i.d. random vectors distributed with $S(t_1, t_2) = P\{T_1 > t_1, T_2 > t_2\}$ where
\[ S(0,0) = 1. \] Below, we refer to \( S \) as a survival function by analogy with the univariate case. Let the marginal distributions be denoted by \( S_1(t) = S(t,0) \) and \( S_2(t) = S(0,t) \). The \( T_i \) are the bivariate failure times.

Let \( L_1, \ldots, L_n \) be an independent sequence of i.i.d. random variables with distribution \( S_0(t) = P\{L > t\} \) where \( S_0(0) = 1 \). These are the censoring times.

To avoid ambiguities in the definition of the \( d_i \) below, assume that \( (A_{S_1} \cup A_{S_2}) \cap A_{S_0} = \emptyset \) where \( A_{S_j} \) is the set of jump points for \( S_j \).

The data will consist of

\[ t_i = (T_{1i}^{L_{1i}}, T_{2i}^{L_{1i}}), \]

\[ d_i = (I(t_{1i} = T_{1i}), I(t_{2i} = T_{2i})). \]

Using these data we want to estimate \( S \).

Note that the observations censored as \((1,0) \cap (0,1)\) lie above [below] the equiangular line of the first quadrant and the ones censored as \((0,0)\) lie on the equiangular line.

The observations which are going to be considered as censored for the estimation of \( S(t_1, t_2) \) are those whose failure time vector could have been an element of \((t_1, \infty) \times (t_2, \infty)\). For example,
if \( t_1 > t_2 \), the censored observations for estimating \( S(t_1, t_2) \) are the \((1,0)\) observations inside the triangle with vertices \((t_2, t_2), (t_1, t_2), (t_1, t_1)\) and the \((0,0)\) observations on the line from the origin to \((t_1, t_1)\).

The estimation of \( S(t_1, t_2) \) is of primary interest, but it should be mentioned that estimation of univariate distributions of functions of \( T_1 \) and \( T_2 \) may be of interest also (e.g., \( T_1, T_2, T_1 \wedge T_2, T_1 \vee T_2 = \max(T_1, T_2), |T_1 - T_2| \)). The KME of such a function may differ from the one derived from the bivariate estimator proposed for \( S(t_1, t_2) \) and based on all the data. For example, using KME for \( P(|T_1 - T_2| > t) \) will amount to discarding (or, equivalently, censoring at zero) the observations with \( \tilde{d} = (0,0) \) whereas the bivariate estimator uses these observations and results in a different estimator for the univariate function. Further, one can construct an example where all the estimators are computed in the Kaplan-Meier way and \( \hat{P}(|T_1 - T_2| > t) \) is not \( \hat{P}(T_1 \vee T_2 - T_1 \wedge T_2 > t) \)
\( = \sum_x \hat{P}(T_1 \vee T_2 > t + x | T_1 \wedge T_2 = x) \hat{P}(T_1 \wedge T_2 = x) \) as it should be. However, for the estimation of \( P(T_1 \wedge T_2 > t) \) both methods will always agree with each other and with the estimator achieved by using competing risks theory (see Peterson (1975)).

Having stated the bivariate problem, we should point out some of its differences from the corresponding univariate problem. The symmetry between the failure times and censoring variables is not present in the bivariate problem. But, more importantly, the
bivariate problem can not be treated as a competing risks problem. Our problem is not a series system in $T_1$ and $T_2$, but a parallel system where the life of the system is given by $\max(T_1,T_2)$ rather than by $\min(T_1,T_2)$.

1.3 Summary of Results

In Chapter 2 we will display the estimator we propose for $S(t_1,t_2)$ and will indicate several ways to compute it.

In Chapter 3 we will define the generalized maximum likelihood estimator. We will prove that our estimator is of this type.

In Chapter 4 the self-consistency property is introduced in the context of our problem and we will prove that the GMLE is the only self-consistent estimator. Campbell (1979) has investigated this self-consistency property for grouped data with censoring where the censoring variables are considered to be of the bivariate type too. Korwar (1980) extended Campbell's results when the data may be both left and right censored.

An example is worked out in Chapter 5.

The consistency of the estimator is proved in a companion paper. (See Muñoz (1980).)
2. THE ESTIMATOR AND ITS COMPUTATION

The estimator \( \hat{S} \) which we propose for \( S \) is the GMLE. The first section of this chapter displays where and how \( \hat{S} \) spreads the probability mass. The derivation of \( \hat{S} \) is left for Chapter 3.

The second section provides an algorithm to compute \( \hat{S} \) directly. This algorithm is an application of the self-consistency property of the GMLE proved in Chapter 4.

The final section provides a third way to compute the GMLE. It is the extension of Efron's (1967) redistribute-to-the-right algorithm for the KME.

2.1 The Generalized Maximum Likelihood Estimator

We first introduce some notation:

\[
(t_1, t_2)^{\infty, t_2} = \{ x : x_1 > t_1, x_2 = t_2 \} = \text{horizontal line at } t_2 \text{ from } t_1 \text{ to } \infty,
\]

\[
(t_1, t_2)^{t_1, \infty} = \{ x : x_1 = t_1, x_2 > t_2 \} = \text{vertical line at } t_1 \text{ from } t_2 \text{ to } \infty,
\]

\[
\#(\sim t_1, \preceq t_2) = \sum_{i=1}^{n} I(t_{1i} \sim t_1, t_{2i} \preceq t_2),
\]

where \( \sim \) could be \( >, \geq, \text{ or } = \).

In Chapter 3 we prove that a GMLE puts positive mass on points \( t_i \) if \( d_i = (1,1) \), on lines \( (t_{1i}, t_{2i})^{\infty, t_{2i}} \) \( [(t_{1i}, t_{2i})(t_{1i}, \infty)] \).
if the lines contain no observations and \(d_{\perp} = (0,1)\) \([d_{\perp} = (1,0)]\),
and on the region \((t_{11}, \infty) \times (t_{21}, \infty)\) if it contains no observations
and \(d_{\perp} = (0,0)\).

We then will prove that the mass \(\hat{S}\) associates with \(\perp\) (on
a point, line, or region) is given by

\[
\frac{\#(=t_{1i},=t_{2j})}{n} \times \prod_{d_{j}=(0,1)} \frac{\#(>t_{1j},=t_{2j})}{\#(>t_{1j},=t_{2j})}
\]

\(t_{2i}=t_{2j}<t_{1j}<t_{1i}\)

\(2.1.1\)

\[
\times \prod_{d_{j}=(1,0)} \frac{\#(=t_{1j},>t_{2j})}{\#(=t_{1j},>t_{2j})} \times \prod_{d_{j}=(0,0)} \frac{\#(>t_{1j},>t_{2j})}{\#(>t_{1j},>t_{2j})}
\]

\(t_{1i}=t_{1j}<t_{2j}<t_{2i}\)

\(t_{1j}=t_{2j}<t_{1i}<t_{1i}\)

where the products are performed for the distinct observations,
modified by the convention that uncensored observations precede censored observations in case of ties. That is, if there are \((1,1)\)
and \((1,0)\) \([(0,1)]\) observations tied at \(\perp\), consider the \((1,0)\)
\([(0,1)]\) censored observations at \((t_{11}, t_{21})\) \([(t_{11}, t_{21})]\); and if there are uncensored or partially censored observations at the
horizontal and vertical lines from \((t_{1i}, t_{1i})\) where \(d_{\perp} = (0,0)\),
consider the \((0,0)\) censored observations at \((t_{1i}, t_{1i})\). The
above convention is the extension of the convention "deaths precede
losses" used for the KME.
One should note that either the second or third term in the expression above is one according to whether \( t_{1i} < t_{2i} \) or \( t_{1i} > t_{2i} \), respectively. Both terms will be one if \( t_{1i} = t_{2i} \).

This formula for the mass that \( \hat{S} \) associates with \( t_{1i} \) shows how the mass carried by an observation (i.e., \( n^{-1} \)) is modified (increased) by the censored observations which affect it. This is exactly analogous to the KME.

### 2.2 Computation of \( \hat{S} \) using the Self-consistency Property

In Chapter 4 we prove that a self-consistent estimator for \( S \) spreads the mass as a GMLE does. In that sense, one says that the GMLE is the only self-consistent estimator. This section gives an algorithm to compute a self-consistent estimator, hence a GMLE.

A self-consistent estimator \( \hat{S} \) for \( S \) is a survival function such that, for all \( t_1, t_2 \),

\[
\begin{align*}
n \hat{S}(t_1, t_2) &= \#(> t_1, > t_2) + \\
                      &+ \sum_{d_1 = (0, 1)} \frac{\hat{S}(t_1, t_{2i}^-) - \hat{S}(t_1, t_{2i}^+)}{\hat{S}(t_{1i}, t_{2i}^-) - \hat{S}(t_{1i}, t_{2i}^+)}
\end{align*}
\]

\[(2.2.1)\]

\[
+ \sum_{d_1 = (1, 0)} \frac{\hat{S}(t_{1i}, t_i) - \hat{S}(t_{1i}, t_{2i}^-)}{\hat{S}(t_{1i}, t_{2i}^-) - \hat{S}(t_{1i}, t_{2i}^+)} + \sum_{d_1 = (0, 0)} \frac{\hat{S}(t_{1i}, t_{2i}^-) - \hat{S}(t_{1i}, t_{2i}^+)}{\hat{S}(t_{1i}, t_{2i}^+)}
\]

where it is assumed without loss of generality that there are no ties among the (partial or double) censored observations.
The following proposition shows that a self-consistent estimator is also self-consistent for the estimation of the conditional probabilities $P\{T_1 > t_1 | T_1 > t_1', T_2 = t_2\}$ and $P\{T_1 > t_1, T_2 > t_2 | T_1 > t_1', T_2 > t_1'\}$.

**Proposition 2.2.1.** If $\hat{S}$ is a survival function, $\hat{S}$ is self-consistent if and only if (wlog) for $t_1 > t_2$ we have

$$\frac{\hat{S}(t_1, t_2) - \hat{S}(t_1', t_2)}{\hat{S}(t_1', t_2) - \hat{S}(t_1, t_2)} = \frac{1}{\#(>t_1', =t_2)} \left[ \#(>t_1', =t_2) + \sum_{d_1 = (0,1)}^{t_1 < t_{1i} < t_1} \frac{\hat{S}(t_1, t_2) - \hat{S}(t_1, t_2)}{\hat{S}(t_{1i}, t_2) - \hat{S}(t_{1i}, t_2)} \right]$$

(2.2.2)

for $t_1' = t_1$

and

$$\frac{\hat{S}(t_1, t_2 v t')} {\hat{S}(t', t')} = \frac{1}{\#(>t', >t')} \left[ \#(>t_1, >t_2 v t') + \sum_{d_1 = (0,1)}^{t_2 v t' < t_{2i} < t_{1i} < t_1} \frac{\hat{S}(t_1, t_{2i}) - \hat{S}(t_1, t_{2i})} {\hat{S}(t_{1i}, t_{2i}) - \hat{S}(t_{1i}, t_{2i})} \right]$$

(2.2.3)

$$+ \sum_{d_1 = (0,0)}^{t' < t_1 = t_{2i} < t_1} \frac{\hat{S}(t_1, t_{2i})} {\hat{S}(t_{1i}, t_{2i})}$$

for $t' < t_1$.
This proposition is restated and proved as Lemma 4.2.1.

Now we proceed to give a recursive formula for equations (2.2.2) and (2.2.3) and hence for \( \hat{S} \). We start with a naive estimator which underestimates the conditional probabilities on these equations and the recursion will increase them using the information the data provide.

As in the proposition above, we give the recursive formula for the computation of \( \hat{S} \) at \((t_1', t_2')\) where \( t_1 > t_2 \). The case \( t_1 < t_2 \) is analogous and the case \( t_1 = t_2 \) is particularly simple. Needless to say, the algorithm is to be used on points \((t_1', t_2')\) for which the estimator is not undefined due to "last" censored observations. (See Chapter 3).

Let

\[
\hat{S}_0(t_1', t_2') = \frac{1}{n} \#(>t_1', t_2') \quad \text{for all } t_1', t_2';
\]

then for \( t_1 \) such that \( d_1 = (0,1), t_2 < t_{21} < t_{11} \leq t_1 \) and \( \#(>t_{11}', =t_{21}') > 0 \)

\[
(2.2.4) \quad \frac{\hat{S}_0(t_1', t_{21}')-\hat{S}_0(t_1', t_{21})}{\hat{S}_0(t_{11}', t_{21}')-\hat{S}_0(t_{11}', t_{21})} = \frac{\#(>t_1', =t_{21}')}{\#(>t_{11}', =t_{21}')},
\]

and for \( t_1 \) such that \( d_1 = (0,0), t_{11} = t_{21} \leq t_1 \) and \( \#(>t_{11}', >t_{21}') > 0 \)
\[
\frac{\hat{S}_0(t_{11}, t_{21})}{\hat{S}_0(t_{11}, t_{21})} = \frac{\#(t_{11} > t_{21})}{\#(t_{11} > t_{21})}.
\]

For \( k \geq 0 \), let \( \hat{S}_{k+1} \) be a survival function such that for the same points as above

\[
\frac{\hat{S}_{k+1}(t_{11}, t_{21}) - \hat{S}_{k+1}(t_{11}, t_{21})}{\hat{S}_{k+1}(t_{11}, t_{21}) - \hat{S}_{k+1}(t_{11}, t_{21})}
\]

\[
\left(2.2.6\right) \quad = \frac{1}{\#(t_{11} = t_{21})} \left[ \#(t_{11} = t_{21}) + \sum_{d_j = (0, 1)} \frac{\hat{S}_{k}(t_{11}, t_{21}) - \hat{S}_{k}(t_{11}, t_{21})}{\hat{S}_{k}(t_{11}, t_{21}) - \hat{S}_{k}(t_{11}, t_{21})} \right]
\]

and

\[
\frac{\hat{S}_{k+1}(t_{11}, t_{21})}{\hat{S}_{k+1}(t_{11}, t_{21})} = \frac{1}{\#(t_{11} > t_{21})} \left[ \#(t_{11} > t_{21}) + \sum_{d_j = (0, 0)} \frac{\hat{S}_{k}(t_{11}, t_{21})}{\hat{S}_{k}(t_{11}, t_{21})} \right]
\]

\[
\left(2.2.7\right) \quad + \sum_{d_j = (0, 0)} \frac{\hat{S}_{k}(t_{11}, t_{21})}{\hat{S}_{k}(t_{11}, t_{21})} \left[ \right]
\]

\[
+ \sum_{d_j = (0, 1)} \frac{\hat{S}_{k}(t_{11}, t_{21}) - \hat{S}_{k}(t_{11}, t_{21})}{\hat{S}_{k}(t_{11}, t_{21}) - \hat{S}_{k}(t_{11}, t_{21})} \right].
\]

12
Therefore, by extending (2.2.7) to \((t_{1i}, t_{2i}) = (0, 0)\)

\[
\hat{S}_{k+1}(t_1, t_2) = \frac{1}{n} \left\{ \#(>t_1, >t_2) \right. 

\left. + \sum_{d_1=(0,1)} \sum_{t_{2i}<t_{2i}<t_{1i}} \frac{\hat{S}_k(t_{1i}, t_{2i}) - \hat{S}_k(t_{1i}, t_{2i})}{\hat{S}_k(t_{1i}, t_{2i}) - \hat{S}_k(t_{1i}, t_{2i})} \right. 

\left. + \sum_{d_1=(0,0)} \sum_{t_{1i}=t_{2i}<t_{1i}} \frac{\hat{S}_k(t_{1i}, t_{2i})}{\hat{S}_k(t_{1i}, t_{2i})} \right. 

\right\}.
\]

(2.2.8)

The algorithm starts with (2.2.4) and (2.2.5) and then recursively uses (2.2.6) and (2.2.7); and when these settle down (i.e., step \(k+1 = \text{step } k\)), one substitutes them in (2.2.8) and, by proposition 2.2.1, the result provides a self-consistent estimator. It is easy to see that such an estimator will be achieved at least at the step equal to the number of censored observations with respect to \((t_1, \infty) \times (t_2, \infty)\). One could also see that equations (2.2.6) and (2.2.7) will first settle down for the closest censored observations with respect to \((t_1, \infty) \times (t_2, \infty)\) and subsequently for censored observations further apart from \((t_1, \infty) \times (t_2, \infty)\).
2.3 A Third Algorithm to Compute $\hat{S}$

Efron (1967) gave an algorithm to compute the K-M estimator by redistributing to the right the mass carried by the censored observations. Peterson (1975) extended it to the competing risks problem. This section extends such an algorithm to the computation of the masses for the bivariate case.

To indicate the two stages of the algorithm we assume, without loss of generality, that there are no ties among the observations. The algorithm redistributes the mass of the censored observations for which $\hat{S}$ does not associate any mass. The reader is referred to Section 2.1 to recall how $\hat{S}$ spreads its mass.

Before any redistribution, one associates a mass of $\frac{1}{n}$ with each observation.

The first stage redistributes the mass of the $(0,0)$-observations for which the regions $(t_{11}^\infty, t_{11}^\infty)$, associated with them, contain observations. If $(t_{11}, t_{11})$ is the $(0,0)$-observation closest to the origin, increase the mass associated with an observation belonging to $(t_{11}^\infty, t_{11}^\infty)$ by $[n \times \#(t_{11}, t_{11})]^{-1}$. Recursively, do the same for the next $(0,0)$-observation next furthest away from the origin. As a consequence, the mass associated with a point $t_{ij}$ will be

$$\frac{1}{n} \prod_{d_j = (0,0)} \frac{1 + \#(t_{1j}, t_{1j})}{\#(t_{1j}, t_{2j})} \cdot \frac{t_{1j} - t_{2j} < t_{1i} \wedge t_{2i}}{t_{1j} - t_{2j}}$$
The second stage redistributes the mass of the (0,1)-observations for which the lines \((t_{11}, t_{21})^{(\infty, t_{21})}\) contain observations. (The algorithm for the (1,0)-observations is similar.) If \((t_{1(1)}, t_{21})\) is among the (0,1)-observations with second component equal to \(t_{2i}\), the closest to the equiangular line, increase the mass associated with an observation belonging to \((t_{1(1)}, t_{21})^{(\infty, t_{21})}\) by

\[
\frac{1}{\sum_{(t_{1(1)}, t_{21})}} \times \frac{1}{n} \times \prod_{d_j=(0,0)} \frac{1 + \#(t_{1j}, t_{2j})}{\#(t_{1j}, t_{2j})}.
\]

Recursively, do the same for the (0,1)-observation next furthest away from the equiangular line with second component equal to \(t_{2i}\). As a consequence, the mass associated with an observation \(t_{1i}\), assuming wlog \(t_{1i} > t_{2i}\), will be

\[
\frac{1}{n} \times \prod_{d_j=(0,1)} \frac{1 + \#(t_{1j}, t_{2j})}{\#(t_{1j}, t_{2j})} \times \prod_{d_j=(0,0)} \frac{1 + \#(t_{1j}, t_{2j})}{\#(t_{1j}, t_{2j})}.
\]

The last expression is just (2.1.1) for the case of no ties as assumed. Hence, the algorithm gives the mass that \(\hat{S}\) associates with an observation \(t_{1i}\) as it was described in Section 2.1.
3. THE GENERALIZED MAXIMUM LIKELIHOOD ESTIMATOR FOR THE BIVARIATE CENSORED DATA PROBLEM.

In this chapter we will show that the estimator described in Chapter 2 is the maximum likelihood estimator.

3.1 Definition of Generalized Maximum Likelihood Estimators (GMLE)

Our goal is to find the maximum likelihood estimator for the joint distribution of \( T_1 \) and \( T_2 \) in the class of all distribution functions. Being a non-dominated family, the classical maximization of densities is not applicable. Following Johansen (1978) and Kiefer and Wolfowitz (1956) we generalize the definition of maximum likelihood as follows.

**Definition.** Let \( \mathcal{P} \) be a family of probability measures. For \( P_1, P_2 \in \mathcal{P} \) let \( f(x, P_1, P_2) = \frac{dP_1}{d(P_1 + P_2)} (x) \). If \( x \) is the observed vector, \( \hat{P} \) is a maximum likelihood estimator if

\[
(3.1.1) \quad f(x, \hat{P}, P) \geq f(x, P, \hat{P}) \quad \text{for all } P \in \mathcal{P}.
\]

As Johansen noted, if the family is dominated, the definition reduces to the classical one and, what is more relevant for our problem, if \( \hat{P} \) gives positive probability to the observed data, then (3.1.1) reduces to

\[
(3.1.2) \quad \hat{P}(x) > P(x) \quad \text{for all } P \in \mathcal{P} \quad \text{such that } P(x) > 0.
\]
It is easily seen that the estimator described in Chapter 2 assigns positive probability to the observed data. To prove that the estimator is a GMLE, we can restrict attention to the sub-family of survival functions with the same property by virtue of (3.1.2).

### 3.2 The GMLE for the Bivariate Censored Data Problem

As was stated in section 1.2, our problem focuses on the estimation of the joint survival function $S$ of $(T_1, T_2)$ when one observes, for $1 \leq i \leq n$

\[
t_i = (T_{1i} \wedge L_i, T_{2i} \wedge L_i)
\]

and

\[
d_i = (I(t_{1i} = T_{1i}), I(t_{2i} = T_{2i})) = \begin{cases} 
(1,1) & \text{if } T_{1i} \lor T_{2i} < L_i, \\
(1,0) & \text{if } T_{1i} < L_i < T_{2i}, \\
(0,1) & \text{if } T_{2i} < L_i < T_{1i}, \\
(0,0) & \text{if } L_i < T_{1i} \lor T_{2i}.
\end{cases}
\]

We need to introduce more notation

\[
(t_{1i}, t_{2i})_{(t_{1i}, \infty)} = \{ t : t_{1i} = t_{1i}, t > t_{2i} \},
\]

17
i.e., the right open line at \( t_{11} \), from \( t_{21} \), and, the right closed line

\[
\overline{(t_{11}, t_{21})} \cap (t_{11}, \infty) = \{ t : t_1 = t_{11}, \ t_2 > t_{21} \}.
\]

Analogously, we have \( (t_{11}, t_{21}) \cap (\infty, t_{21}) \) and \( \overline{(t_{11}, t_{21})} \cap (\infty, t_{21}) \).

\[
(t_{11}, \infty) \cap (t_{11}, t_{11}) \cap (t_{11}, \infty) \cap (t_{11}, t_{11}) = \overline{(t_{11}, t_{11})} \cap (t_{11}, \infty) \cap (t_{11}, t_{11}) = (t_{11}, \infty) 
\]

and

\[
(t_{11}, \infty) \times (t_{21}, \infty) = \{ t : t_1 > t_{11}, \ t_2 > t_{21} \}.
\]

Let

\[ A_{11} = \{ t_i : d_i = (1,1) \} \]

\[ A_{10} = \{ t_i : d_i = (1,0) \ \text{and} \ \overline{(t_{11}, t_{21})} \cap (t_{11}, \infty) \ \text{contains no observations} \} \]

\[ A_{01} = \{ t_i : d_i = (0,1) \ \text{and} \ (t_{11}, t_{21}) \cap (\infty, t_{21}) \ \text{contains no observations} \} \]

\[ A_{00} = \{ t_i : d_i = (0,0) \ \text{and} \ (t_{11}, \infty) \times (t_{11}, \infty) \ \text{contains no observations} \} \]

\[ A_1 = A_{11} \cup A_{10} \ \text{and} \ A_{-1} = A_{11} \cup A_{01} \]

\[ A = A_{11} \cup A_{10} \cup A_{01} \cup A_{00}. \]
$A_{11}$ is the set of the double uncensored observations, $A_{10}$ [$A_{01}$] is the set of "last" (1,0) [(0,1)] observations and $A_{00}$ is the set of the "last" (0,0) observation.

It is easily seen that a survival function, which assigns positive probability to the data, has to assign positive mass to the elements of $A_{11}$, to the lines associated with the elements of $A_{10}$ and $A_{01}$ and to the region associated with the element of $A_{00}$. Let $S^+$ denote the subclass of such survival functions; thus, $S^+$ is the set of survival functions, $S$, such that for all $t_i \in A$

\[
S(t_i) I(t_i \in A_{11}) + S\{(t_{11}, t_{21})(t_{11}', \infty)\} I(t_i \in A_{10})
\]

(3.2.1)

\[
+ S\{(t_{11}, t_{21})(\infty, t_{21})\} I(t_i \in A_{01}) + S(t_i) I(t_i \in A_{00}) > 0,
\]

where $S(\cdot)$ denotes the mass $S$ puts on $\cdot$, and $S(\cdot)$ the value of $S$ at $\cdot$. In seeking a GMLE, one can restrict attention to $S^+$.

Analogously, in seeking a GMLE for $S_0$ (the survival function of $L$) one can restrict attention to $S_0^+$ which would be the set of univariate survival functions, $S_0$, such that $S_0(t_{11}, t_{21}) > 0$ for $t_i \in A_{11}$ and $S_0\left(\max_{d_i=(1,1)} \{t_{11}, t_{21}\}\right) > 0$. 

19
We now proceed to write the likelihood of the data for a pair of survival functions \((S, S_0)\) in \(S^+ \times S_0^+\). If \(p_{S, S_0}\) denotes the probability when the survival functions are \(S\) and \(S_0\); the likelihood of the data is

\[
\prod_{i=1}^{n} \prod_{j=0}^{1} \prod_{k=0}^{1} \left[ p_{S, S_0}(T_1^L = t_{1i}, T_2^L = t_{2i}, d_i = (j,k)) \right]^{I(d_i = (j,k))}
\]

\[= \prod_{i=1}^{n} \left[ p_{S, S_0}(T_1 = t_{1i}, T_2 = t_{2i}, L > t_{1i} \lor t_{2i}) \right]^{I(d_i = (1,1))}
\]

\[
\times \left[ p_{S, S_0}(T_1 = t_{1i}, L = t_{2i}, T_2 > t_{2i}) \right]^{I(d_i = (1,0))}
\]

\[
\times \left[ p_{S, S_0}(L = t_{1i}, T_2 = t_{2i}, T_1 > t_{1i}) \right]^{I(d_i = (0,1))}
\]

\[
\times \left[ p_{S, S_0}(L = t_{1i}, T_1 > t_{1i}, T_2 > t_{1i}) \right]^{I(d_i = (0,0))}
\]

\[= \prod_{i=1}^{n} \left[ s(t_{1i}) s_0(t_{1i} \lor t_{2i}) \right]^{I(d_i = (1,1))}
\]

\[
\times \left[ s((t_{1i}, t_{2i})_{(t_{1i}, \infty)}) s_0(t_{2i}) \right]^{I(d_i = (1,0))}
\]

\[
\times \left[ s((t_{1i}, t_{2i})_{(\infty, t_{2i})}) s_0(t_{1i}) \right]^{I(d_i = (0,1))}
\]

\[
\times \left[ s(t_{1i}) s_0(t_{1i}) \right]^{I(d_i = (0,0))}
\]
\[
= \prod_{i=1}^{n} \left[ S(t_{i1}) \right]^{I(d_{i1}=(1,1))} \left[ S((t_{11},t_{21})(t_{11},\omega)) \right]^{I(d_{i1}=(0,1))} \times \left[ S((t_{11},t_{21})(\omega,t_{21})) \right]^{I(d_{i1}=(0,0))} \left[ S(t_{i1}) \right]^{I(d_{i1}=(0,0))} \\
\times \prod_{i=1}^{n} \left[ S_{0}(t_{11},v_{21}) \right]^{I(d_{i1}=(1,1))} \left[ S_{0}(t_{11},v_{t_{21}}) \right]^{I(d_{i1}=(1,1))} \left[ S_{0}(t_{11},v_{21}) \right]^{I(d_{i1}=(1,1))}
\]

where the second equality follows from the independence assumption between \( \tau \) and \( L \). The first [second] product in the last expression depends only on \( S \) [\( S_{0} \)] and so it will be denoted by \( L(S) [L_{0}(S_{0})] \). Consequently, finding \( (\hat{S},\hat{S}_{0}) \) which maximize the likelihood above reduces to finding \( \hat{S} [\hat{S}_{0}] \) which maximizes \( L(S) [L_{0}(S_{0})] \). Our problem focuses on \( S \).

The next lemma allows us to restrict the class, where the maximizer of \( L \) is going to be found, to a smaller class than \( \mathcal{G}^{+} \).

**Lemma 3.2.1.** Let \( \mathcal{G}^{++} = \{ S \in \mathcal{G}^{+} : \sum_{t_{i1} \in A} \text{(expression (3.2.1))} = 1 \} \). Then, for all \( S \in \mathcal{G}^{+} \) there exists an \( S^{*} \in \mathcal{G}^{++} \) such that \( L(S^{*}) \geq L(S) \).

**Proof.** The idea of the proof is to derive \( S^{*} \) from \( S \) by increasing the mass at points, lines, and region which are farther away from the origin.

Let \( \tau_{i} = t_{11}^{t_{21}} \) for \( t_{i} \in A \), and \( \tau_{(1)} < \ldots < \tau_{(m)} \) the order of the distinct \( \tau_{i} \)'s.
Let $1 \leq k < m$. One of the following three cases must occur:

(i) $\overline{(\tau(k), \tau(k)) (\tau(k), \infty)} \cap A \neq \emptyset$ and $\overline{(\tau(k), \tau(k)) (\infty, \tau(k))} \cap A \neq \emptyset$.

Let $(\tau(k), t_2(k))$ be the element in $A_{11}$, with first component equal to $\tau(k)$, and the largest second component. If $(\tau(k), t_2(k)) \in A_{11}$, move to this point the mass $S$ puts on $\overline{(\tau(k), \tau(k)) (\tau(k), \infty)} \cap A^c$. If $(\tau(k), t_2(k)) \in A_{10}$, move to the line $\overline{(\tau(k), t_2(k)) (\tau(k), \infty)}$ the mass $S$ puts on $\overline{(\tau(k), t_2(k)) (\tau(k), \infty)} \cap A^c$.

Similarly, let $(t_1(k), \tau(k))$ be the element in $A_{11}$ with second component equal to $\tau(k)$, and the largest first component. If $(t_1(k), \tau(k)) \in A_{11}$, move to this point the mass $S$ puts on $\overline{(\tau(k), \tau(k)) (\infty, \tau(k))} \cap A^c$. If $(t_1(k), \tau(k)) \in A_{01}$, move to the line $\overline{(t_1(k), \tau(k)) (\infty, \tau(k))}$ the mass $S$ puts on $\overline{(t_1(k), \tau(k)) (\tau(k), \infty)} \cap A^c$.

(ii) $\overline{(\tau(k), \tau(k)) (\tau(k), \infty)} \cap A = \emptyset$ (hence, $\overline{(\tau(k), \tau(k)) (\infty, \tau(k))} \cap A \neq \emptyset$).

Let $(t_1(k), \tau(k))$ be defined as in (i) above. If $(t_1(k), \tau(k)) \in A_{11}$, move to this point the mass $S$ puts on $\overline{(\tau(k), \infty) (\tau(k), \tau(k)) (\infty, \tau(k))} \cap A^c$. If $(t_1(k), \tau(k)) \in A_{01}$, move to the line $\overline{(t_1(k), \tau(k)) (\infty, \tau(k))}$ the mass $S$ puts on...
\[
\left( \tau(k), \infty \right) \times \left( \tau(k), \infty \right) \cap A^c.
\]

(iii) \[
\left( \tau(k), \infty \right) \times \left( \tau(k), \infty \right) \cap A = \phi \] (hence, \[
\left( \tau(k), \infty \right) \times \left( \tau(k), \infty \right) \cap A \neq \phi \]).
\]

This case is similar to (ii).

Finally, for \( k = m \) if \( A_{00} = \phi \), do the same as above, but additionally move to the point or line the mass \( S \) puts outside

\[
\bigcup_{k=1}^{m} \left( \tau(k), \infty \right) \times \left( \tau(k), \infty \right). \quad \text{Otherwise, } \quad \left( \tau(m), \infty \right) \times \left( \tau(m), \infty \right) \in A_{00}
\]

move to the region \( \left( \tau(m), \infty \right) \times \left( \tau(m), \infty \right) \) the mass \( S \) puts outside

\[
\bigcup_{k=1}^{m-1} \left( \tau(k), \infty \right) \times \left( \tau(k), \infty \right).
\]

Let \( S^* \) be the survival function one gets by modifying \( S \) as above. Then \( S^* \in S^{++} \), and

\[
S^* \{ \tau_1 \} > S \{ \tau_1 \} \text{ for } \tau_1 \text{ such that } d_1 = (1,1),
\]

\[
S^* \{ (\tau_{11}, \tau_{21}) \} \geq S \{ (\tau_{11}, \tau_{21}) \} \text{ for } \tau_1 \text{ such that } d_1 = (1,0),
\]

\[
S^* \{ (\tau_{11}, \tau_{21}) \} \geq S \{ (\tau_{11}, \tau_{21}) \} \text{ for } \tau_1 \text{ such that } d_1 = (0,1),
\]

\[
S^* \{ \tau_1 \} > S \{ \tau_1 \} \text{ for } \tau_1 \text{ such that } d_1 = (0,0).
\]

Hence, \( L(S^*) \geq L(S) \). □
As a consequence of the lemma, the maximizer of \( L \) is going
to be an element of \( S^{++} \) whose elements spread the whole mass at
the points belonging to \( A_{11} \), at the lines associated with the
elements of \( A_{10} \) and \( A_{01} \), and at the region associated with
\( A_{00} \). Hence, an element of \( S^{++} \) can be parametrized by the value
of those masses.

As before, let \( \tau_k = t_{1k} \wedge t_{2k} \) for \( t_k \in \Lambda \) and
\[ \tau(1) < \ldots < \tau(m) \]
be the ordered distinct \( \tau_k \)'s.

Next, we describe how to label (order) the elements of \( \Lambda \).

One starts at the origin and moves on the equiangular line un-
til \( (t, \infty)(t, t)(\infty, t) \) contains elements of \( \Lambda \). First, one labels
the elements of \( \Lambda \) on the horizontal line \( (t, t)(\infty, t) \) by going
from \( (t, t) \) to \( (\infty, t) \) and then, the elements of \( (t, t)(t, \infty) \) by
going from \( (t, t) \) to \( (t, \infty) \). (If only one of the lines contain
elements of \( \Lambda \), obvious simplification results.) The process is
repeated by moving further away from the origin on the equiangular
line. More precisely,

**Step 0:** Let \( k = 1, \ell = 1 \).

**Step 1:** If \( (\tau(k), t(\ell))(\infty, t(k)) \cap \Lambda = \emptyset \), go to Step 2;
otherwise, let \( t(\ell) = t(k) \) and order, according to the first
component, the elements of \( \Lambda \) with second component equal to
\( t(\ell) \) and first component greater than \( t(\ell) \). Label them as
(t(1), t(2), t(3), ..., t(m)), t(2)). Increase \( \ell \) by one. If \((t(1), t(2)) \cap A \neq \emptyset\), go to Step 2; otherwise, go to Step 3.

Step 2: Let \( t(\ell) = t(k) \) and order, according to the second component, the elements of \( A \) with first component equal to \( t(\ell) \) and second component greater than or equal to \( t(\ell) \). Label them as \((t(1), t(2)), (t(2), t((2))), \ldots, (t(\ell), t((m)))\). Increase \( \ell \) by one and go to Step 3.

Step 3: Increase \( k \) by one. If \( k < m \), go back to Step 1; otherwise, change \( m \) to \( \ell - 1 \) and stop.

Recalling the convention (stated in section 2.1) to avoid ties between censored and uncensored components, we let, for \( 1 \leq i \leq m \) and \( 1 \leq j \leq m \),

\[
\delta_{ij} = \# \text{ of (tied) observations at } (t(i), t((j))) \text{ [or } t((j)), t(i)]
\]

if the \((i,j)\) point is of the form \((t(i), t((j)))\) [or \((t((j)), t(i))\)],

\[
n_{ij} = \begin{cases} 
\# \text{ of } (1,0) \text{ observations in } \frac{(t(i), t(i))(t(i), t((j)))}{(t(i), t(i))(t(i), t((j)))}, \\
\# \text{ or } (0,1) \text{ observations in } \frac{(t(i), t(i))(t((j)), t(i))}{(t(i), t(i))(t((j)), t(i))}, \\
\end{cases}
\]

\[
n_1 = \# \text{ of } (0,0) \text{ observations in } \frac{(0,0)(t(i), t(i))}{(0,0)(t(i), t(i))}.
\]
Define \( n_{10} \equiv 0 = n_0 \). For \( S \in \mathbb{S}^{++} \), we parametrize its masses as

\[
P_{ij} = \begin{cases} 
S\{(t_{(i)}, t_{(j)})\} & \text{if } (t_{(i)}, t_{(j)}) \in A_{11} , \\
or & \\
S\{(t_{(i)}, t_{(j)})(t_{(i)}, \infty)\} & \text{if } (t_{(i)}, t_{(j)}) \in A_{10} , \\
or & \\
S(t_{(i)}, t_{(j)}) & \text{if } (t_{(i)}, t_{(j)}) \in A_{00} , \\
or & \\
S\{(t_{(j)}, t_{(i)})\} & \text{if } (t_{(j)}, t_{(i)}) \in A_{11} , \\
or & \\
S\{(t_{(j)}, t_{(i)})(\infty, t_{(i)})\} & \text{if } (t_{(j)}, t_{(i)}) \in A_{01} .
\end{cases}
\]

Hence, the part of the likelihood which depends on \( S \) (denoted by \( L(S) \)) can be written in terms of the \( P_{ij} \) as follows:

\[
L(S) = L(p) = \prod_{i=1}^{m} \prod_{j=1}^{m_i} \delta_{ij} \prod_{i=1}^{m} \prod_{j=1}^{m_i} \left( \sum_{k=j}^{m_i} p_{ik}^* \right)^{n_{ij} - n_{i,j-1}} \\
\prod_{i=1}^{m} \left( \sum_{k=1}^{m_k} p_{i}^* \right)^{n_i - n_{i-1}}
\]

(3.2.2)

where \( p_{k}^* = \sum_{l=1}^{m_k} p_{kl} \).

The first product stands for the contribution to the likelihood of the elements of \( A \); the second stands for the contribution of the partial censored observations which are not in \( A_{10} \cup A_{01} \).
and the third stands for the contribution of the double censored observations which are not in \( A_{00} \).

The next theorem gives the \( P_{ij} \) which maximize the above expression. They define the element(s) of \( \mathbf{g}^{++} \) (hence of \( \mathbf{g} \)) that give the GMLE.

**Theorem 3.2.1.** Let \( U_{ij} = n_{i,m_i} + \sum_{k=j}^{m_i} \delta_{ik} \) for \( 1 \leq i \leq m \), \( 1 \leq j \leq m_i \) and \( U_1 = n_m + \sum_{k=1}^{m} U_{kl} \). The \( P_{ij} \) which maximize (3.2.2) are

\[
(3.2.3) \quad \hat{P}_{ij} = \frac{\delta_{ij}}{n} \times \prod_{k=1}^{i} \frac{U_k - n_{k-1}}{U_k - n_k} \times \prod_{l=1}^{j} \frac{U_{i,l} - n_{i,l-1}}{U_{i,l} - n_{i,l}}.
\]

**Remark:**

\( U_{i,n_i-1} = \#(>t_{(i-1)}, t_{(i-1)}) \) where \( t(0) = 0 \),

\( U_{i,n_i} = \#(\leq t_{(i)}, >t_{(i)}) \),

\( U_{ij,n_{i,j-1}} = \#(>t_{((j-1))}, = t_{(i)}) \) or \( \#(=t_{(i)}, >t_{((j-1))}) \). where \( (t(0), t_{(i)}) = (t_{(i)}, t_{(i)}) \) or \( (t_{(i)}, t(0)) = (t_{(i)}, t_{(i)}) \) accordingly.

\( U_{ij,n_{i,j}} = \#(>t_{((j))}, = t_{(i)}) \) or \( \#(=t_{(i)}, >t_{((j))}) \).

Therefore, it is easy to verify that (2.1.1) is the same as (3.2.3).
Proof. From (3.2.2) the likelihood is

\[
L(p) = \prod_{i=1}^{m} \left( \prod_{j=1}^{m_i-1} p_{ij} \left( \sum_{k=j}^{m_i} p_{ik} \right)^{n_{ij} - n_{i,j-1}} \right) \delta_{i,m_i} \times \left( \sum_{k=1}^{m} p_k \right)^{n_i - n_i-1}.
\]

Let

\[
\theta_{ij} = \sum_{\ell=j}^{m_i} p_{i\ell} / \sum_{\ell=j}^{m_i} p_{i\ell} \quad \text{for} \quad 1 \leq i \leq m, \quad 1 \leq j \leq m_i - 1
\]

so

\[
\sum_{\ell=j}^{m_i} p_{i\ell} = p_i \cdot \theta_{i1} \ldots \theta_{i,j-1} = p_i \cdot \prod_{\ell=1}^{j-1} \theta_{i\ell}
\]

and

\[
(3.2.4) \quad p_{ij} = \begin{cases} 
p_i \cdot \left( \prod_{\ell=1}^{j-1} \theta_{i\ell} \right) (1-\theta_{ij}) & \text{if} \quad 1 \leq j \leq m_i - 1 \\
p_i \cdot \prod_{\ell=1}^{m_i-1} \theta_{i\ell} & \text{if} \quad j = m_i
\end{cases}
\]

We then rewrite the likelihood as
\[
\prod_{i=1}^{m} \prod_{j=1}^{m_i - 1} (1 - \theta_{ij}) \delta_{ij} \left( \prod_{k=1}^{j-1} \theta_{i\ell} \right) n_{ij} - n_{i,j-1} + \delta_{ij} \\
\times \left( \prod_{k=1}^{m_i - 1} \theta_{i\ell} \right) n_{i,m_i - n_i,i,m_i - 1} + \delta_{i,m_i} \left( \sum_{k=1}^{m} P_k \right) n_{i,n_i - 1} \\
= \prod_{i=1}^{m} \prod_{j=1}^{m_i - 1} U_{i1} \left( \sum_{k=1}^{m} P_k \right) n_{i,n_i - 1} \left( \prod_{j=1}^{m} (1 - \theta_{ij}) \delta_{ij} \theta_{i,j+1,n_{ij}} \right).
\]

We need another change of variable. Let

\[
\gamma_i = \sum_{k=1}^{m} P_k. \quad \text{for } 1 \leq i \leq m-1.
\]

Then, since \( \sum_{k=1}^{m} P_k = 1 \),

\[
\sum_{k=1}^{i} P_k = \prod_{k=1}^{i-1} \gamma_k \quad \text{for } 1 \leq i \leq m,
\]

and

\[
(3.2.5) \quad P_i = \begin{cases} \\
\prod_{k=1}^{i-1} \gamma_k (1 - \gamma_i) & \text{if } 1 \leq i \leq m-1, \\
\prod_{k=1}^{m-1} \gamma_k & \text{if } i = m.
\end{cases}
\]

29
Therefore, the likelihood becomes

\[
L(S) = \left[ \prod_{i=1}^{m-1} (1-\gamma_i) \prod_{k=1}^{i-1} \gamma_k \right]^{n_i-n_i-1+U_{il}} \times \left( \prod_{j=1}^{m_i-1} (1-\theta_i \delta_{ij}) \delta_{ij} U_{i,j+1-n_{ij}} \right) \times \left( \prod_{j=1}^{m-1} (1-\theta_m \delta_{mj}) \delta_{mj} U_{m,j+1-n_{mj}} \right) \]

\[
= \left[ \prod_{i=1}^{m-1} (1-\gamma_i) \gamma_i \right]^{n_i-n_i+U_{il}} \left( \prod_{j=1}^{m_i-1} (1-\theta_i \delta_{ij}) \delta_{ij} U_{i,j+1-n_{ij}} \right) \times \prod_{j=1}^{m-1} (1-\theta_m \delta_{mj}) \delta_{mj} U_{m,j+1-n_{mj}} .
\]

It is well known from binomial sampling theory that each product is maximized by

\[
\hat{\theta}_{ij} = \frac{(U_{i,j+1-n_{ij}})}{(U_{i,j+1-n_{ij}} + \delta_{ij})} .
\]

\[
= \frac{(U_{i,j+1-n_{ij}})}{(U_{i,j-n_{ij}})} ,
\]

and
\[ \hat{\gamma}_i = \frac{(U_{i+1} - n_i)}{(U_{i+1} - n_i + U_{il})} = \frac{(U_{i+1} - n_i)}{(U_i - n_i)} . \]

Thus, using (3.2.5) and noting that \( U_1 - n_0 = U_1 = n \) and \( U_m - n_m = U_m l \), we have

\[ \hat{p}_{i} = \frac{U_{il}}{n} \times \prod_{k=1}^{i} \frac{U_k - n_{k-1}}{U_k - n_k} \quad \text{for} \quad 1 \leq i \leq m \]

and finally, replacing on (3.2.4) and noting that \( n_{10} = 0 \) and \( U_{i,m_i} - n_{i,m_i} = \delta_{i,m_i} \), we have, for \( 1 \leq i \leq m \) and \( 1 \leq j \leq m_i \),

\[ \hat{p}_{ij} = \frac{\delta_{ij}}{n} \times \prod_{\ell=1}^{j} \frac{U_{i \ell} - n_{i \ell}}{U_{i \ell} - n_{i \ell}} \times \prod_{k=1}^{i} \frac{U_k - n_{k-1}}{U_k - n_k} \]

which is (3.2.3). \( \square \)

Due to the asymmetry of the estimation problem with respect to \( S \) and \( S_0 \), we want to indicate a GME for \( S_0 \).

**Theorem 3.2.2.** A GME for \( S_0 \) is the KME associated with the times \( t_{1i} \sqrt{t_{2i}} \) which will be considered censored if \( d_i = (1,1) \).

**Proof.** The proof follows by noting that the part of the likelihood which involves \( S_0 \) is given by

\[ L_0(S_0) = \prod_{i=1}^{n} \left[ S_0(t_{1i} \sqrt{t_{2i}}) \right]^{1-I(d_i = (1,1))} \left[ S_0(t_{1i} \sqrt{t_{2i}}) \right]^{I(d_i = (1,1))}. \] \( \square \)
3.3 Comparing the Univariate and Bivariate GMLE

In this section we want to point out some of the similarities and the dissimilarities of the KME (which is the univariate GMLE) and the bivariate solution discussed in the previous section.

(i) Roughly speaking, they jump at the uncensored times and at the last censored times, if any. They assign the masses in the same fashion. This is evident by comparing (2.1.1) with the corresponding one for the KME which assigns mass

$$\frac{\#(=t_{(i)})}{n} \times \prod_{c} \frac{\#(>t_{i})}{\#(>t_{j})}$$

at the $i$th uncensored time $t_{(i)}$, where $\prod_{c}$ denotes the product over censored observations.

(ii) The non-uniqueness of the estimators is due to the existence of last censored observations.

(iii) In case of no censoring observations, both reduce to the empirical survival functions.

(iv) The bivariate GMLE can not be achieved by estimating the products of conditional probabilities as in the univariate case where the KME is a product-limit estimator.

(v) In the bivariate case $\hat{S}(t_1, t_2) \hat{S}_0(t_1 \land t_2)$ is not the empirical survival function of the observed vector as in the univariate case. This makes the subsurvival function approach
used by Peterson (1977) not directly extendible to the bivariate case.

(vi) In Chapter 5 we give an example which illustrates how the KME for the marginals are not the marginals of $\hat{S}$. 
4. SELF-CONSISTENCY

The self-consistency property of an estimator as defined by Efron (1967) is considered in this chapter. We will prove that the GMLE is the only self-consistent estimator, where uniqueness is to be understood up to the corresponding uniqueness of the GMLE. In Chapter 2 we discussed an algorithm for computing the GMLE based on its self-consistency property. In a companion paper this property is used again to prove the consistency of the GMLE. (See Muñoz (1980).) The EM algorithm (Dempster, Laird and Rubin (1977)) can be used as an alternative and more general approach to self-consistency. It is used by Hanley and Parnes (1980) to handle the estimation problem when the censoring is of the bivariate type too.

4.1 Definition of Self-consistency

As for the univariate problem (Efron (1967), Peterson (1975)) one wants to estimate $S(t_1,t_2)$ as the proportion of failure time vectors on $(t_1,\infty) \times (t_2,\infty)$. For the uncensored (complete) data problem this is just the empirical bivariate survival function. For the censored data problem we are lead to the bivariate self-consistent estimator. Given the censoring information, one must estimate the conditional probability that the failure vector is in the set whose probability one is estimating. We have then an estimation problem within an estimation problem. An estimator is self-consistent if, when used for both, there is no change.

Definition 4.1.1. A survival function $\hat{S}(t_1,t_2)$ is a self-consistent estimator of $S(t_1,t_2)$ if
\[ n \hat{S}(t_1, t_2) = \#(t_1 > t_2, t_2 > t_1) + \sum_{d_1=(0,1), \ t_2 < t_{21} < t_{11} < t_1} \frac{\hat{S}(t_1, t_{21}-) - \hat{S}(t_1, t_{21})}{\hat{S}(t_{11}, t_{21}-) - \hat{S}(t_{11}, t_{21})} \]

\[ + \sum_{d_1=(1,0), \ t_1 < t_{11} < t_{21} < t_2} \frac{\hat{S}(t_{11}, t_2) - \hat{S}(t_{11}, t_2)}{\hat{S}(t_{11}, t_{21}) - \hat{S}(t_{11}, t_{21})} \]

\[ + \sum_{d_1=(0,0), \ t_{11} = t_{21} < t_1 < t_2} \frac{\hat{S}(t_1, t_{21}, t_2, v_{21})}{\hat{S}(t_{11}, t_{21})} . \]

The second summand in (4.1.1) estimates \( P(T_{11} > t_1, T_{21} > t_2 | T_{11} > t_{11}, T_{21} = t_{21}) \); the third estimates \( P(T_{11} > t_1, T_{21} > t_2 | T_{11} = t_{11}, T_{21} > t_{21}) \) and the last one estimates \( P(T_{11} > t_1, T_{21} > t_2 | T_{11} > t_{11}, T_{21} > t_{21}) \). Needless to say, an estimator to be self-consistent has to assign positive probability to lines associated with the one component censored observations and to regions associated with the two component censored observations.

Another way to look at self-consistency is as follows: given an estimator \( \hat{S}(0) \) one can use (4.1.1) to achieve a new estimator \( \hat{S}(1) \) and continue this process until \( \hat{S}(k) = \hat{S}(k-1) \). The result is a self-consistent estimator.

It should be noted that the second [third] sum is empty if \( t_2 > t_1 [t_2 < t_1] \). Both will be empty if \( t_1 = t_2 \). Due to this symmetry, we will (wlog) take \( t_1 > t_2 \) for the discussion below.
Finally, the condition that \( \hat{S} \) be a survival function is needed to avoid odd behavior of \( \hat{S} \) after the "last" censored observations in any direction (see Peterson (1975)).

4.2 Self-consistency and Generalized Maximum Likelihood Estimation

In this section we prove the main theorem of the chapter. Previously, we stated and used the following lemma, which we now prove.

**Lemma 4.2.1.** If \( \hat{S} \) is a survival function, \( \hat{S} \) is self-consistent if and only if (wlog) for \( t_1 > t_2 \) we have

\[
\frac{\hat{S}(t_1, t_2^-) - \hat{S}(t_1, t_2)}{\hat{S}(t_1', t_2^-) - \hat{S}(t_1', t_2)} = \frac{1}{\#(t_1', t_2')},
\]

for \( t_1' \leq t_1 \)

\[
\times \left[ \frac{\hat{S}(t_1, t_2^-) - \hat{S}(t_1, t_2)}{\hat{S}(t_1', t_2^-) - \hat{S}(t_1', t_2)} \right] \sum_{d_1 = (0,1)}^{t_1 < t_1 < t_1 < t_1} \frac{\hat{S}(t_1, t_2^-) - \hat{S}(t_1, t_2)}{\hat{S}(t_1', t_2^-) - \hat{S}(t_1', t_2)}
\]

and

\[
\frac{\hat{S}(t_1, t_2^\infty t')}{\hat{S}(t', t')} = \frac{1}{\#(t', t')} \left[ \#(t_1, t_2^\infty t') \right]
\]

\[
+ \sum_{d_1 = (0,1)}^{t_2^\infty t' < t_2^\infty t_1 < t_1} \frac{\hat{S}(t_1, t_2^\infty t_1) - \hat{S}(t_1, t_2)}{\hat{S}(t_1', t_2^\infty t_1) - \hat{S}(t_1', t_2)} \sum_{d_1 = (0,0)}^{t' < t_1^\infty t_2 < t_1} \frac{\hat{S}(t_1, t_2^\infty t_2)}{\hat{S}(t_1', t_2^\infty t_2)}
\]

for \( t' \leq t_1 \) .
Proof. (=). Assume \( \hat{S} \) is self-consistent. Let \( t_1 > t_2 \).

Using (4.1.1) for \((t,t)\) one gets

\[
(4.2.3) \quad \frac{\hat{S}(t,t)}{\#(>t,>t)} = \left[ n - \sum_{d_1=(0,0)}^{t_1=t_{21}<t} \frac{1}{\hat{S}(t_{11},t_{21})} \right]^{-1},
\]

also, using (4.1.1) for \((t_1,t_2^-)\) and \((t_1,t_2)\) and subtracting the latter from the former one gets

\[
n[\hat{S}(t_1,t_2^-) - \hat{S}(t_1,t_2)] = \#(>t_1,=t_2) + \sum_{d_1=(0,1)}^{t_2=t_{21}<t_1<1} \frac{\hat{S}(t_1,t_2^-) - \hat{S}(t_1,t_2)}{\hat{S}(t_{11},t_{21}) - \hat{S}(t_{11},t_2)} + \sum_{d_1=(0,0)}^{t_1=t_{21}<t_2} \frac{\hat{S}(t_1,t_2^-) - \hat{S}(t_1,t_2)}{\hat{S}(t_{11},t_{21})},
\]

then by using (4.2.3) for \( t_2^- \)

\[
\frac{\hat{S}(t_1,t_2^-) - \hat{S}(t_1,t_2)}{\#(>t_1,=t_2)}
\]

(4.2.4)

\[
= \left[ \frac{\#(\geq t_1, t_2^-)}{\hat{S}(t_2-, t_2^-)} - \sum_{d_1=(0,1)}^{t_2=t_{21}<t_1<1} \frac{1}{\hat{S}(t_{14},t_{22}) - \hat{S}(t_{14},t_2)} \right]^{-1}.
\]

Moreover, from (4.1.1) one can write
\[ \hat{S}(t_1, t_2) \left[ n - \sum_{d_i=(0,0), t_{1i}=t_{2i}<t_1} \frac{1}{\hat{S}(t_{1i}, t_{2i})} \right] = \#(>t_1, >t_2) \]

\[ + \sum_{d_i=(0,1), t_2<t_{2i}<t_{1i}<t_1} \frac{\hat{S}(t_{1i}, t_{2i}) - \hat{S}(t_{1i}', t_{2i})}{\hat{S}(t_{1i}, t_{2i}) - \hat{S}(t_{1i}', t_{2i})} + \sum_{d_i=(0,0), t_2<t_{1i}=t_{2i}<t_1} \frac{\hat{S}(t_1, t_{2i})}{\hat{S}(t_{1i}, t_{2i})}, \]

then by (4.2.3)

\[(4.2.5) \quad \hat{S}(t_1, t_2) = \frac{\hat{S}(t_1, t_2)}{\#(>t_1, >t_2)} \quad [\text{R.H.S. of previous identity}]. \]

Let \( t_1' \leq t_1 \). From (4.2.4) we have

\[ \frac{\hat{S}(t_1, t_2) - \hat{S}(t_1', t_2)}{\#(>t_1', >t_2)} \cdot \frac{\#(>t_1', =t_2)}{\hat{S}(t_1', t_2) - \hat{S}(t_1', t_2)} \]

\[ = \frac{\#(>t_2', >t_2)}{\hat{S}(t_2', t_2)} - \sum_{d_i=(0,1), t_2=t_{2i}<t_{1i}<t_1} \frac{1}{\hat{S}(t_{1i}, t_{2i}) - \hat{S}(t_{1i}', t_2)} \]

\[ - \frac{\#(>t_2', >t_2)}{\hat{S}(t_2', t_2)} - \sum_{d_i=(0,1), t_2=t_{2i}<t_{1i}<t_1} \frac{1}{\hat{S}(t_{1i}, t_{2i}) - \hat{S}(t_{1i}', t_2)}. \]
\[
1 + \frac{\sum_{d_1 = (0,1)} \frac{1}{\hat{S}(t_{11}, t_{21}^-) - \hat{S}(t_{11}, t_2)}}{\frac{\#(\rightarrow t_2, \rightarrow t_2)}{\hat{S}(t_{21}^-, t_2^-)} - \sum_{d_1 = (0,1)} \frac{1}{\hat{S}(t_{11}, t_2^-) - \hat{S}(t_{11}, t_2)}}}_{t_2 = t_{21} < t_{11} < t_1}
\]

\[
= 1 + \frac{\hat{S}(t_{11}, t_{21}^-) - \hat{S}(t_{11}, t_2)}{\#(>t_{11}, >t_2)} \times \sum_{d_1 = (0,1)} \frac{1}{\hat{S}(t_{11}, t_2^-) - \hat{S}(t_{11}, t_2)}
\]

\[
\times \sum_{d_1 = (0,1)} \frac{\hat{S}(t_{21}, t_{21}^-) - \hat{S}(t_{21}, t_2)}{\hat{S}(t_{11}, t_{21}^-) - \hat{S}(t_{11}, t_2)}
\]

\[
\times \left( \frac{\#(t_{21}, t_{21}^-)}{\hat{S}(t_{11}, t_{21}^-) - \hat{S}(t_{11}, t_2)} \right)\]

and (4.2.1) follows.

To show that self-consistency implies (4.2.2), let \( t' < t_1 \).

Using (4.2.3) in an analogous manner one obtains:

\[
\frac{\hat{S}(t_{21}, t_{21}^-)}{\hat{S}(t', t_2)} = \frac{1}{\#(>t', >t')}
\]

\[
\times \left( \frac{\#(t_{21}, t_{21}^-)}{\hat{S}(t_{11}, t_{21}^-) - \hat{S}(t_{11}, t_2)} \right)\]

\[
\times \sum_{d_1 = (0,0)} \frac{\hat{S}(t_{21}, t_{21}^-) - \hat{S}(t_{21}, t_2)}{\hat{S}(t_{11}, t_{21}^-) - \hat{S}(t_{11}, t_2)}
\]

thus,

\[
(4.2.6) \quad \frac{\hat{S}(t_{21}, t_{21}^-)}{\#(>t_{21}, >t_{21})} = \left( \frac{\#(t', t_2)}{\hat{S}(t', t_2)} - \sum_{d_1 = (0,0)} \frac{1}{\hat{S}(t_{11}, t_{21})} \right)\]

\[
\times \left( \frac{\#(>t', >t')}{\hat{S}(t', t_2)} \right) - \sum_{d_1 = (0,0)} \frac{1}{\hat{S}(t_{11}, t_{21})}
\]

\[
\times \left( \frac{\hat{S}(t_{21}, t_{21}^-)}{\hat{S}(t_{11}, t_{21}^-) - \hat{S}(t_{11}, t_2)} \right)\]

\[
\times \sum_{d_1 = (0,0)} \frac{\hat{S}(t_{21}, t_{21}^-) - \hat{S}(t_{21}, t_2)}{\hat{S}(t_{11}, t_{21}^-) - \hat{S}(t_{11}, t_2)}
\]

\[
\times \left( \frac{\#(t_{21}, t_{21}^-)}{\hat{S}(t_{11}, t_{21}^-) - \hat{S}(t_{11}, t_2)} \right)\]

\[
- \sum_{d_1 = (0,0)} \frac{1}{\hat{S}(t_{11}, t_{21})}
\]
Now, using (4.2.5) for the point \((t_1, t_2 v t')\) one obtains

\[
\hat{S}(t_1, t_2 v t') \cdot \frac{\#(> t_2 v t', > t_2 v t')}{\hat{S}(t_2 v t', t_2 v t')} = \#(> t_1, > t_2 v t')
\]

\[
+ \sum_{d_i=(0,1)} \frac{\hat{S}(t_1, t_2 v t') - \hat{S}(t_1, t_2 v t'}{\hat{S}(t_1, t_2 v t') - \hat{S}(t_1, t_2 v t')}
\]

\[
t_2 v t' < t_2 v t' < t_1 v t' < t_2 v t'
\]

\[
+ \sum_{d_i=(0,0)} \frac{\hat{S}(t_1, t_2 v t')}{\hat{S}(t_1, t_2 v t')}
\]

\[
t_2 v t' < t_2 v t' < t_1 v t' < t_2 v t'
\]

and then by (4.2.6) one has:

\[
\hat{S}(t_1, t_2 v t') \frac{\#(> t_1', > t')}{\hat{S}(t_1', t')} = \#(> t_1, > t_2 v t')
\]

\[
+ \sum_{d_i=(0,1)} \frac{\hat{S}(t_1, t_2 v t') - \hat{S}(t_1, t_2 v t')}{\hat{S}(t_1, t_2 v t') - \hat{S}(t_1, t_2 v t')}
\]

\[
t_2 v t' < t_2 v t' < t_1 v t' < t_2 v t'
\]

\[
+ \sum_{d_i=(0,0)} \frac{\hat{S}(t_1, t_2 v t')}{\hat{S}(t_1, t_2 v t')} + \sum_{d_i=(0,0)} \frac{\hat{S}(t_1, t_2 v t')}{\hat{S}(t_1, t_2 v t')}
\]

\[
t_2 v t' < t_2 v t' < t_1 v t' < t_2 v t'
\]

Since the last two terms on the R.H.S. expression equal

\[
\sum_{d_i=(0,0)} \frac{\hat{S}(t_1, t_2 v t')}{\hat{S}(t_1, t_2 v t')}
\]

\[
t_1 < t_1 v t' < t_2 v t'
\]

\[
t_1 < t_1 v t' < t_2 v t'
\]

40
(4.2.2) follows, and thus the first half of the proof.

(⇐) The proof is completed by taking $t' = 0$ on (4.2.2).

The next theorem states the relation between a self-consistent estimator and a GMLE.

**Theorem 4.2.1.** A survival function $\hat{S}$ is a self-consistent estimator if and only if $\hat{S}$ is a GMLE.

**Proof.** (⇐) Assume $\hat{S}$ is self-consistent. Using (4.2.2) for $t_1 = t_2 (=t, \text{say})$ one gets

\[
(4.2.7) \quad \frac{\hat{S}(t, t)}{\hat{S}(t', t')} = \frac{1}{\#(t', t')} \left[ \#(t', t') + \sum\limits_{\substack{d_j=(0, 0) \\text{such that} \ t'<t_1<e_2\leq t}} \frac{\hat{S}(t, t)}{\hat{S}(t_1, t_2)} \right],
\]

for $t' \leq t$.

If $t_k$ is a double censored observation (i.e., $t_{1k} = t_{2k}$ and $d_k = (0, 0)$), then, by (4.2.7)

\[
\frac{\hat{S}(t_{1k}, t_{1k})}{\hat{S}(t_{1k}^-, t_{1k}^-)} = \frac{1}{\#(t_{1k}^-, t_{1k}^-)} \left[ \#(t_{1k}^-, t_{1k}^-) + \#(t_{1k}^-, t_{1k}^-) \frac{\hat{S}(t_{1k}, t_{1k})}{\hat{S}(t_{1k}, t_{1k})} \right],
\]

and we have

\[
(4.2.8) \quad \hat{S}(t_{1k}, t_{1k}) = \hat{S}(t_{1k}^-, t_{1k}^-).
\]
On the other hand, if \( (t_{1k}^{\infty})(t_{1k}, t_{1k}^{\infty})(t_{1k}^{\infty}, t_{1k}) \) contains non-double censored observations, then by (4.2.7):

\[
(4.2.9) \quad \frac{\hat{S}(t_{1k}', t_{1k})}{\hat{S}(t_{1k}^-, t_{1k}^-)} = \frac{\#(t_{1k}', t_{1k})}{\#(t_{1k}', t_{1k})} .
\]

Therefore, a self-consistent estimator behaves as a GMLE with respect to the equiangular line. We proceed to prove an analogous result for horizontal (and vertical) lines.

If \( t_k \) with \( t_{1k} > t_{2k} \) is a \((0,1)\) censored observation, then by (4.2.1),

\[
\frac{\hat{S}(t_{1k}', t_{2k}^-) - \hat{S}(t_{1k}, t_{2k})}{\hat{S}(t_{1k}^-, t_{2k}^-) - \hat{S}(t_{1k}^-, t_{2k})} = \frac{1}{\#(t_{1k}^{\infty}, t_{2k})} \left[ \frac{\#(t_{1k}^-, t_{2k})}{\#(t_{1k}^-, t_{2k})} \right.
\]

\[
+ \frac{\hat{S}(t_{1k}', t_{2k}^-) - \hat{S}(t_{1k}, t_{2k})}{\hat{S}(t_{1k}', t_{2k}^-) - \hat{S}(t_{1k}, t_{2k})} \right] 
\]

so

\[
(4.2.10) \quad \hat{S}(t_{1k}', t_{2k}^-) - \hat{S}(t_{1k}, t_{2k}) = \hat{S}(t_{1k}^-, t_{2k}^-) - \hat{S}(t_{1k}^-, t_{2k}).
\]

On the other hand, if \( t_k \), with \( t_{1k} > t_{2k} \), is a \((1,1)\) observation, then by (4.2.1):

\[
(4.2.11) \quad \frac{\hat{S}(t_{1k}', t_{2k}^-) - \hat{S}(t_{1k}, t_{2k})}{\hat{S}(t_{1k}^-, t_{2k}^-) - \hat{S}(t_{1k}^-, t_{2k})} = \frac{\#(t_{1k}', t_{2k})}{\#(t_{1k}', t_{2k})}. 
\]
Thus, by (4.2.8) to (4.2.11), a self-consistent estimator spreads the mass as a GMLE does. Hence, it is a GMLE.

(⇒) Assume \( \hat{S} \) is a GMLE. Then \( \hat{S} \) satisfies (4.2.8) to (4.2.11) and so \( \hat{S} \) is self-consistent for \( \hat{S}(t,t) \) and \( \hat{S}(t_1,t_2) - \hat{S}(t_1,t_2) \) for \( t_1 > t_2 \). It is

\[
(4.2.12) \quad n \hat{S}(t,t) = \#(t,t) + \hat{S}(t,t) \sum_{d_1=(0,0)}^{1} \frac{1}{\hat{S}(t_{1i},t_{2i})},
\]

and

\[
(4.2.13) \quad n[\hat{S}(t_1,t_2) - \hat{S}(t_1,t_2)] = \#(t_1,t_2) + [\hat{S}(t_1,t_2) - \hat{S}(t_1,t_2)]
\]

\[
\times \left\{ \sum_{d_1=(0,1)}^{1} \frac{1}{\hat{S}(t_{1i},t_{2i})} - \hat{S}(t_{1i},t_{2i}) + \sum_{d_1=(0,0)}^{1} \frac{1}{\hat{S}(t_{1i},t_{2i})} \right\}.
\]

Let \( t_2 < y(1) < \ldots < y(k) \leq t_1 \) such that \( (t_1,y(1))^{(t_1,\infty)} \) contains observations. Then, by the way \( \hat{S} \) spreads its mass, one can write

\[
n \hat{S}(t_1,t_2) = n \hat{S}(t_1,y(1))^{-}
\]

\[
= \sum_{j=1}^{k} n[\hat{S}(t_1,y(j))^{-} - \hat{S}(t_1,y(j))] + n \hat{S}(t_1,t_1) .
\]
Using (4.2.12) and (4.2.13) one gets

\[ n \hat{S}(t_1, t_2) = \sum_{j=1}^{k} \#(t_1, y(j)) \]

\[ + \sum_{j=1}^{k} \sum_{d_i=(0,1)}^{t_2 \leq t_1} \frac{\hat{S}(t_1, y(j)) - \hat{S}(t_1, y(j))}{\hat{S}(t_1, t_2)} \]

\[ + \sum_{j=1}^{k} \sum_{d_i=(0,0)}^{t_1 = t_2, y(j)} \frac{\hat{S}(t_1, y(j)) - \hat{S}(t_1, y(j))}{\hat{S}(t_1, t_2)} \]

\[ + \#(t_1, t_2) + \sum_{d_i=(0,0)}^{t_1 = t_2, y(j)} \frac{\hat{S}(t_1, t_2)}{\hat{S}(t_1, t_2)} \]

By interchanging the sums

\[ n \hat{S}(t_1, t_2) = \#(t_1, t_2) + \sum_{d_i=(0,1)}^{t_2 < t_1 < t_1} \frac{\hat{S}(t_1, t_2) - \hat{S}(t_1, t_2)}{\hat{S}(t_1, t_2)} \]

\[ + \sum_{d_i=(0,0)}^{t_1 = t_2, y(j)} \frac{1}{\hat{S}(t_1, t_2)} \left( \sum_{j} \left[ \frac{\hat{S}(t_1, y(j)) - \hat{S}(t_1, y(j))}{\hat{S}(t_1, y(j))} \right] \right) \]

44
\[ + \sum_{d_1=(0,0)} \frac{\hat{S}(t_1,t_1)}{\hat{S}(t_{11},t_{21})} \]

\[= \#(t_1 > t_2) + \sum_{d_1=(0,1)} \frac{\hat{S}(t_1,t_{21}) - \hat{S}(t_1,t_{11})}{\hat{S}(t_{11},t_{21}) - \hat{S}(t_{11},t_{21})} \]

\[+ \sum_{d_1=(0,0)} \frac{1}{\hat{S}(t_{11},t_{21})} (\hat{S}(t_1,t_2) - \hat{S}(t_1,t_1)) \]

\[+ \sum_{d_1=(0,0)} \frac{1}{\hat{S}(t_{11},t_{21})} (\hat{S}(t_1,t_{11}) - \hat{S}(t_1,t_1)) \]

\[+ \sum_{d_1=(0,0)} \frac{\hat{S}(t_1,t_1)}{\hat{S}(t_{11},t_{21})} \]

\[= \#(t_1 > t_2) + \sum_{d_1=(0,1)} \frac{\hat{S}(t_1,t_{21}) - \hat{S}(t_1,t_{11})}{\hat{S}(t_{11},t_{21}) - \hat{S}(t_{11},t_{21})} \]

45
\[ + \sum_{d_1=(0,0)} \frac{\hat{S}(t_1, t_2 v t_{24})}{\hat{S}(t_{11}, t_{21})}, \]
\[ t_{11} = t_{24} < t_1 \]

which is (4.1.1) for \( t_1 > t_2 \). Hence \( \hat{S} \) is self-consistent. \( \square \)

**Remark.** The point \((t_1, t_2)\) in the second part of the proof above is meant to be a point where the GMLE \( \hat{S} \) is not ambiguous due to possible mass on lines and/or region. Alternatively, following Efron (1967), one may treat the "last" censored observations as double uncensored.

As a final comment, due to the fact that \( \hat{S}(t_1, t_2) \hat{S}_0(t_1 v t_2) \) is not necessarily equal to \( \#(>t_1, >t_2)/n \), one says that there is no mutual consistency of the self-consistent estimators (see Peterson (1975)).
5. EXAMPLE

We illustrate our estimator by using data on times of recurrence and death for patients with stage III malignant melanoma collected by Lynn E. Spitler. We point out how the Kaplan–Meier estimator for the survival time is not the marginal of our estimator. In addition, a way to allocate line masses to points is indicated so as to achieve a GMLE precisely defined at most points. In particular, the marginals of this estimator will have the same ambiguity (if any) as the Kaplan–Meier estimators. Finally, we illustrate the discrepancy between \( \hat{S}(t_1,t_2) \hat{S}_0(t_1 \vee t_2) \) and the empirical survival function of the data.

5.1 The Data

After treatment (surgery), twenty-five patients with stage III malignant melanoma were free of any clinical evidence of disease (said to be "in remission"). They were followed up with regard to the course of their disease, and the recorded dates for each patient were: the date of surgery (i.e., date patient went into remission), the date of recurrence, the date of death and, in the absence of any of the former two, the date of last follow up.

With the first component for recurrence and the second for death, the data, in days, are:
<table>
<thead>
<tr>
<th>i</th>
<th>( t_{i1} = (t_{i1}, t_{21}) )</th>
<th>( d_{i1} = (d_{i1}, d_{21}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( (220, 253) )</td>
<td>( (1, 1) )</td>
</tr>
<tr>
<td>2</td>
<td>( (262, 1225) )</td>
<td>( (1, 1) )</td>
</tr>
<tr>
<td>3</td>
<td>( (1569, 1569) )</td>
<td>( (0, 0) )</td>
</tr>
<tr>
<td>4</td>
<td>( (2233, 2233) )</td>
<td>( (0, 0) )</td>
</tr>
<tr>
<td>5</td>
<td>( (1305, 1305) )</td>
<td>( (0, 0) )</td>
</tr>
<tr>
<td>6</td>
<td>( (348, 658) )</td>
<td>( (1, 1) )</td>
</tr>
<tr>
<td>7</td>
<td>( (816, 894) )</td>
<td>( (1, 0) )</td>
</tr>
<tr>
<td>8</td>
<td>( (249, 380) )</td>
<td>( (1, 1) )</td>
</tr>
<tr>
<td>9</td>
<td>( (182, 189) )</td>
<td>( (1, 1) )</td>
</tr>
<tr>
<td>10</td>
<td>( (313, 1088) )</td>
<td>( (1, 0) )</td>
</tr>
<tr>
<td>11</td>
<td>( (326, 719) )</td>
<td>( (1, 1) )</td>
</tr>
<tr>
<td>12</td>
<td>( (100, 1326) )</td>
<td>( (1, 1) )</td>
</tr>
<tr>
<td>13</td>
<td>( (1336, 1336) )</td>
<td>( (0, 0) )</td>
</tr>
<tr>
<td>14</td>
<td>( (473, 568) )</td>
<td>( (1, 1) )</td>
</tr>
<tr>
<td>15</td>
<td>( (32, 66) )</td>
<td>( (1, 1) )</td>
</tr>
<tr>
<td>16</td>
<td>( (602, 602) )</td>
<td>( (0, 0) )</td>
</tr>
<tr>
<td>17</td>
<td>( (180, 782) )</td>
<td>( (1, 0) )</td>
</tr>
<tr>
<td>18</td>
<td>( (121, 773) )</td>
<td>( (1, 0) )</td>
</tr>
<tr>
<td>19</td>
<td>( (333, 453) )</td>
<td>( (1, 0) )</td>
</tr>
<tr>
<td>20</td>
<td>( (140, 144) )</td>
<td>( (1, 1) )</td>
</tr>
<tr>
<td>21</td>
<td>( (424, 743) )</td>
<td>( (1, 0) )</td>
</tr>
<tr>
<td>22</td>
<td>( (303, 303) )</td>
<td>( (0, 0) )</td>
</tr>
<tr>
<td>23</td>
<td>( (88, 249) )</td>
<td>( (1, 1) )</td>
</tr>
<tr>
<td>24</td>
<td>( (88, 207) )</td>
<td>( (1, 1) )</td>
</tr>
<tr>
<td>25</td>
<td>( (116, 116) )</td>
<td>( (0, 0) )</td>
</tr>
</tbody>
</table>
One should point out that this set of data is a particular case of our problem in the sense that the first component cannot be larger than the second; therefore, no $d = (0,1)$ observations are possible.

The data are plotted in Figure 1. Arrows are attached to the censored points, and are meant to suggest the sets where the times of recurrence and death could happen.

5.2 The Estimator $\hat{S}$

The data points can be ordered according to the system of Section 3.2. Figure 2 shows the points, lines and region where $\hat{S}$ assigns positive probabilities. By (3.2.3) (or equivalently (2.1.1)), the values of the probabilities $\hat{S}$ attaches to these points, lines and region are as follows

\[
\frac{1}{25} \quad \text{at} \ (32,66); \ (88,207); \ (88,249); \ (100,1326);
\]

\[
\frac{1}{25} \times \frac{21}{20} = \frac{1.05}{25} \quad \text{at} \ (140,144); \ (182,189); \ (220,353); \ (249,380);
\]

\[
\text{(262,1225) and at the lines} \ (121,773)(121,\infty);
\]

\[
(180,782)(180,\infty);
\]

\[
\frac{1}{25} \times \frac{21}{20} \times \frac{13}{12} = \frac{1.1375}{25} \quad \text{at} \ (326,719); \ (348,658); \ (473,568) \text{ and at the lines} \ (313,1088)(313,\infty);
\]

\[
(333,453)(333,\infty);
\]

\[
(424,743)(424,\infty);
\]
\[
\frac{1}{25} \times \frac{21}{20} \times \frac{13}{12} \times \frac{6}{5} = \frac{1.365}{25}
\quad \text{at the line } (816, 894)(816, \infty);
\]

\[
\frac{1}{25} \times \frac{21}{20} \times \frac{13}{12} \times \frac{6}{5} \times \frac{4}{1} = \frac{5.46}{25}
\quad \text{at the region } \{t : 2233 < t_1 < t_2\}.
\]

5.3 The Kaplan–Meier Estimator for the Survival vs. the Marginal Estimator from \(\hat{S}\)

Let us first compute the marginal \(\hat{S}_2\) of \(\hat{S}\). Projecting \(\hat{S}\) on the \(T_2\)-axis, one gets masses as follows

\[
\frac{1}{25}
\quad \text{at 66; 207; 249; 1326;}
\]

\[
\frac{1.05}{25}
\quad \text{at 144; 189; 353; 380; 1225;}
\quad \text{and at lines from 773 and 782;}
\]

\[
\frac{1.1375}{25}
\quad \text{at 568; 658; 719;}
\quad \text{and at lines from 453; 743; 1088;}
\]

\[
\frac{1.365}{25}
\quad \text{at the line from 894;}
\]

\[
\frac{5.46}{25}
\quad \text{at the line from 2233}.
\]

If one interprets the above as a censored random sample with the "number of observations" at a particular time equal to the
numerator of the above ratios, then one can redistribute the masses on lines (i.e., from "censored observations") in the Kaplan-Meier way. \( \hat{S}_2 \), then, puts masses as follows

\[
\begin{align*}
\frac{1}{25} & \quad \text{at 66; 207; 249;} \\
\frac{1.05}{25} & \quad \text{at 144; 189; 353; 380;} \\
\frac{1.2151}{25} & \quad \text{at 568; 658; 719;} \\
\frac{1.979}{25} & \quad \text{at 1225;} \\
\frac{1.885}{25} & \quad \text{at 1326;} \\
\frac{10.291}{25} & \quad \text{at the line from 2233.}
\end{align*}
\]

On the other hand, the Kaplan-Meier estimator \( \hat{S}_2^{KM} \) for \( S_2 \) is found by first projecting the data on the \( T_2 \)-axis (i.e., ignoring the information on \( T_1 \)) and then finding the Kaplan-Meier estimator of the projected data.

The estimators are
<table>
<thead>
<tr>
<th>t</th>
<th>(\hat{S}_{KM}^2(t))</th>
<th>(\hat{S}_2(t))</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,66)</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>[66,144)</td>
<td>.9600</td>
<td>.9600</td>
<td>0</td>
</tr>
<tr>
<td>[144,189)</td>
<td>.9183</td>
<td>.9180</td>
<td>-.0003</td>
</tr>
<tr>
<td>[189,207)</td>
<td>.8765</td>
<td>.8760</td>
<td>-.0005</td>
</tr>
<tr>
<td>[207,249)</td>
<td>.8348</td>
<td>.8360</td>
<td>.0012</td>
</tr>
<tr>
<td>[249,353)</td>
<td>.7930</td>
<td>.7960</td>
<td>.0030</td>
</tr>
<tr>
<td>[353,380)</td>
<td>.7490</td>
<td>.7540</td>
<td>.0050</td>
</tr>
<tr>
<td>[380,568)</td>
<td>.7049</td>
<td>.7120</td>
<td>.0071</td>
</tr>
<tr>
<td>[568,658)</td>
<td>.6579</td>
<td>.6634</td>
<td>.0055</td>
</tr>
<tr>
<td>[658,719)</td>
<td>.6073</td>
<td>.6148</td>
<td>.0075</td>
</tr>
<tr>
<td>[719,1225)</td>
<td>.5567</td>
<td>.5662</td>
<td>.0095</td>
</tr>
<tr>
<td>[1225,1326)</td>
<td>.4639</td>
<td>.4870</td>
<td>.0231</td>
</tr>
<tr>
<td>[1326,2233]</td>
<td>.3479</td>
<td>.4116</td>
<td>.0637</td>
</tr>
<tr>
<td>(2233,\infty)</td>
<td>undefined</td>
<td>undefined</td>
<td></td>
</tr>
</tbody>
</table>

It should be pointed out that the difference is not due to the way the projected line masses of \(\hat{S}\) were redistributed for the computation of \(\hat{S}_2\). The value of the estimator at, say, 380 is unchanged by such redistribution and it is different from \(\hat{S}_2^{KM}(380)\).

Thus, we see that the generalized maximum likelihood estimator of the survival time distribution does change, when recurrence times are taken into account.
Finally, in contrast with the survival marginal, the recurrence marginal is estimated equally by both methods. This is because \( T_1 = T_1 \wedge T_2 \) and, as discussed in Chapter 2, all of the methods - Kaplan-Meier, competing risks theory, and our own will lead to the same estimator for \( P\{T_1 \wedge T_2 > t\} \).

5.4 Redistributing the Line Masses of \( \hat{S} \)

From Figure 2 one sees that the estimator \( \hat{S} \) is undefined at many places. This section indicates a way to assign the masses on lines to points, for the most part, while preserving the maximum likelihood properties of the bivariate estimator and also preserving the marginal distributions as calculated in the previous section.

The idea is based on a generalization of Efron's redistribute-to-the-right algorithm. Both the KME and our estimator can be achieved by redistributing the mass attached to a censored observation according (i.e., proportionally) to the masses on the set associated with the censored observation (i.e., the set where the failure time(s) could occur). One starts by attaching mass \( n^{-1} \) to all observations.

To use the above idea for redistributing the line masses of \( \hat{S} \) one proceeds as follows: Select the line with largest second component (i.e., (313,1088)(313,\( \varnothing \))) and redistribute the mass in that line proportionally to the masses of the \( T_2 \)-marginal after
such second component. Repeat the procedure until you get to the line with the least second component. For problems with horizontal lines the procedure is analogous.

As an illustration, the mass $1.1325/25$ at the line $(313,1088)(313,\infty)$ should be redistributed to the points $(313,1225); (313,1326)$ and the line $(313,2233)(313,\infty)$ according to the masses $1.05/25; 1/25$ and $5.46/25$ attached to $1225,1326$ and the line from $2233$ by the $T_{2}\text{-marginal}$. Therefore, one should assign mass

$$\frac{1.1325}{25} \times \left(\frac{1.05}{1.05+1+5.46}\right) \text{ to } (313,1225),$$

$$\frac{1.1325}{25} \times \left(\frac{1}{1.05+1+5.46}\right) \text{ to } (313,1326),$$

$$\frac{1.1325}{25} \times \left(\frac{5.46}{1.05+1+5.46}\right) \text{ to } (313,2233)(313,\infty).$$

After all the masses on lines are redistributed as above, the achieved GMLE puts masses as follows

$$\frac{1}{25} \text{ at } (32,66); (88,207); (88,249); (100,1326);$$

$$\frac{1.1468}{25} \text{ at } (121,1225); (180,1225);$$

$$\frac{1.1398}{25} \text{ at } (121,1326); (180,1326);$$
\[ \frac{.7634}{25} \text{ at } (121,2233)(121,\infty); (180,2233)(180,\infty); \]

\[ \frac{1.05}{25} \text{ at } (140,144); (182,189); (220,253); (249,280); (262,1225); \]

\[ \frac{.1590}{25} \text{ at } (313,1225); (424,1225); \]

\[ \frac{.1515}{25} \text{ at } (313,1326); (424,1326); \]

\[ \frac{.827}{25} \text{ at } (313,2233)(313,\infty); (424,2233)(424,\infty); \]

\[ \frac{1.1375}{25} \text{ at } (326,719); (348,658); (473,568); \]

\[ \frac{.0776}{25} \text{ at } (333,568); (333,658); (333,719); \]

\[ \frac{.1264}{25} \text{ at } (333,1225); \]

\[ \frac{.1204}{25} \text{ at } (333,1326); \]

\[ \frac{.6576}{25} \text{ at } (333,2233)(333,\infty); \]

\[ \frac{.1908}{25} \text{ at } (816,1225); \]

\[ \frac{.1818}{25} \text{ at } (816,1326); \]
\[
\frac{.9924}{25} \text{ at } (816, 2233)(816, \infty); \\
\frac{5.46}{25} \text{ at } \{t : 2233 < t_1 < t_2 \}. 
\]

Figure 3 shows the places where the estimator above assigns positive probabilities.

It is easy to verify that the marginals of this estimator are those described in Section 5.3.

5.5 Mutual Inconsistency of \( \hat{S} \) and \( \hat{S}_0 \)

The univariate censored data \((\ell_i; \epsilon_i)\) where \(\ell_i = t_{i1} \vee t_{2i}\) and \(1 - \epsilon_i = I(d_i = (1,1))\), are the basis for the computation of the GMLE for the censoring distribution \( \hat{S}_0 \). For our example we have

\[
\begin{array}{cc}
t & \hat{S}_0(t) \\
[0, 116) & 1 \\
[116, 303) & .9583 \\
[303, 453) & .9079 \\
[453, 602) & .8511 \\
[602, 743) & .7903 \\
[743, 773) & .7185 \\
[773, 782) & .6466 \\
[782, 894) & .5748 \\
[894, 1088) & .5029 \\
[1088, 1305) & .4311 \\
[1305, 1356) & .3449 \\
[1356, 1569) & .2300 \\
[1569, 2233] & .1150 \\
[2233, \infty) & 0 \\
\end{array}
\]
The generalized maximum likelihood estimator for 
\( P\{T_1^{\land}>t_1, T_2^{\land}>t_2\} \) is \( \hat{S}(t_1, t_2) \hat{S}_0(t_1 \vee t_2) \). On the other hand, a natural estimator for such probability is the proportion of observations with first [and second] component greater than \( t_1 \) [\( t_2 \)]. However, for \( t_1 = 400 \) and \( t_2 = 700 \), \( \hat{S}(400, 700) \hat{S}_0(700) \) is

\[
\frac{1}{25} \left( 1.1375 + 1.365 + 5.46 \right) \times .7903 = \frac{1}{25} \times 6.2927,
\]

but the proportion of observations with times of recurrence greater than 400 and death greater than 700 days is 6/25.
Figure 1. Times to first recurrence and death of twenty-five patients with malignant melanoma. Arrows are attached to censored observations.
Figure 2. Points, lines and region where the GMLE assigns positive probabilities.
Figure 3. Points, lines and region where the GMLE assigns positive probabilities after redistributing the masses on lines to points. Point masses generated by the redistribution are indicated by circles rather than dots.
REFERENCES


likelihood estimator in the presence of infinitely many 
incidental parameters. *Annals of Mathematical Statistics* 
27, 887-906.

survivorship function with doubly censored data. Un-
published manuscript.

Langberg, N. A., Proschans, F. and Quinzì, A. J. (1981). Estima-
ting dependent life lengths, with applications to the theory 

Meier, P. (1975). Estimation of a distribution function from 
incomplete observations. *Perspectives in Probability and 
Statistics (J. Gani, Ed.)*, 67-87. Applied Probability 
Trust, Sheffield, England.

of the distribution function from censored observations. 
Technical Report (to appear), Division of Biostatistics, 
Stanford University, Stanford, California.

Peterson, A. (1975). Nonparametric estimation in the competing 
risks problem. Technical Report No. 13 (R01 GM21215), 
Division of Biostatistics, Stanford University, Stanford, 
California.

(1977). Expressing the Kaplan-Meier estimator as a function 
of empirical subsurval functions. *Journal of the American 
Statistical Association* 72, 854-858.