THE UNIVERSITY OF CALGARY

ESTIMATION OF BIVARIATE SURVIVAL FUNCTION UNDER RANDOM CENSORSHIP

by

BONAVENTURE ANANI ANTHONIO

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

September, 1988

© B.A. ANTHONIO, 1988

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-46555-7

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled, "*Estimation of Bivariate Survival Function Under Random Censorship*" submitted by Bonaventure Anani Anthonio in partial fulfillment of the requirements for the degree of Master of Science.

M. D. Burke

M.D. Burke, Chairman
Department of Mathematics & Statistics

John R Collins

J.R. Collins
Department of Mathematics & Statistics

Dudley Paul Johnson

D.P. Johnson
Department of Mathematics & Statistics

E A Johnson

E.A. Johnson
Department of Biology

DATE: Oct 7, 1988

ABSTRACT

In many estimation problems it is inconvenient or impossible to make complete measurements on all members of a random sample. The problem is to estimate the distribution of the life times under random censoring. The one-dimensional random censoring model has been treated in great detail in the recent literature beginning with the landmark papers of Kaplan and Meier (1958).

For bivarite observations, the censoring may be univariate (homogeneous) or bivariate (heterogeneous). Various methods for the nonparametric estimation of a bivariate survival function in the presence of censoring have been proposed by a number of authors. The aim of this thesis is to compare the performance of some of the proposed estimators.

In Chapter I, we consider a reduced-sample estimator and a self-consistent one, both of which were proposed by Campbell (1980). An extension of the Kaplan and Meier (1958) estimator to the case of bivariate censored data as was proposed by Korwar and Dahiya (1982) is considered in Chapter II. In Chapter III, the line of reasoning by Hanley and Parnes (1983) is followed to develop a nonparametric maximum likelihood estimate of the underlying survival function under homogeneous censoring. Path dependent and closed form estimators, due

to Campbell and Földes (1980) are considered in Chapter IV. These estimators are modified in Chapter V to satisfy the monotonicity requirements of a survival function as was presented in a paper by Burke (1988). An approach due to Tsai, Leurgans and Crowley (1986) is employed in Chapter VI. This involves a decomposition of the bivariate survival function in terms of estimable functions.

A simple scheme for generating the random variables which were used in this study as well as a comparison of the performance of the estimators discussed in Chapters II, IV, V and VI are given in Chapter VII.

ACKNOWLEDGEMENT

It is a pleasure to thank my supervisor Dr. M.D. Burke for his crucial support and suggestions. He corrected innumerable errors of style and thought. Without his help I could not have brought this work to its completion. I wish to thank Dr. P.C. Consul for the guidance he gave me whilst serving as my interim adivsor.

I owe a special debt to a friend, Anup Gupta for teaching me all the programming I know. I am extremely grateful to him for the many hours he spent at the computer terminal with me. I also wish to express my indebtedness to Tim Nugent and Doug Taylor for the time they spent with me in the statistics laboratory.

Special thanks are also due to Messrs. Tak Fung, Kim Wagstaff and Don Kirkham, all of the Academic Computing Services, for their invaluable assistance.

I would like to acknowledge the financial assistance of the University of Calgary which enabled me to further my education here.

Finally, I am extremely grateful to Maria Dourado who gave much of her time in typing this thesis.

Dedicated to my beloved parents

TABLE OF CONTENTS

INTRODUCTION

In lifetesting, medical follow-up and other fields, the observation of the occurrence time of the event of interest (called a death) may be prevented for some of the items of the sample by the previous occurrence of some other event (called a loss). In many estimation problems incomplete measurements are thus made on some of the members of a random sample. For example, observation of the life of a vacuum tube may be ended by the breakage of the tube, or a need to use the test facilities for other purposes. In medical follow-up studies to determine the distribution of survival times after treatment, contact with some individuals may be lost before their death, and others may die from causes which we may desire to exclude from consideration. In the above examples, incomplete observations may also result from a need to complete a study within a certain time. An incomplete observation is said to be censored and its numerical value can be referred to as a limit of observation. These limits of observations are constants or values of other random variables, which are assumed independent of the complete observations.

In this thesis we shall consider the case of censored bivariate data. Two types of censoring arise in bivariate failure time data: univariate censoring and bivariate censoring. The former arises naturally in two contexts. Firstly, the experimental units may contain two similar components (such as ears, eyes, knees, etc.) whose joint survival is being studied. Alternatively, the experimental units may contain two dissimilar components whose joint survival is being studied. In both cases, censoring occurs when the experimental unit is removed from observation before both components have been observed to fail. Random variables need not be times in the usual sense. Variables could be cumulative dose or cumulative cost. In this case censoring occurs when an experimental unit (or component) is removed from observation for reasons independent of both responses.

To model the censoring, consider a sequence of independent random censoring vectors $\{(C_i, D_i)\}$, $i = 1,\dots,n$ from the bivariate distribution $G(s,t) = P(C > s, D > t)$. While such an assumption may not always be valid, it permits censoring times to differ. Further, let $\left\{(X_i^0, Y_i^0)\right\}$, $i = 1,\dots,n$ be independent pairs of random variables from a joint distribution function $F^d(s,t) = P(X^0 \le s, Y^0 \le t)$ where (X_i^0, Y_i^0), $i = 1,\dots,n$ is a random sample of pairs of lifetimes from (X^0, Y^0). The (X_i^0, Y_i^0), $i = 1,\dots,n$ are subject to random censorship from the right. Thus independent censoring vectors (C_i, D_i) exist for each bivariate vector (X_i^0, Y_i^0). The observed quantities are $(X_i, Y_i, \xi_{1i}, \xi_{2i})$, $i = 1,\dots,n$ where

$$X_i = \min(X_i^0,\ C_i), \qquad Y_i = \min(Y_i^0,\ D_i);$$

$$\xi_{1i} = \begin{Bmatrix} 1 & (X_i = X_i^0) \\ 0 & (X_i < X_i^0) \end{Bmatrix}, \qquad \xi_{2i} = \begin{Bmatrix} 1 & (Y_i = Y_i^0) \\ 0 & (Y_i < Y_i^0) \end{Bmatrix}.$$

The objective is to estimate the distribution function and its corresponding survival function $F(s,t) = P(X^0 > s,\ Y^0 > t)$ in the presence of censoring. A number of authors have proposed estimators for the survival function above. We shall introduce the work of some of these authors with a view towards comparing the proposed estimators. Throughout this thesis, we shall denote the various estimators by their designated numbers. For example, estimator (4.1.8) is an indication that this estimator is completely specified by equation (4.1.8) of Chapter IV, Section 1.

In Chapter I, we consider two estimators which were proposed by Campbell (1980) for the estimation problem with randomly censored discrete data. The first, which is estimator (1.1.1) is a reduced-sample estimator. The second, which is estimator (1.2.2) is a self-consistent one and is developed from a nonparametric likelihood function. In Chapter II, we consider estimator (2.2.3) which was proposed by Korwar and Dahiya (1982). This is an extension of the Kaplan and Meier (1958) estimator of a univariate survival function and involves an iterative approach to the bivariate estimation problem at hand. As in the case of estimator (1.2.2), it is shown that this estimator is self-consistent. A distinction is made between what we term univariate and bivariate censoring in Chapter III, as was proposed by Hanley and Parnes (1983). It is shown how a multivariate empirical survival function must be constructed in order to be considered a (non-parametric) maximum likelihood estimate of the underlying survival function. This construction results in estimator (3.2.6) and applies only to the case of univariate censoring. It is a closed form solution, similar to the product limit estimate of Kaplan and Meier (1958). Other closed-form estimators, due to Campbell and Földes (1980), for the bivariate model are considered in Chapter IV. Estimators (4.1.8) and (4.1.10) which are path dependent are also introduced. A hazard function approach is further employed to estimate $-\ln F(s,t)$ and hence $F(s,t)$. Consequently, two path-dependent estimators of $-\ln F(s,t)$ are proposed and these lead to estimators (4.2.19) and (4.2.21). It is also shown that this class of estimators may not necessarily be monotone nonincreasing in both coordinates.

In Chapter V, we follow the line of thinking of Burke (1988) to arrive at estimator (5.1.3) and (5.2.2). This approach involves a suitable modification of the Campbell and Földes (1982) estimators discussed in Chapter IV, to satisfy the important monotonicity requirements of a survival function. Changing the role of X and Y in each of these two estimators results in an additional estimator. Chapter VI, contains estimator (6.2.3) which was developed by Tsai, Leurgans and Crowley (1986). The approach employed here is to decompose the bivariate survival function in terms of estimable functions, some of which are kernel and bandwidth dependent.

A simple scheme for generating the random variables which were used in this study is given in Chapter VII as well as a comparison of the performance of estimators (4.1.8), (4.2.19), (5.1.3), (5.2.2), (6.3.2) and (2.2.3). Results from a computer algorithm written for the numerical solution of the said estimators is provided in tabular form and graphically for different sample sizes and censoring schemes. For each sample size, three different censoring schemes are considered:

(i) 10% censoring,

(ii) 40% censoring,

(iii) 50% censoring.

In the case of 40% censoring and estimators (4.1.8), (4.2.19), (5.1.3), (5.2.2) and (6.3.2), simulation results are provided for sample sizes of 10, 30, 50, ..., 170, where as in the case of 40% censoring and estimator (2.2.3) such results are only provided for sample sizes of 10, 30, ..., 90. For 10% and 50% censoring, simulation results are given for sample sizes of 10, 30, ..., 90 for all the said estimators.

CHAPTER I

NONPARAMETRIC BIVARIATE ESTIMATION WITH RANDOMLY CENSORED DISCRETE DATA

1.0 INTRODUCTION

In this chapter, we shall develop two estimators of a bivariate distribution function using randomly censored discrete data as presented in a paper by Campbell (1980).

It is assumed that the censoring occurs independently of the lifetimes and that deaths and losses which occur simultaneously can be separated. In such cases, it is conventional to assume that deaths precede losses.

In Section 1, a reduced-sample estimator is given. A nonparametric approach and the related self-consistency technique of Efron (1967) is focused on in Section 2 to arrive at a self-consistent estimator. Section 2 also contains a discussion on the relative merits of the proposed estimators.

1.1 THE BIVARIATE REDUCED-SAMPLE ESTIMATOR

Let the censoring variables $\{(C_i, D_i)\}$, $i = 1,2,\ldots,n$ be from the continuous joint survival distribution

$$G(s,t) = P(C > s, D > t)$$

Also let the pairs of observed lifetimes $\{(X_i, Y_i)\}$, $i = 1,\ldots,n$ be a random sample from a continuous joint survival distribution

$$H(s,t) = P(X > s, Y > t).$$

Due to the earlier assumption that (C_i, D_i) is independent of (X_i^0, Y_i^0) for $i = 1,\ldots,n$, we have

$$H(s,t) = F(s,t)\, G(s,t),$$

which implies

$$F(s,t) = H(s,t)/G(s,t).$$

In a case where not only are X_i, Y_i, ξ_{1i}, ξ_{2i} available for each pair, but also C_i, D_i, then an estimate of the survival function is given by

$$\tilde{F}(s,t) = \frac{H_n(s,t)}{\tilde{G}_n(s,t)} \tag{1.1.1}$$

where $n\,\tilde{G}_n(s,t)$ is the number of pairs such that $C_i \geq s$ and $D_i \geq t$ and $n\,H_n(s,t)$ is the number of pairs for which $X_i^0 > s$, $Y_i^0 > t$ with $C_i \geq s$ and $D_i \geq t$. This is appropriate because when $X_i = s$ and $\xi_{1i} = 0$ then $X_i^0 > s$. Similarly when $Y_i = t$ and $\xi_{2i} = 0$ then $Y_i^0 > t$. $\tilde{F}(s,t)$ in (1.1.1) is the bivariate reduced-sample estimate.

1.2 A BIVARIATE SELF-CONSISTENT ESTIMATOR WITH DISCRETE TIMES OF DEATH OR LOSSES.

In this section we shall consider the bivariate estimation problem with discrete times of death or losses as was considered by Campbell (1980). An extension of the self-consistent approach of Efron (1967) shall be used.

Let $S_1, \ldots, S_I$ be the distinct times for losses or deaths for the first item of the pair, and let $t_1, \ldots, t_J$ be the corresponding distinct times for the second item.

Let us define δ_{ij}, α_{ij}, β_{ij} and λ_{ij} as follows:

δ_{ij} is the number of pairs for which $X_k = S_i$, $Y_k = t_j$, $\xi_{1k} = 1$, $\xi_{2k} = 1$;

α_{ij} is the number of pairs for which $X_k = S_i$, $Y_k = t_j$, $\xi_{1k} = 0$, $\xi_{2k} = 1$;

β_{ij} is the number of pairs for which $X_k = S_i$, $Y_k = t_j$, $\xi_{1k} = 1$, $\xi_{2k} = 0$;

λ_{ij} is the number of pairs for which $X_k = S_i$, $Y_k = t_j$, $\xi_{1k} = 0$, $\xi_{2k} = 0$.

The convention that deaths precede losses is adopted to separate deaths and losses which occur at any point (s_i, t_j). It is therefore in order to say that in the first coordinate, the pairs δ_{ij} and β_{ij} precede α_{ij} and λ_{ij}; in the second coordinate δ_{ij} and α_{ij} precede β_{ij} and λ_{ij}.

Let $F_{ij} = F(s_i, t_j) = P(X^0 > S_i, Y^0 > t_j)$

for $i = 1, \ldots, I, \quad j = 1, \ldots, J$.

Also let $\Delta_{ij} = F_{ij} + F_{i-1,j-1} - F_{i,j-1} - F_{i-1,j}$ be the probability of death in rectangle $(S_{i-1}, S_i] \times (t_{j-1}, t_j]$, $Q_{ij} = F_{i,j-1} - F_{ij}$ is the probability of death in $(S_i, \infty) \times (t_{j-1}, t_j]$, and $R_{ij} = F_{i-1,j} - F_{ij}$ is the probability of death in $(S_{i-1}, S_i] \times (t_j, \infty)$. The graph below is a pictorial representation of the above.

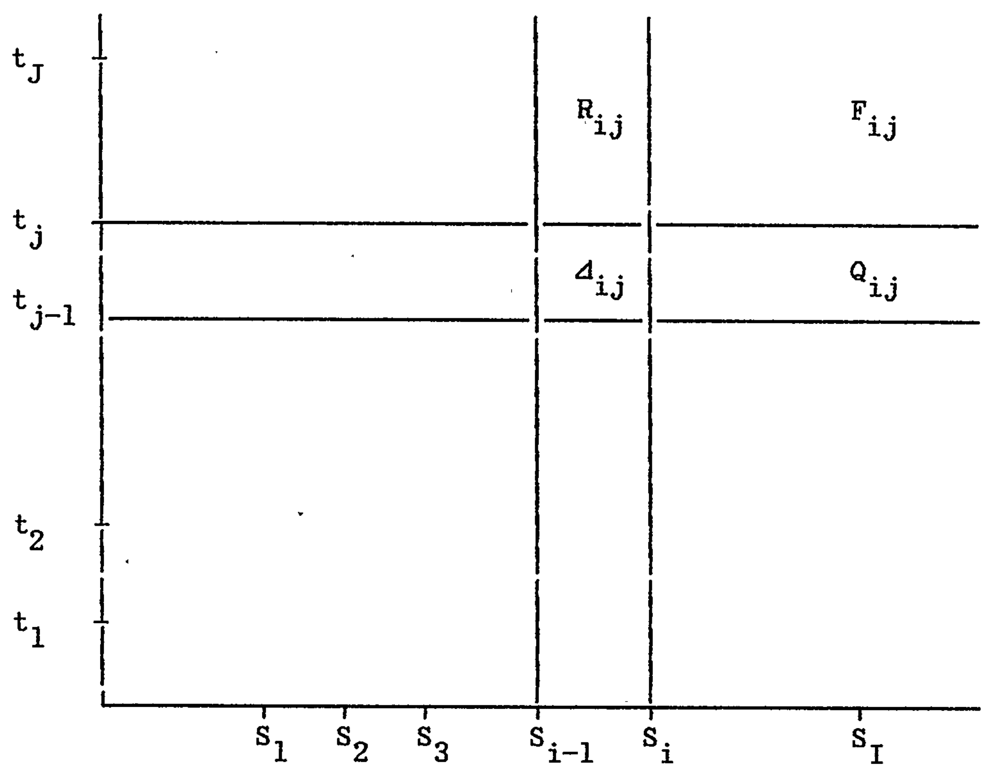

Figure 1.1.1

PLEASE NOTE THERE ARE TWO(2) PAGES NUMBERED 10. THE TEXT, HOWEVER, IS DIFFERENT.

VEUILLEZ NOTER QU'IL Y A DEUX FEUILLETS NUMEROTES NO.10. CEPENDANT, LE CONTENU EST SENSIBLEMENT DIFFERENT.

The nonparametric likelihood function is thus given by:

$$L = \prod_{i=1}^{I} \prod_{j=1}^{J} \Delta_{ij}^{\delta_{ij}} F_{ij}^{\lambda_{ij}} Q_{ij}^{\alpha_{ij}} R_{ij}^{\beta_{ij}}.$$

The value of F_{ij} which maximizes L for fixed values of δ_{ij}, λ_{ij}, α_{ij} and β_{ij} is given by setting $\frac{d}{dF_{ij}} \log L$ equal to zero to get the likelihood equation (1.2.1). Hence

$$\text{Log } L = \sum_{i=1}^{I} \sum_{j=1}^{J} \left[\delta_{ij} \log \Delta_{ij} + \lambda_{ij} \log F_{ij} + \alpha_{ij} \log Q_{ij} + \beta_{ij} \log R_{ij}\right].$$

Keeping in mind that;

$$\begin{aligned}
\Delta_{ij} &= F_{ij} + F_{i-1,j-1} - F_{i,j-1} - F_{i-1,j} \\
\Delta_{i+1,j+1} &= F_{i+1,j+1} + F_{ij} - F_{i+1,j} - F_{i,j+1} \\
\Delta_{i,j+1} &= F_{i,j+1} + F_{i-1,j} - F_{ij} - F_{i-1,j+1} \\
\Delta_{i+1,j} &= F_{i+1,j} + F_{i,j-1} - F_{i+1j-1} - F_{ij} \\
Q_{ij} &= F_{i,j-1} - F_{ij} \\
Q_{i,j+1} &= F_{ij} - F_{i,j+1} \\
R_{ij} &= F_{i-1,j} - F_{ij} \\
R_{i+1,j} &= F_{ij} - F_{i+1,j}.
\end{aligned}$$

We have,

$$\frac{d}{dF_{ij}} \text{Log } L = \frac{\delta_{ij}}{\Delta_{ij}} + \frac{\delta_{i+1,j+1}}{\Delta_{i+1,j+1}} - \frac{\delta_{i,j+1}}{\Delta_{i,j+1}} - \frac{\delta_{i+1,j}}{\Delta_{i+1,j}} + \frac{\lambda_{ij}}{F_{ij}}$$

$$+ \frac{\beta_{i+1,j}}{R_{i+1,j}} - \frac{\beta_{ij}}{R_{ij}} + \frac{\alpha_{i,j+1}}{Q_{i,j+1}} - \frac{\alpha_{ij}}{Q_{ij}} .$$

The nonparametric likelihood function is thus given by:

$$L = \prod_{i=1}^{I} \prod_{j=1}^{J} \Delta_{ij}^{\delta_{ij}} F_{ij}^{\lambda_{ij}} Q_{ij}^{\alpha_{ij}} R_{ij}^{\beta_{ij}}.$$

The value of F_{ij} which maximizes L for fixed values of δ_{ij}, λ_{ij}, α_{ij} and β_{ij} is given by setting $\frac{d}{dF_{ij}} \log L$ equal to zero to get the likelihood equation (1.2.1). Hence

$$\text{Log } L = \sum_{i=1}^{I} \sum_{j=1}^{J} \left[\delta_{ij} \log \Delta_{ij} + \lambda_{ij} \log F_{ij} + \alpha_{ij} \log Q_{ij} + \beta_{ij} \log R_{ij}\right].$$

Keeping in mind that;

$$\begin{aligned}
\Delta_{ij} &= F_{ij} + F_{i-1,j-1} - F_{i,j-1} - F_{i-1,j} \\
\Delta_{i+1,j+1} &= F_{i+1,j+1} + F_{ij} - F_{i+1,j} - F_{i,j+1} \\
\Delta_{i,j+1} &= F_{i,j+1} + F_{i-1,j} - F_{ij} - F_{i-1,j+1} \\
\Delta_{i+1,j} &= F_{i+1,j} + F_{i,j-1} - F_{i+1j-1} - F_{ij} \\
Q_{ij} &= F_{i,j-1} - F_{ij} \\
Q_{i,j+1} &= F_{ij} - F_{i,j+1} \\
R_{ij} &= F_{i-1,j} - F_{ij} \\
R_{i+1,j} &= F_{ij} - F_{i+1,j}.
\end{aligned}$$

We have,

$$\frac{d}{dF} \text{ Log } L = \frac{\delta_{ij}}{\Delta} + \frac{\delta_{i+1,j+1}}{\Delta} - \frac{\delta_{i,j+1}}{\Delta} - \frac{\delta_{i+1,j}}{\Delta} + \frac{\lambda_{ij}}{F}$$

Therefore the likelihood equation is

$$(1.2.1) \qquad \frac{\delta_{ij}}{\Delta_{ij}} + \frac{\delta_{i+1,j+1}}{\Delta_{i+1,j+1}} - \frac{\delta_{i,j+1}}{\Delta_{i,j+1}} - \frac{\delta_{i+1,j}}{\Delta_{i+1,j}} + \frac{\lambda_{ij}}{F_{ij}} + \frac{\beta_{i+1,j}}{R_{i+1,j}} - \frac{\beta_{ij}}{R_{ij}} + \frac{\alpha_{i,j+1}}{Q_{i,j+1}} - \frac{\alpha_{ij}}{Q_{ij}} = 0.$$

We shall now consider a self-consistent approach to the problem of estimating F_{ij}. To estimate F_{ij} at time (s_i,t_j), we would count all N_{ij} pairs that are known to be alive at (s_i,t_j). This figure excludes deaths at S_i or at t_j but not losses. If $k < i$ and $\ell > j$, the probability that a pair which was censored in the first coordinate at S_k but died in the second coordinate at time t_ℓ will survive to (s_i,t_ℓ) is $Q_{i\ell}/Q_{k\ell}$. Therefore the expected number of the $\alpha_{k\ell}$ pairs to survive to (s_i, t_ℓ) is $\alpha_{k\ell}\,\frac{Q_{i\ell}}{Q_{k\ell}}$. Similarly if $k > i$, $\ell < j$, for the $\beta_{k\ell}$ pairs censored in the second coordinate, the expected number to survive to (s_k, t_j) is $\beta_{k\ell}\; R_{kj}/R_{k\ell}$. Finally, if $k < i$ or $\ell < j$, for the $\lambda_{k\ell}$ doubly censored pairs, the expected number to survive to (s_i,t_j) is $\lambda_{k\ell}\; F_{\max(i,k),\max(j,\ell)}/F_{k\ell}$.

We can thus define an estimate $\hat{F}_{ij}$ of F_{ij} to be self-consistent if it satisfies the equation

$$(1.2.2) \qquad n\,\hat{F}_{ij} = N_{ij} + \sum_{\ell>j,k<i} \alpha_{k\ell}\,\frac{\hat{Q}_{i\ell}}{\hat{Q}_{k\ell}} + \sum_{k>i,\ell<j} \beta_{k\ell}\,\frac{\hat{R}_{kj}}{\hat{R}_{k\ell}} + \sum_{k<i \text{ or } \ell<j} \lambda_{k\ell}\,\frac{\hat{F}_{\max(i,k),\max(j,\ell)}}{\hat{F}_{k\ell}},$$

where

$$\hat{Q}_{ij} = \hat{F}_{i,j-1} - \hat{F}_{ij}, \quad \hat{R}_{ij} = \hat{F}_{i-1,j} - \hat{F}_{ij},$$

$$N_{ij} = \sum_{k>i,\ell>j} \delta_{k\ell} + \sum_{k\geq i,\ell>j} \alpha_{k\ell} + \sum_{k>i,\ell\geq j} \beta_{k\ell} + \sum_{k\geq i,\ell\geq j} \lambda_{k\ell}.$$

Equation (1.2.2) is called a bivariate self-consistent equation and a function satisfying it is called a self-consistent estimate of F_{ij}.

Theorem 1.1

An estimate $\hat{F}_{ij}$ satisfying (1.2.2) is also a solution of (1.2.1) with F_{ij} replaced by $\hat{F}_{ij}$ and Δ_{ij} by $\hat{\Delta}_{ij} = \hat{F}_{ij} + \hat{F}_{i-1,j-1} - \hat{F}_{i-1,j} - \hat{F}_{i,j-1}$.

Proof: We have

$$\hat{\Delta}_{ij} = \hat{F}_{ij} + \hat{F}_{i-1,j-1} - \hat{F}_{i-1,j} - \hat{F}_{i,j-1},$$

which implies

$$n\,\hat{\Delta}_{ij} = n\left[\hat{F}_{ij} + \hat{F}_{i-1,j-1} - \hat{F}_{i-1,j} - \hat{F}_{i,j-1}\right].$$

From (1.2.2) we have

$$n\,\hat{\Delta}_{ij} = \left[N_{ij} + \sum_{\ell>j,k<i} \alpha_{k\ell} \frac{\hat{Q}_{i\ell}}{\hat{Q}_{k\ell}} + \sum_{k>i,\ell<j} \beta_{k\ell} \frac{\hat{R}_{kj}}{\hat{R}_{k\ell}} \right.$$

$$\left. + \sum_{k<i \text{ or } \ell<j} \lambda_{k\ell} \frac{\hat{F}_{\max(i,k),\max(j,\ell)}}{\hat{F}_{k\ell}}\right] +$$

$$\left[N_{i-1,j-1} + \sum_{\ell>j-1,k<i-1} \alpha_{k\ell} \frac{\hat{Q}_{i-1,\ell}}{\hat{Q}_{k\ell}} + \sum_{k>i-1,\ell<j-1} \beta_{k\ell} \frac{\hat{R}_{k,j-1}}{\hat{R}_{k\ell}}\right.$$

$$\left. + \sum_{k<i-1 \text{ or } \ell<j-1} \lambda_{k\ell} \frac{\hat{F}_{\max(i-1,k),\max(j-1,\ell)}}{\hat{F}_{k\ell}}\right]$$

$$- \left[N_{i-1,j} + \sum_{\ell>j,k<i-1} \alpha_{k\ell} \frac{\hat{Q}_{i-1,\ell}}{\hat{Q}_{k\ell}} + \sum_{k>i-1,\ell<j} \beta_{k\ell} \frac{\hat{R}_{kj}}{\hat{R}_{k\ell}}\right.$$

$$\left. + \sum_{k<i-1 \text{ or } \ell<j} \lambda_{k\ell} \frac{\hat{F}_{\max(i-1,k),\max(j,\ell)}}{\hat{F}_{k\ell}}\right]$$

$$- \left[N_{i,j-1} + \sum_{\ell>j-1,k<i} \alpha_{k\ell} \frac{\hat{Q}_{i\ell}}{\hat{Q}_{k\ell}} + \sum_{k>i,\ell<j-1} \beta_{k\ell} \frac{\hat{R}_{k,j-1}}{\hat{R}_{k\ell}}\right.$$

$$\left. + \sum_{k<i \text{ or } \ell<j-1} \lambda_{k\ell} \frac{\hat{F}_{\max(i,k),\max(j-1,\ell)}}{\hat{F}_{k\ell}}\right]$$

$$= \delta_{ij} + \sum_{k\leq i-1} \alpha_{kj} \left[\frac{\hat{Q}_{i-1,j}-Q_{ij}}{\hat{Q}_{kj}}\right] + \sum_{\ell\leq j-1} \beta_{i\ell} \left[\frac{\hat{R}_{i,j-1}-\hat{R}_{ij}}{\hat{R}_{i\ell}}\right]$$

$$+ \sum_{k\leq i-1,\ell\leq j-1} \lambda_{k\ell} \frac{\hat{R}_{i,j-1}-\hat{R}_{ij}}{\hat{F}_{k\ell}} .$$

Dividing this equation by $\hat{\Delta}_{ij} = R_{i,j-1} - \hat{R}_{ij} = \hat{Q}_{i-1,j} - \hat{Q}_{ij}$ gives

$$\text{(1.2.3)} \qquad n = \frac{\delta_{ij}}{\hat{\Delta}_{ij}} + \sum_{k\leq i-1} \frac{\alpha_{kj}}{\hat{Q}_{kj}} + \sum_{\ell\leq j-1} \frac{\beta_{i\ell}}{\hat{R}_{i\ell}} + \sum_{k\leq i-1,\ell\leq j-1} \frac{\lambda_{k\ell}}{\hat{F}_{k\ell}}.$$

We can form three other equations similar to (1.2.3) based on $\hat{\Delta}_{i+1,j}$, $\hat{\Delta}_{i,j+1}$, $\hat{\Delta}_{i+1,j+1}$.

Adding equations based on $\hat{\Delta}_{ij}$ and $\hat{\Delta}_{i+1,j+1}$ and subtracting those based on $\hat{\Delta}_{i+1,j}$ and $\hat{\Delta}_{i,j+1}$, we have

$$\frac{\delta_{ij}}{\hat{\Delta}_{ij}} + \frac{\delta_{i+1,j+1}}{\hat{\Delta}_{i+1,j+1}} - \frac{\delta_{i,j+1}}{\hat{\Delta}_{i,j+1}} - \frac{\delta_{i+1,j}}{\hat{\Delta}_{i+1,j}} + \frac{\lambda_{ij}}{\hat{F}_{ij}}$$

$$+ \frac{\beta_{i+1,j}}{\hat{R}_{i+1,j}} - \frac{\beta_{ij}}{\hat{R}_{ij}} + \frac{\alpha_{i,j+1}}{\hat{Q}_{i,j+1}} - \frac{\alpha_{ij}}{\hat{Q}_{ij}} = 0.$$

The resultant equation is (1.2.1) with F_{kl} replaced by $\hat{F}_{kl}$. This completes the proof.

The bivariate self-consistent estimator (1.2.2), while more difficult to compute than $\tilde{F}$ (estimator (1.1.1)), does not require the complete censoring information concerning the C_i's and D_i's. Further, it is always a distribution function; in the event of no censoring it reduces to the empirical distribution function. The estimator jumps only at the points of double deaths or final censored values in any dimension. The fact that estimator (1.2.2) depends on several estimable functions is likely to be a drawback in this case. The estimated survival probabilities provided by this estimator may often be greater than the real probabilities for most of the sample points even when this point is not censored. The reason being that in deriving the self-consistent estimator the weight of the censored observations is spread on all the points beyond the censored point.

CHAPTER II

AN ITERATIVE APPROACH TO THE BIVARIATE ESTIMATION PROBLEM

2.0 INTRODUCTION

Kaplan and Meier (1958) give a maximum likelihood estimator of the distribution function based on a univariate right censored sample. In the present chapter, we will investigate the extension of their results to the case of bivariate right censored samples as was presented in a paper by Ramesh M. Korwar and Ram C. Dahiya (1982). Following Efron (1967), we will also provide "self-consistent" estimators for the bivariate distribution function.

It is appropriate at this stage to give the univariate Kaplan-Meier PL (Product Limit) estimator.

Let X_i^0 be the lifetimes and C_i the corresponding censoring variables or limits of observation, where $i = 1,2,\ldots,n$. Also let X_i denote the observed lifetimes. Then,

$$X_i = \min(X_i^0, C_i),$$

$$\xi_{1i} = \begin{cases} 1 & (X_i = X_i^0), \\ 0 & (X_i < X_i^0). \end{cases}$$

The Kaplan Meier PL estimator of $F(s) = P(X^0 > s)$ is given by

$$\text{(2.0.1)} \qquad \hat{F}(s) = \begin{cases} \prod_{i=1}^{k-1} \left[\frac{n-i}{n-i+1}\right]^{\xi_{1i}}, & \text{if } s \in \left[X_{(k-1)}, X_{(k)}\right] \\ 0 & , \text{ if } s > X_{(n)}, \end{cases}$$

and where

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$$

is an arrangement in an ascending order of magnitude of the X_i values.

We will now introduce the idea of a "self-consistent" estimator.

Let

$$N_X(s) = \text{number of } X_i > s$$

and

$$N_{X^0}(s) = \text{number of } X_i^0 > s.$$

We would like to define $\hat{F}^0(s) = N_{X^0}(s)/n$, which is a binomial estimate of $F(s)$. Due to censorship, the function $N_{X^0}(s)$ is not available to us. Since $X_i^0 \geq X_i$ for every i, we do know that $X_i > s$ implies $X_i^0 > s$, so $N_{X^0}(s) \geq N_X(s)$ for every s. For an $X_i \leq s$ that is uncensored, $X_i^0 = X_i \leq s$, and X_i^0 cannot contribute to $N_{X^0}(s)$. The ambiguous situation is $X_i \leq s$, X_i censored ($\xi_{1i} = 0$), in which case X_i^0 will be greater than s with conditional probability $F^0(s)/F^0(X_i)$. We do not know $F^0(s)$, but given any initial estimate of it, say $F_1^0(s)$, it seems natural to estimate the conditional probability $P(X_i^0 > S | X_i \leq s, \xi_{1i} = 0)$ by $\hat{F}_1^0(s)/F_1^0(X_i)$ and define an improved estimate of $F^0(s)$ by

$$n\,\hat{F}_2^0(s) = N_X(s) + \sum_{\substack{X_i \le s \\ \xi_{1i}=0}} \frac{\hat{F}_1^0(s)}{\hat{F}_1^0(X_i)} \tag{2.0.2}$$

$$= N_X(s) + \sum_{X_i \le s} (1-\xi_{1i}) \frac{\hat{F}_1^0(s)}{\hat{F}_1^0(X_i)} \,.$$

Iterating, we could then use $\hat{F}_2^0$ in place of $\hat{F}_1^0$ above to get another improved estimate $\hat{F}_3^0$, and so forth. The question arises whether the sequence $\hat{F}_1^0$, $\hat{F}_2^0$, $\hat{F}_3^0$ ... would converge to a function $\hat{F}_0$ which could then not be further improved by application of (2.0.2). Such a function would have to satisfy

$$n\,\hat{F}^0(s) = N_X(s) + \sum_{X_i \le s} (1-\xi_{1i}) \frac{\hat{F}^0(s)}{\hat{F}^0(X_i)} \tag{2.0.3}$$

for all s. Efron (1967) called a function satisfying (2.0.3) a "self consistent" estimate of $F^0(s)$.

For the bivariate censored data, we shall consider the case of one variable being censored in Section 1, where it will be shown that there does exist a unique "self consistent" estimator. We will also provide an estimator of F when both variables are censored in Section 2 of this chapter.

2.1 THE CASE OF ONLY ONE VARIABLE BEING CENSORED

Suppose that only one of the random variables X^0, Y^0, say Y^0 is censored. Then we have

$$
\begin{aligned}
F(s,t) &= P(X^0 > s,\ Y^0 > t) \\
&= P(X^0 > s)\ P(Y^0 > t \mid X^0 > s) \\
&= H(s)\ G(t \mid s)
\end{aligned}
$$

where $G(t|s) = P(Y^0 > t \mid X^0 > s)$, and $H(s) = P(X^0 > s)$. A "self consistent" estimator of $F(s,t)$ can now be obtained as the censoring only affects the estimation of $G(t|s)$ which is a univariate estimation problem.

Since X^0 is uncensored, an estimate of $H(s)$ can be given by

$$\hat{H}(s) = n_s/n$$

where

$$n_s = \#\text{ of } X_i\text{'s each of which is greater than } s.$$

Similarly, an estimate of $G(t|s)$ is given by the Kaplan-Meier PL estimator (2.0.1) as

$$
(2.1.1) \qquad \hat{G}(t/s) = \begin{cases} \prod_{i=1}^{k-1} \left[\dfrac{n_s - i}{n_s - i + 1}\right]^{\xi_{2i}}, & \text{if } t \in \left[Y^{(s)}_{(k-1)}, Y^{(s)}_{(k)}\right] \\ 0, & \text{if } t > Y^{(s)}_{(n_s)} \end{cases}
$$

where

$$Y^{(s)}_{(1)} \le \ldots\ldots \le Y^{(s)}_{(n_s)}$$

is an arrangement in an ascending order of magnitude of those ordinates of the pairs (X_i, Y_i) for which $X_i > s$. Thus for each $s \ge 0$, we look at the pairs (X_i, Y_i) for which $X_i > s$ and form $\hat{G}(t|s)$, the univariate Kaplan-Meier PL estimator, using the ordinates of these pairs. Therefore an estimator of $F(s,t)$ is given by

$$\hat{F}(s,t) = \frac{n_s}{n}\hat{G}(t|s). \tag{2.1.2}$$

Theorem 2.1

The estimator (2.1.2) is unique and "self-consistent" and satisfies the equation

$$n\,F(s,t) = n_{s,t} + \sum_{\substack{X_i>s\\ Y_i\le t}} (1-\xi_{2i})\frac{\hat{F}(s,t)}{\hat{F}(s,Y_i)}, \tag{2.1.3}$$

where,

$$n_{s,t} = \#\text{ of pairs } (X_i,\ Y_i) \text{ such that } X_i > s,\ Y_i > t.$$

Proof:

Since $\hat{G}(t/s)$ is the univariate Kaplan-Meier PL estimator which was formed using the ordinates of the pairs $(X_i,\ Y_i)$ for which $X_i > s$, then it has to satisfy (2.0.3) with n replaced by n_s and $N_X(s)$ replaced by $n_{s,t}$.

Thus,

$$n_s\,\hat{G}(t|s) = n_{s,t} + \sum_{\substack{X_i>s\\ Y_i\le t}} (1-\xi_{2i})\frac{\hat{G}(t|s)}{\hat{G}(Y_i|s)}.$$

Which is equivalent to

$$n\cdot\frac{n_s}{n}\hat{G}(t|s) = n_{s,t} + \sum_{\substack{X_i>s\\ Y_i\le t}} (1-\xi_{2i})\frac{\frac{n_s}{n}\hat{G}(t|s)}{\frac{n_s}{n}\hat{G}(Y_i|s)}.$$

But $\frac{n_s}{n}\hat{G}(t|s) = \hat{F}(s,t)$, thus we have

$$n\,\hat{F}(s,t) = n_{s,t} + \sum_{\substack{X_i > s \\ Y_i \le t}} (1-\delta_{2i}) \frac{\hat{F}(s,t)}{\hat{F}(s,Y_i)} . \tag{2.1.4}$$

Equation (2.1.4) is essentially (2.0.3). Thus it follows that $F(s,t) = n_s\, G(t|s)/n$ satisfies (2.1.3). To show uniqueness, let if possible $\tilde{F}(s,t)$ be another "self-consistent" estimator satisfying (2.1.3). Define a function $\hat{G}^*(t|s)$ by

$$\frac{n_s}{n} \hat{G}^*(t|s) = \tilde{F}(s,t).$$

Since $\tilde{F}(s,t)$ satisfies (2.1.3) it follows that $\hat{G}^*(t|s)$ must satisfy (2.1.4). But (2.1.4) has the unique solution $\hat{G}(t/s)$. Thus

$$G^*(t/s) = \hat{G}(t|s) \quad \text{and} \quad \tilde{F}(s,t) = \frac{n_s}{n} \hat{G}(t|s).$$

We will now discuss the large-sample properties of the estimator (2.1.2).

Theorem 2.1.2

The estimator given by (2.1.2) is pointwise consistent.

Proof:

The theorem follows by noting that; (1) the estimator is a product of two one-dimensional Kaplan-Meier (P-L) estimators and (2) the one dimensional (P-L) estimator is pointwise consistent.

For stating theorem (2.1.3) we need more notation.

Let $Y_i = \min(Y_i^0, D_i)$.

where $D_1,\ldots,D_n$ are the censoring variables. We assume that $D_1,\ldots,D_n$ are i.i.d and distributed independent of (X_i^0, Y_i^0) $i = 1,\ldots,n$. Let

$$G_0(t) = P(D \geq t),\ t \geq 0$$

and

$$H_0(s,t) = P(X \geq s,\ Y \geq t).$$

Theorem 2.1.3

If G_0 and $F(s,t)$ are continuous and if F is such that log F is absolutely continuous with partial derivatives that exist almost everywhere, and if, for $0 < S,\ T < \infty$, $H_0(S,T) > 0$, then

$$\sup_{\substack{0\leq s\leq S \\ 0\leq t\leq T}} |\hat{F}(s,t) - F(s,t)| = 0\left[\sqrt{\frac{\log\log n}{n}}\right].$$

Proof:

This is corollary 5.2 of Campbell and Földes (1982) as applied to the situation at hand.

2.2 THE CASE OF BOTH VARIABLES BEING CENSORED

Using Efron's (1967) concept of "self-consistency", we approach the problem of estimating $F(s,t)$, when both variables are subject to right-censoring, as follows. The contribution of $n_{s,t}$ pairs (X_i, Y_i) to $\hat{F}(s,t)$, is clear whether or not for these pairs X_i, Y_i or both are censored. What is not clear is the contribution of the following:

1) $(X_i > s \cdot Y_i \le t,\ \xi_{1i} = 1,\ \xi_{2i} = 0)$,

2) $(X_i > s \cdot Y_i \le t,\ \xi_{1i} = 0,\ \xi_{2i} = 0)$,

3) $(X_i \le s \cdot Y_i > t,\ \xi_{1i} = 0,\ \xi_{2i} = 1)$,

4) $(X_i \le s \cdot Y_i > t,\ \xi_{1i} = 0,\ \xi_{2i} = 0)$,

and

5) $(X_i \le s,\ Y_i \le t,\ \xi_{1i} = 0,\ \xi_{2i} = 0)$.

In these ambiguous cases, we proceed as follows.

We estimate the conditional probability that $X_i^0 > s$ and $Y_i^0 > t$ given $X_i > s$, $Y_i \le t$, $\xi_{1i} = 1$, $\xi_{2i} = 0$ by

$$(2.2.1)\qquad P(X_i^0 > s,\ Y_i^0 > t \mid X_i > s,\ Y_i \le t,\ \xi_{1i} = 1,\ \xi_{2i} = 0)$$
$$= F(s,t)/F(s,Y_i).$$

Of course, we do not know $F(s,t)$. If $\hat{F}_1(s,t)$ is an initial estimator of $F(s,t)$, it seems natural to estimate the conditional probability in (2.2.1) by $\hat{F}_1(s,t)/\hat{F}_1(s,Y_i)$. Similarly we estimate

$$P(X_i^0>s,\ Y_i^0>t \mid X_i>s,\ Y_i\le t,\ \xi_{1i} = 0,\ \xi_{2i} = 0) \text{ by } \hat{F}_1(X_i,t)/\hat{F}_1(X_i,\ Y_i),$$

$$P(X_i^0>s,\ Y_i^0>t \mid X_i\le s,\ Y_i>t,\ \xi_{1i} = 0,\ \xi_{2i} = 1) \text{ by } \hat{F}_1(s,t)/\hat{F}_1(X_i,t),$$

$$P(X_i^0>s,\ Y_i^0>t \mid X_i\le s,\ Y_i>t,\ \xi_{1i} = 0,\ \xi_{2i} = 0) \text{ by } \hat{F}_1(s,Y_i)/\hat{F}_1(X_i,Y_i)$$

and

$$P(X_i^0>s,\ Y_i^0>t \mid X_i\le s,\ Y_i\le t,\ \xi_{1i} = 0,\ \xi_{2i} = 0) \text{ by } \hat{F}_1(s,t)/\hat{F}_1(X_i,Y_i).$$

Thus we can define an improved estimator $\hat{F}_2(s,t)$ of F by

$$(2.2.2) \quad n\,\hat{F}_2(s,t) = n_{s,t} + \sum_{\substack{X_i>s \\ Y_i\le t}} \xi_{1i}\,(1-\xi_{2i})\,\hat{F}_1(s,t)/\hat{F}_1(s,Y_i)$$

$$+ \sum_{\substack{X_i>s \\ Y_i\le t}} (1-\xi_{1i})\,(1-\xi_{2i})\,\hat{F}_1(X_i,t)/\hat{F}_1(X_i,Y_i)$$

$$+ \sum_{\substack{X_i\le s \\ Y_i>t}} (1-\xi_{1i})\,\xi_{2i}\,\hat{F}_1(s,t)/\hat{F}_1(X_i,t)$$

$$+ \sum_{\substack{X_i\le s \\ Y_i>t}} (1-\xi_{1i})\,(1-\xi_{2i})\,\hat{F}_1(s,Y_i)/\hat{F}_1(X_i,Y_i)$$

$$+ \sum_{\substack{X_i\le s \\ Y_i\le t}} (1-\xi_{1i})\,(1-\xi_{2i})\,\hat{F}_1(s,t)/\hat{F}_1(X_i,Y_i).$$

For iterating we could then use $\hat{F}_2$ in place of F_1 above and obtain yet another improved estimator $\hat{F}_3$, and thus obtain a sequence of estimators $\hat{F}_1$, $\hat{F}_2$, $\hat{F}_3$, ... This sequence will converge to a function $\hat{F}$ which satisfies

$$(2.2.3) \qquad n\,\hat{F}(s,t) = n_{s,t} + \sum_{\substack{X_i>s \\ Y_i\le t}} \xi_{1i}\,(1-\xi_{2i})\,\hat{F}(s,t)/\hat{F}(s,Y_i)$$

$$+ \sum_{\substack{X_i > s \\ Y_i \le t}} (1-\xi_{1i})\,(1-\xi_{2i})\,\hat{F}(X_i,t)/\hat{F}(X_i,Y_i)$$

$$+ \sum_{\substack{X_i \le s \\ Y_i > t}} (1-\xi_{1i})\,\xi_{2i}\,\hat{F}(s,t)/\hat{F}(X_i,t)$$

$$+ \sum_{\substack{X_i \le s \\ Y_i > t}} (1-\xi_{1i})\,(1-\xi_{2i})\,\hat{F}(s,Y_i)/\hat{F}(X_i,Y_i)$$

$$+ \sum_{\substack{X_i \le s \\ Y_i \le t}} (1-\xi_{1i})\,(1-\xi_{2i})\,\hat{F}(s,t)/\hat{F}(X_i,Y_i).$$

Following Efron (1967) we will call a function $\hat{F}$ satisfying (2.2.3) a "self-consistent" estimator of F.

The possible drawbacks of estimator (2.2.3) are similar to those of estimator (1.2.2) in the sense that it also depends on several estimable fuctions. Here again, there is the possibility for a consistent over estimation of survival probabilities as indicated by the simulation results of Chapter VII. The obvious explanation for this is the fact that at most sample points, the weight of the censored observations is spread on all the points beyond the censored point. As is typical of convergent related problems it was realized that the computer algorithm written for the numerical solution of estimator (2.2.3) took several iterations to arrive at the estimated survival probabilities. In particular, for a sample size of 30 it took 22 iterations whereas in the case of a sample size of 150, it took about twenty five minutes of cpu time.

CHAPTER III

NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATE OF A SURVIVAL FUNCTION IN THE PRESENCE OF CENSORING

3.0 INTRODUCTION

In this chapter, we will consider the nonparametric estimation of a bivariate distribution in the presence of censoring as was presented in a paper by Hanley and Parnes (1983).

New definitions such as homogeneous and heterogeneous censoring which are synonymous to univariate and bivariate censoring respectively are introduced. In Section 1, it is shown how a bivariate empirical survival function must be constructed in order to be considered a (nonparametric) maximum likelihood estimate of the underlying survival function. In Section 2, it is shown that a closed-form solution, similar to the product-limit estimate of Kaplan and Meier, is possible with homogeneous censoring.

3.1 FORMULATION

Let $T = (T_1,T_2)$ represent a bivariate random variable denoting the durations before two events occur. This notation is equivalent to (X^0, Y^0) which was used in the previous chapters. Let $t_1, t_2, \ldots, t_n$ represent n independent realizations of T. We can form the empirical 'survival function' $\hat{F}(t)$ as an estimate of the underlying 'survival' function $F(t) = F(t_1,t_2) = P(T_1 > t_1; T_2 > t_2)$. Let L_{ij}

denote the limit of observation for the j^{th} ($j = 1,2$) of the two events to occur in subject i ($i = 1,\ldots,n$). Thus, the observable data for subject i are contained in the two vectors t_i^* and Z_i, where, for $j = 1,2$, $t_{ij}^* = \min(t_{ij}, L_{ij})$ and $Z_{ij} = 1$ if $t_{ij}^* = t_{ij}$ and 0 otherwise. As in the previous chapters, each t_{ij}^* for which $Z_{ij} = 0$ is called a censored observation. One can represent the data on subject i by noting that t_i belongs to a region or subset R_i of the space of $T \equiv \mathbb{R}^2$. The region R_i will be an elemental rectangle, a horizontal or vertical strip, or an open quadrant, depending on whether $Z_i = (1,1)$, $(0,1)$, $(1,0)$ or $(0,0)$. See Figure 3.1.1.

We will consider choosing, from among all admissible survival functions $F(t)$, one denoted by $\hat{F}(t)$, which maximizes the likelihood of the observed data. Thus for any specified probability distribution $p(t) = d\,F(t)$ on T,

(3.1.1) $$\mathcal{L} \propto \prod_i \int_{R_i} p(t) \equiv \prod_i P_i .$$

Integrals and differentials are being used for both discrete and continuous-type random variables.

In order to maximize $\mathcal{L}$, $\hat{F}$ or equivalently $\hat{P}(t)$ must be constructed as follows.

(i) the entire probability mass must be distributed within $\cup_i R_i$; mass placed outside the data-defined regions R_i will not contribute to any of the terms of $\mathcal{L}$, and will not help to maximize it.

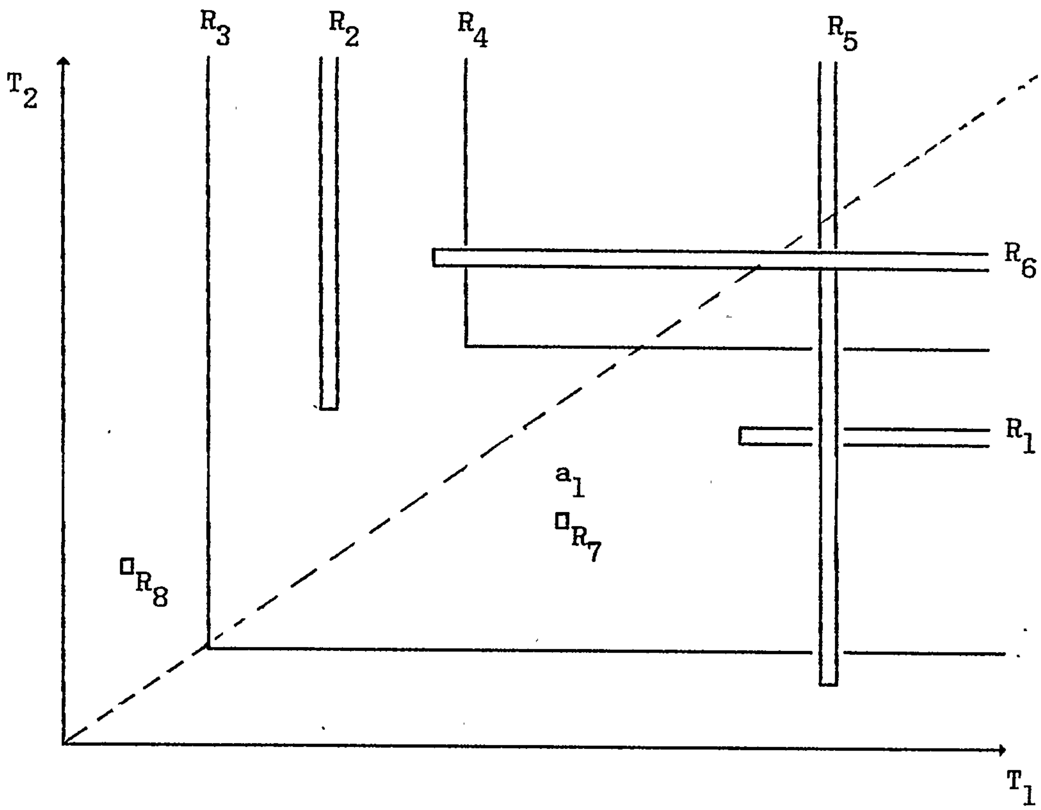

Figure 3.1.1 Data-defined regions corresponding to complete (Regions 7 and 8), half-censored (Regions 1, 2, 5 and 6) and doubly-censored (Regions 3 and 4) observations. Regions 4, 5 and 6 arose from heterogeneous censoring.

(ii) each R_i must receive some probability mass, otherwise $\mathcal{L}$ will vanish;

(iii) if either component of t_i is censored, its contribution $P_i = \int_{R_i} P(t)$ to $\mathcal{L}$ is not affected by how $P(\cdot)$ is distributed within R_i; thus, P_i should be arranged so that it is maximally shared by other regions, R_j, that are contained in, or intersect with, R_i. In this way, the contributions P_j of these other regions will be increased without changing P_i.

Stated in set-theoretic terms, this implies that the total probability mass should be distributed over the maximal intersections $\mathcal{A}_1, \ldots, \mathcal{A}_m$ of the R_i. By a maximal intersection $\mathcal{A}$ we mean nonempty finite intersection of the R_i such that for each i, $\mathcal{A} \cap R_i = \phi$ or $\mathcal{A}$. Some of the maximal intersections will each contain just one point, which is either an observed (uncensored) t or possibly an intersection of two 'half-censored' observations. The single points in these sets form unambiguous support points for $\hat{p}(\cdot)$. In the case where a maximal intersection $\mathcal{A}$ consists of more than a single point, there is no unique choice of specific support points from $\mathcal{A}$, and we can without any loss of generality refer to $\mathcal{A}$ itself or choose a point a from $\mathcal{A}$ to represent it. For simplicity we will write $P(a)$ rather than $P(\mathcal{A})$.

The above guidelines are readily illustrated by the example in Figure 3.1.1. From the n = 8 regions R_i shown, we can contruct a support consisting of m = 5 sets $\mathcal{A}_1, \ldots, \mathcal{A}_5$. The observations which generated R_7 and R_8 must form two of the support points which we arbitrarily label a_1 and a_2, and that the probability mass $p(a_1)$ will contribute to both P_7 and P_3 and thus to $\mathcal{L}$. $a_3 = R_5 \cap R_6$, will contribute to P_4, P_5, P_3 and P_6, while $a_4 = R_5 \cap R_1$, will contribute to P_5, P_3 and P_1. For the fifth component of the support for $\hat{p}(\cdot)$, one can take either the entire region R_2 or any arbitrary point $a_5 \in R_2$; $P(a_5)$ will contribute to both P_3 and P_2.

Once a support set $A = \{a_1, \ldots, a_m\}$ has been chosen for $\hat{p}(\cdot)$, the likelihood $\mathcal{L}$ can be written as

$$\mathcal{L} \propto \prod_i \left\{ \sum_{a_k \in R_i} P(a_k) \right\} . \tag{3.1.2}$$

In the example of figure 3.1.1 (with the m = 5 support points $a_1, \ldots, a_5$ receiving probability masses of $p(a_1) = p_1, \ldots, p(a_5) = p_5$), we have

$$\mathcal{L} \propto P_4\, P_5\, (P_1 + P_3 + P_4 + P_5)\, P_3(P_3 + P_4)\, P_3 P_1 P_2.$$

To determine the magnitudes of $P(a_1), \ldots, P(a_m)$, it is helpful to distinguish two censoring patterns which we will call 'homogeneous' and 'heterogeneous'.

3.2 ESTIMATING F(t) FROM HOMOGENEOUSLY CENSORED DATA.

We call the censoring 'homogeneous' if every two data-defined regions R_i and R_j are either disjoint or nested one within the other. This pattern occurs for example when one follows a subject for equal lengths of time towards each endpoint, i.e. if $L_{i1} = L_{i2}$. Incomplete observations can thus be represented by regions which are either (i) horizontal strips lying entirely to the right of the diagonal $T_1 = T_2$, (ii) vertical strips lying entirely above this diagonal, or (iii) squares which are open to the right and have their lower left corner on the diagonal (Regions 1, 2 and 3, respectively in Figure 3.1.1).

To simplify the presentation, we will introduce additional support points, which will receive zero mass in the ML estimation, but which allow us to speak of a grid of K^2 support points a_{11} to a_{kk}. The figure below is a presentation of the augmentation and relabeling of original support points a_1, a_2, a_3 and a_4 (solid circles) as a rectangular grid (open and solid circles). The Grid is formed from (i) intersections of vertical and horizontal lines through original points and (ii) intersections of vertical and horizontal lines through points where diagonal line crosses the lines through these original points. The augmented set of points is then relabeled a_{11} to a_{66} with the first subscript referring to T_1 and the second to T_2.

With an augmented and relabeled A, and abbreviating $P(a_{rs})$ to P_{rs}, P_i can be written as a sum over a rectangular grid

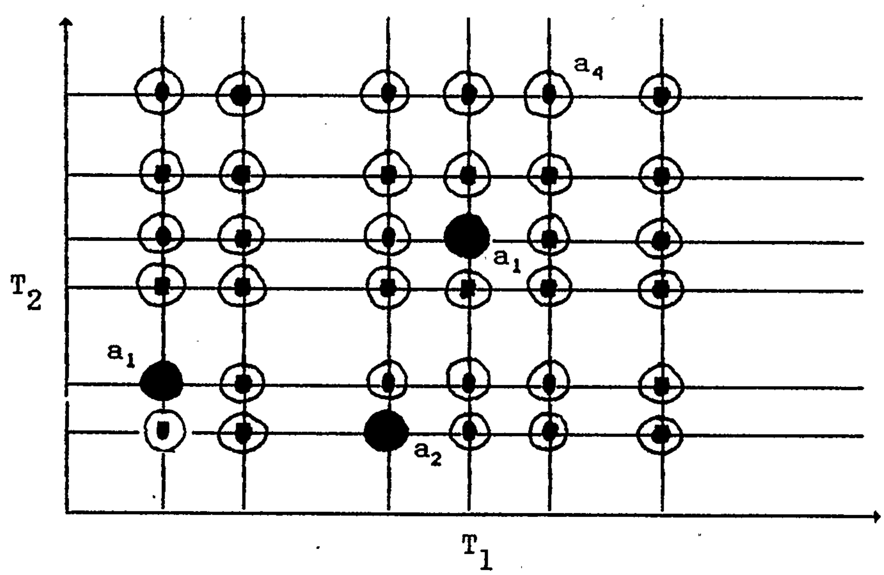

Figure 3.2.1

(3.2.1) $$P_i = \int_{R_i} P = \sum_r \sum_s P_{rs},$$

where the summation index r(or s) runs from the left most (lowest to highest) a_{rs} in R_i.

Letting $X(a)$ and $Y(a)$ refer to the first and second coordinate values of a, the probabilities can be written as:

(3.2.2)
$$\phi_{kk} = P_r\left\{T_1 > X(a_{kk});\ T_2 > Y(a_{kk}) \mid T_1 \ge X(a_{kk});\ T_2 \ge Y(a_{kk})\right\}$$
$$= \sum_{r>k} \sum_{s>k} P_{rs} / \sum_{r\ge k} \sum_{s\ge k} P_{rs},$$

$$\phi_{k.} = P_r\left\{T_1 = X(a_{kk});\ T_2 > Y(a_{kk}) \mid T \ge X(a_{kk});\ T_2 \ge Y(a_{kk})\right\}$$
$$= \sum_{s>k} P_{ks} / \sum_{r\ge k} \sum_{s\ge k} P_{rs},$$

$$\phi_{.k} = P_r\left\{T_1 > X(a_{kk});\ T_2 = Y(a_{kk}) \mid T_1 \ge X(a_{kk});\ T_2 \ge Y(a_{kk})\right\}$$
$$= \sum_{r>k} P_{rk} / \sum_{r\ge k} \sum_{s\ge k} P_{rs},$$

in all the above $1 \le k \le K-1$.

(3.2.3) $$\Psi_{k\ell} = P_r\left\{T_2 > Y(a_{k\ell}) \mid T_1 = X(a_{k\ell});\ T_2 \ge Y(a_{k\ell})\right\}$$
$$= \sum_{s>\ell} P_{ks} / \sum_{s\ge\ell} P_{ks}, \qquad 1 \le k < \ell \le K-1,$$

(3.2.4) $$\Psi_{k\ell} = \sum_{r>\ell} P_{r\ell} / \sum_{r\ge k} P_{r\ell}, \qquad 1 \le \ell < k \le K-1.$$

Each P_i can be written as a product of ϕ and Ψ terms. Thus, we have

$$\mathcal{L} = \prod_k \left\{ \phi_{kk}^{N_{kk}} \, \phi_{k.}^{n_{k.}} \, \phi_{.k}^{n_{.k}} \left[1 - \phi_{kk} - \phi_{k.} - \phi_{.k}\right]^{n_{kk}} \right. \tag{3.2.5}$$

$$\left. \left\{ \prod_{\ell>k} \Psi_{k\ell}^{m_{k\ell}} \left[1 - \Psi_{k\ell}\right]^{n_{k\ell}} \right\} \left\{ \prod_{k>\ell} \Psi_{k\ell}^{m_{k\ell}} \left[1 - \Psi_{k\ell}\right]^{n_{k\ell}} \right\} \right\},$$

where, as is depicted in Figure 3.2.2 the exponents refer to the following counts of sample members:

$n_{k\ell}$, those where the two events occured at $X(a_{k\ell})$ and $Y(a_{k\ell})$;

N_{kk}, those proceeding through a_{kk} without either event taking place;

$n_{k.}$, those where the 'X' events occurs at $X(a_{kk})$ but the other occurs after $Y(a_{kk})$;

$n_{.k}$, the converse of $n_{k.}$;

$m_{k\ell}(\ell > k)$, those where, the 'X' event already having taken place at $X(a_{kk})$, the 'Y' component proceeds through $Y(a_{k\ell})$;

$m_{k\ell}(\ell < k)$, the converse.

Let $R_k = N_{kk} + n_{k.} + n_{.k} + n_{kk}$
and $R_{k\ell} = m_{k\ell} + n_{k\ell}$, then

the ML estimates of the ϕ_{kk}, $\phi_{k.}$ and $\phi_{.k}$ are simply the proportions N_{kk}/R_k, $n_{k.}/R_k$ and $n_{.k}/R_k$, respectively. Similarly, $\hat{\Psi}_{k\ell} = m_{k\ell}/R_{k\ell}$.

The estimates of P_{rs} are obtained from (3.2.2), (3.2.3) and (3.2.4),

$$(3.2.6)\qquad \hat{P}_{rs} = \begin{cases} \left[\prod_{k=1}^{r-1} \hat{\phi}_{kk}\right] \left[1 - \hat{\phi}_{rr} - \hat{\phi}_{r.} - \hat{\phi}_{.r}\right], & s = r \\ \left[\prod_{k=1}^{r-1} \hat{\phi}_{kk}\right] \hat{\phi}_{r.} \left[\prod_{\ell=r+1}^{s-1} \hat{\Psi}_{r\ell}\right] \left[1 - \hat{\Psi}_{rs}\right], & s > r \\ \left[\prod_{k=1}^{s-1} \hat{\phi}_{kk}\right] \hat{\phi}_{.s} \left[\prod_{k=s+1}^{r-1} \hat{\Psi}_{ks}\right] \left[1 - \hat{\Psi}_{rs}\right], & r > s. \end{cases}$$

The ML estimate of $F(t_1,t_2)$ can be obtained by summing the $\hat{P}_{rs}$ in the open rectangle $(t_1,\infty) \times (t_2,\infty)$.

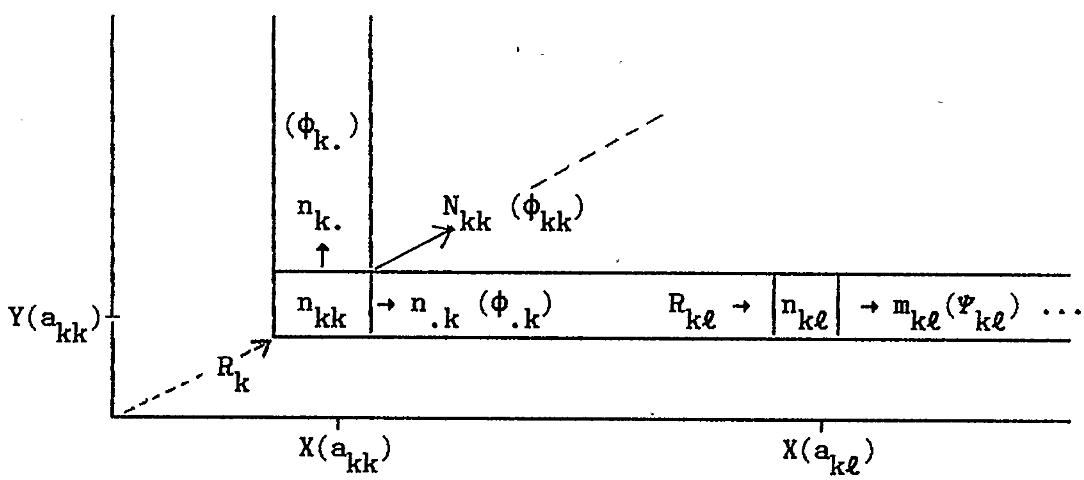

Figure 3.2.2

If T_1 and T_2 are continuous-type variables, then $n_{kk} + n_{.k} + n_{k.} \le 1$; thus at least two of the corresponding ϕ terms will be estimated as zero. If $n_{k.} = 1$, then the $m_{k\ell}(\ell = k+1, \ldots)$ will each be unity until the other event takes place somewhere beyond $Y(a_{kk})$, after which they will equal zero. If the observation time runs out before this second event occurs, the Ψ parameters beyond the last Y observation on this subject cannot be uniquely estimated. This nonuniqueness is similar to the problem that occurs in the univariate Kaplan-Meier (1958) survival curve when the largest observation is censored. In the multivariate case it occurs each time a pair of values, recorded on a continuous scale is 'half-censored', and means that one cannot supply a unique estimate for $F(t_1,t_2)$ when the region $(T_1 > t_1;\ T_2 > t_2)$ contains such observations. This shortcoming can be lessened by discretizing or grouping the data into intervals, as is commonly done in univariate life-tables, so that there are fewer ϕ and Ψ parameters to be estimated, and from larger, more stable denominators. In fact, in many studies subjects are followed up on a fixed schedule so that t actually takes discrete values.

CHAPTER IV

PATH DEPENDENT ESTIMATORS FOR THE BIVARIATE SURVIVAL FUNCTION

4.0 INTRODUCTION

The bivariate estimation problem with discrete times of death or losses was considered in Chapter I, using an extension of the self-consistent approach of Efron. A self-consistent approach for the continuous case was treated in Chapter II. In the previous Chapter, we treated the maximum likelihood approaches to the bivariate estimation problem. In contrast to the iterative estimators of the earlier chapters, we will consider several closed-form estimators for the bivariate model and prove strong uniform consistency to the true bivariate distribution of the lifetimes as was presented in a paper by Campbell and Földes (1980).

Two path-dependent estimators are introduced in Section 1. Each is the product of two one-dimensional Kaplan-Meier product limit estimators. A hazard function approach is employed in Section 2 to estimate $-\ell n\ F(s,t)$ and hence $F(s,t)$. Two path-dependent estimators of $-\ell n\ F(s,t)$ are proposed and these lead to estimators of the bivariate distribution function.

4.1 TWO PATH-DEPENDENT PRODUCT-LIMIT ESTIMATORS

We will adhere to the notation established in the main introduction of this thesis unless otherwise stated. The functions; $F(s,t)$, $G(s,t)$ and $H(s,t)$ carry with them the same interpretations as before.

$$F(s,t) = P(X^0 > s,\ Y^0 > t),$$

$$G(s,t) = P(C > s,\ D > t),$$

$$H(s,t) = P(X > s,\ Y > t).$$

Since $\left\{X_i^0,\ Y_i^0\right\}_{i=1}^{\infty}$ and $\left\{C_i,\ D_i\right\}_{i=1}^{\infty}$ are mutually independent, we have

$$H(s,t) = F(s,t)\ G(s,t). \tag{4.1.1}$$

Based on the elementary observation

$$F(s,t) = F(s,0)\ F_1(t|s), \tag{4.1.2}$$

where $F_1(t|s) = P(Y^0 > t|X^0 > s)$.

The survival function $F(s,t)$ is thus estimated by separately estimating each of the two terms on the right of (4.1.2). This leads to an estimator $\hat{F}_1(s,t)$ based on the path from (0,0) to (s,t) which is linear from (0,0) to (s,0) and linear from (s,0) to (s,t).

Let us now define the following notations which we shall be using:

$$N_n(s,t) = N(s,t) = \sum_{i=1}^{n} I_{\{X_i>s,\ Y_i>t\}}, \tag{4.1.3}$$

$$\alpha_i(s,t) = I_{\{X_i\le s,\ Y_i>t,\ \delta_{1i}=1\}} \qquad (i = 1,\ldots,n), \tag{4.1.4}$$

$$\beta_j(s,t) = I_{\{X_j>s,\ Y_j\le t,\ \delta_{2j}=1\}} \qquad (j = 1,\ldots,n). \tag{4.1.5}$$

To estimate $F(s,0)$, we will calculate the Kaplan-Meier product limit estimator of $F(s,0)$ using the one-dimensional censored sample $\left\{X_i, \xi_{1i}\right\}_{i=1}^{n}$. This produces the estimator

$$\hat{F}_{1n}(s,0) = \begin{cases} \prod_{i=1}^{n} \left[\frac{N(X_i,0)}{N(X_i,0)+1}\right]^{\alpha_i(s,0)} & \text{if } s \le \tau_{1n} \\ 0, & \text{otherwise} \end{cases}$$

where $\tau_{1n} = \max_{1\le i\le n} \{X_i\}$.

(The one-dimensional convention that the last observation is converted to a death (if it is censored) is adhered to here.) To estimate $F_1(t|s)$, the second term of (4.1.2), project all points for which $X_i > s$ horizontally to the line $X = s$, and ignoring the (X_i, ξ_{1i}) values calculate the Kaplan-Meier product-limit estimator based on the data $\left\{Y_j, \xi_{2j}\right\}_{j=1}^{n}$ for which $X_i > s$ (see Figure 4.1.2). The estimate obtained by this method is $P(Y^0 > t|X > s)$ but

$$P(Y^0 > t|X > s) = P(Y^0 > t|X^0 > s, C > s)$$

$$= \frac{P(Y^0 > t, X^0 > s, C > s)}{P(X^0 > s, C > s)}$$

$$= \frac{P(Y^0 > t, X^0 > s)\, P(C > s)}{P(X^0 > s)\, P(C > s)}$$

$$= \frac{P(Y^0 > t,\ X^0 > s)}{P(X^0 > s)}$$

$$= P(Y^0 > t \mid X^0 > s), \tag{4.1.6}$$

since C is independent of the pair (X^0, Y^0). Thus the following estimator of $F_1(t|s)$ is obtained:

$$\hat{F}_{1n}(t|s) = \begin{cases} \prod_{j=1}^{n} \left[\frac{N(s,Y_j)}{N(s,Y_j)+1}\right]^{\beta_j(s,t)} & \text{if } t \le \tau_{2n}(s) \\ 0, & \text{otherwise,} \end{cases} \tag{4.1.7}$$

where $\tau_{2n}(s) = \max_{i \le i \le n} \left\{Y_i:\ X_i > s\right\}$.

Consequently, the estimator for $F(s,t)$ is given by

$$\hat{F}_{1n}(s,t) = \begin{cases} \prod_{i=1}^{n} \left[\frac{N(X_i,0)}{N(X_i,0)+1}\right]^{\alpha_i(s,0)} \prod_{j=1}^{n} \left[\frac{N(s,Y_j)}{N(s,Y_j)+1}\right]^{\beta_j(s,t)} & \text{if } N(s,t)>0 \\ 0, & \text{otherwise.} \end{cases} \tag{4.1.8}$$

By changing the role of s and t it is possible to develop our estimator $\hat{F}_{2n}(s,t)$ based on the relation

$$F(s,t) = F(0,t)\ F_2(s|t) = F(0,t)\ P(X^0 > s \mid Y^0 > t). \tag{4.1.9}$$

Using the linear path from (0,0) to (0,t) and (0,t) to (s,t). The corresponding estimator is

$$(4.1.10)\quad \hat{F}_{2n}(s,t) = \begin{cases} \prod_{j=1}^{n}\left[\frac{N(0,Y_j)}{N(0,Y_j)+1}\right]^{\beta_j(0,t)} \prod_{i=1}^{n}\left[\frac{N(X_i,t)}{N(X_i,t)+1}\right]^{\alpha_i(s,t)} & \text{if } N(s,t)>0 \\ 0, & \text{otherwise.} \end{cases}$$

The question arises as to whether $\hat{F}_{1n}(s,t)$ and $\hat{F}_{2n}(s,t)$ are necessarily distribution functions. We shall illustrate with the example of Figure 4.1.1 that the above estimators do not satisfy the monotonicity requirements of a distribution function.

Figure 4.1.1 consists of four points. ℓ and d denote loss and death respectively. Therefore, (d,ℓ) at (x_3,y_3) denotes a point which is a death in the first corindate and censored in the second. At any point in the rectangle $[0, x_4] \times [0, y_4]$ the estimator $\hat{F}_{14}$ can be calculated.

$$\xi_{11} = 0 \qquad \xi_{21} = 0$$
$$\xi_{12} = 1 \qquad \xi_{22} = 1$$
$$\xi_{13} = 1 \qquad \xi_{23} = 0$$
$$\xi_{14} = 0 \qquad \xi_{24} = 1$$

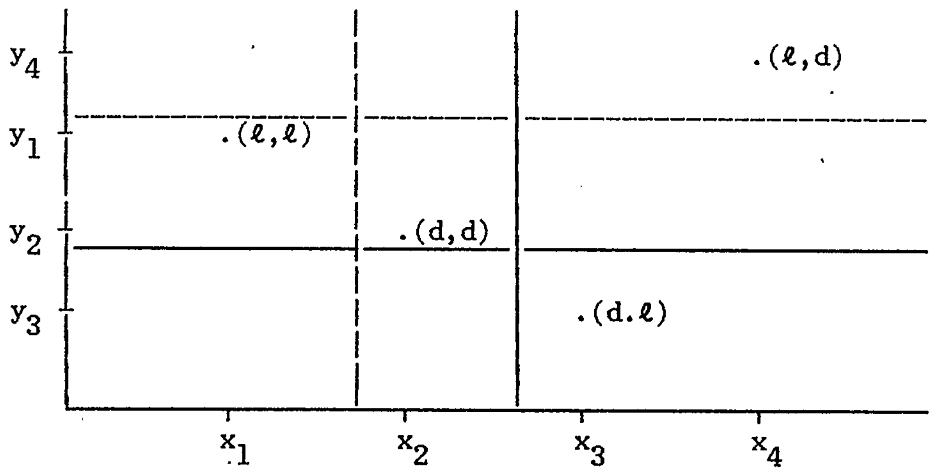

Figure 4.1.1

d = death

ℓ = loss

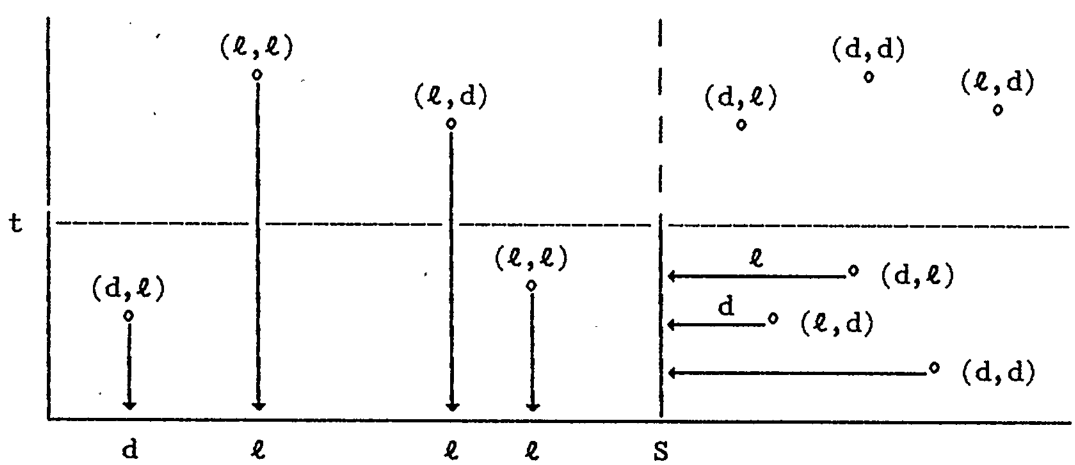

Figure 4.1.2

The convention in the one-dimensional Kaplan-Meier estimator that the final loss is converted to a death is adhered to.

Suppose we wish to compute $\hat{F}_1(s,t)$ where $s \in (x_2,x_3)$ and $t \in (y_3,y_2)$. Then from 4.1.8 we have

$$\hat{F}_{14}(s,t) = \left[\frac{N(x_1,0)}{N(x_1,0)+1}\right]^{\xi_{11}} \left[\frac{N(x_2,0)}{N(x_2,0)+1}\right]^{\xi_{12}} \left[\frac{N(s,y_3)}{N(s,y_3)+1}\right]^{\xi_{23}}$$

$s \in (x_2,x_3)$
$t \in (y_3,y_2)$

$$= \left[\frac{3}{4}\right]^0 \left[\frac{2}{3}\right]^1 \left[\frac{1}{2}\right]^0$$

$$= \frac{2}{3} .$$

Similarly $\hat{F}_1(s,t)$ where $S \in (x_1,x_2)$ and $t \in (y_1,y_4)$ is given by

$$\hat{F}_{14}(s,t) = \left[\frac{N(x_1,0)}{N(x_1,0)+1}\right]^{\xi_{11}} \left[\frac{N(s,y_3)}{N(s,y_3)+1}\right]^{\xi_{23}} \left[\frac{N(s,y_2)}{N(s,y_2)+1}\right]^{\xi_{22}}$$

$s \in (x_1,x_2)$
$t \in (y_1,y_4)$

$$= \left[\frac{3}{4}\right]^0 \left[\frac{2}{3}\right]^0 \left[\frac{1}{2}\right]^1$$

$$= \frac{1}{2} .$$

The entire estimator $\hat{F}_{14}$ can be calculated for various points $(s,t) \in [0,x_4] \times [0,y_4]$. The results of these calculations are displayed in Figure 4.1.3) below.

	0	0	0	0	0
y_4	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$	0
y_1	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$	0
y_2	1	1	$\frac{2}{3}$	$\frac{1}{3}$	0
y_3	1	1	$\frac{2}{3}$	$\frac{1}{3}$	0
	x_1	x_2	x_3	x_4	

Figure 4.1.3

From the above calculations one can conclude that $\hat{F}_{14}$ does not satisfy the monotonicity requirement of a distribution function.

The estimator $\hat{F}_{24}$ which is based on the alternate path (4.1.10) can also be calculated for any point $(s,t) \in [0,x_4] \times [0,y_4]$.

Figure (4.1.4) below is a presentation of the estimator $\hat{F}_{24}$ for the example of Figure 4.1.1.

	0	0	0	0	0
y_4	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	0
y_1	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	0
y_2	1	1	$\frac{1}{2}$	$\frac{1}{2}$	0
y_3	1	1	$\frac{2}{3}$	$\frac{1}{3}$	0
	x_1	x_2	x_3	x_4	

Figure 4.1.4

4.2 ESTIMATORS BASED ON THE BIVARIATE HAZARD FUNCTION

The multivariate hazard approach of Marshall is employed to develop bivariate survival function estimators based on the hazard function. Define the hazard function $R(s,t)$ as

$$R(s,t) = -\log F(s,t). \tag{4.2.1}$$

Assume that R is absolutely continuous with partial derivatives that exist almost everywhere.

Let,

$$\gamma(z) = (\gamma_1(z),\ \gamma_2(z)), \tag{4.2.2}$$

where

$$\gamma_1(z) = \frac{\partial}{\partial z_1} R(z) \quad \text{and} \quad \gamma_2(z) = \frac{\partial}{\partial z_2} R(z). \tag{4.2.3}$$

$\gamma(z)$ thus denotes the gradient of $R(z)$ for the point $z = (z_1,z_2)$ $R(s,t)$ can be constructed as the path integral of $\gamma(z)$ from (0,0) to (s,t). By path independence one can write

$$R(s,t) = \int_{(0,0)}^{(s,t)} \gamma(z)\, dz. \tag{4.2.4}$$

If the path linear from (0,0) to (s,0) and linear from (s,0) to (s,t) is considered, then

$$R(s,t) = \int_0^s \gamma_1(u,0)\,du + \int_0^t \gamma_2(s,v)\,dv \tag{4.2.5}$$

By (4.2.1) and (4.2.2) and from (4.2.5) we have

$$R(s,t) = -\log F(s,t)$$

$$= -\int_0^s \frac{1}{F(u,0)}\,d_u\,F(u,0) - \int_0^t \frac{1}{F(s,v)}\,d_v\,F(s,v), \tag{4.2.6}$$

where $d_u\,F(u,t)$ denotes Lebesgue-Stieltjes integration over u for t fixed.

Using (4.1.1) we have,

$$-\log F(s,t) = -\int_0^s \frac{G(u,0)}{H(u,0)}\,\frac{\partial F(u,0)}{\partial u}\,du - \int_0^t \frac{G(s,v)}{H(s,v)}\,\frac{\partial F(s,v)}{\partial v}\,dv. \tag{4.2.7}$$

Introduce the following functions:

$$\tilde{K}(s,t) = \int_0^s G(u,t)\,\frac{\partial P(X^0 \le u,\ Y^0 > t)}{\partial u}\,du; \tag{4.2.8}$$

$$\tilde{L}(s,t) = \int_0^t G(s,v)\,\frac{\partial P(X^0 > s,\ Y^0 \le v)}{\partial v}\,dv. \tag{4.2.9}$$

But,

$$P(X^0 \le u,\ Y^0 > t) = P(Y^0 > t) - P(X^0 > u,\ Y^0 > t)$$

which implies

$$\frac{\partial P(X^0 \le u,\ Y^0 > t)}{\partial u} = -\frac{\partial P(X^0 > u,\ Y^0 > t)}{\partial u}\ . \tag{4.2.10}$$

Similarly we have,

$$\frac{\partial P(X^0 > s,\ Y^0 \le v)}{\partial v} = -\frac{\partial P(X^0 > s,\ Y^0 > v)}{\partial v}\ . \tag{4.2.11}$$

Applying (4.2.8), (4.2.9), (4.2.10) and (4.2.11) to (4.2.7) we have

$$-\log F(s,t) = \int_0^s \frac{1}{H(u,0)}\, d_u\, \tilde{K}(u,0) + \int_0^t \frac{1}{H(s,t)}\, d_v\, \tilde{L}(s,v). \tag{4.2.12}$$

We can thus estimate H, $\tilde{K}$ and $\tilde{L}$ first. An estimator of $H(u,v)$ is the empirical survival function:

$$H_n(s,t) = \frac{1}{n}\, \Sigma\, I_{\{X_i > s,\, Y_i > t\}} = \frac{N(s,t)}{n}\ . \tag{4.2.13}$$

It can also be shown that,

$$\tilde{K}(s,t) = E[\alpha_i(s,t)].$$

We know that

$$E[\alpha_i(s,t)] = P(C > X^0,\ D > t,\ X^0 \le s,\ Y^0 > t)$$

(because $Y_i > t \Leftrightarrow D_i > t,\ Y_i^0 > t$)

($X_i \le s,\ \xi_{1i} = 1 \Leftrightarrow X_i^0 \le s,\ C_i > X^0$)

$$= \int_{v=t}^{\infty} \int_{u=0}^{s} P(C>X^0,\ D>t,\ X^0 \le s,\ Y^0>t \mid X^0=u,\ Y^0=v)\ d\ F(u,v)$$

$$= \int_{v=t}^{\infty} \int_{u=0}^{s} P(C>u,\ D>t \mid X^0=u,\ Y^0=v)\ d\ F(u,v)$$

$$= \int_{v=t}^{\infty} \int_{u=0}^{s} P(C>u,\ D>t)\ d\ F(u,v)$$

(because (C,D) and (X^0,Y^0) are mutually independent)

$$= \int_{v=t}^{\infty} \int_{u=0}^{s} G(u,t)\ d\ F(u,v)$$

$$= \int_{u=0}^{s} G(u,t) \int_{v=t}^{\infty} d\ F(u,v)$$

$$= - \int_{u=0}^{s} G(u,t)\ d_u\ F(u,t)$$

$$= - \int_{u=0}^{s} G(u,t)\ d_u\ P(X^0 > u,\ Y^0 > t)$$

$$= - \int_{u=0}^{s} G(u,t)\, d_u\, [P(Y^0 > t) - P(X^0 \le u,\ Y^0 > t)]$$

$$= \int_{u=0}^{s} G(u,t)\, d_u\, P(X^0 \le u,\ Y^0 > t)$$

$$= \tilde{K}(s,t)$$

$$\therefore\ \tilde{K}(s,t) = E[\alpha_i(s,t)].$$

Similarly we have,

$$\tilde{L}(s,t) = E(\beta_j(s,t)). \tag{4.2.15}$$

Hence the natural estimators of $\tilde{K}(s,t)$ and $\tilde{L}(s,t)$ are

$$\tilde{K}_n(s,t) \simeq \frac{1}{n} \sum_{i=1}^{n} \alpha_i(s,t) \quad \text{and} \tag{4.2.16}$$

$$\tilde{L}_n(s,t) \simeq \frac{1}{n} \sum_{i=1}^{n} \beta_j(s,t). \tag{4.2.17}$$

Consequently the estimate of $R(s,t) = -\log F(s,t)$ is given by

$$(4.2.18) \qquad \tilde{R}_{1n}(s,t) = \int_0^s \frac{1}{H_n(u,0)} d_u \tilde{K}_n(u,0) + \int_0^t \frac{1}{\overline{H}_n(s,v)} d_v \tilde{L}_n(s,v)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{\alpha_i(s,0)}{H_n(X_i,0)} + \frac{1}{n} \sum_{j=1}^{n} \frac{\beta_j(s,t)}{H_n(s,Y_j)} ,$$

if $N(s,t) > 0$ and $\tilde{R}_1(s,t) = +\infty$ otherwise.

Hence

$$(4.2.19) \qquad \tilde{F}_{1n}(s,t) = \exp\left[- \tilde{R}_{1n}(s,t)\right].$$

Similarly, an estimate of $R(s,t)$ is given by

$$(4.2.20) \qquad \tilde{R}_{2n}(s,t) = \frac{1}{n} \Sigma \frac{\beta_j(0,t)}{H_n(0,Y_i)} + \frac{1}{n} \Sigma \frac{\alpha_i(s,t)}{H_n(X_i,t)},$$

if $N(s,t) > 0$ and $\tilde{R}_{2n}(s,t) = +\infty$ otherwise.

This result in another estimator of $F(s,t)$, which is given by

$$(4.2.21) \qquad \tilde{F}_{2n}(s,t) = \exp\left[- \tilde{R}_{2n}(s,t)\right].$$

4.3 RELATIONSHIP OF THE PRODUCT-LIMIT AND THE HAZARD FUNCTION ESTIMATORS

Lemma 4.3.1

$$(4.3.1) \qquad \sup_{\substack{0 \le s < \infty \\ 0 \le t < \infty}} |H_n(s,t) - H(s,t)| = 0\left[\sqrt{\frac{\log \log n}{n}}\right] \text{ a.s.}$$

$$(4.3.2) \qquad \sup_{\substack{0 \le s < \infty \\ 0 \le t < \infty}} |\tilde{K}_n(s,t) - \tilde{K}(s,t)| = 0\left[\sqrt{\frac{\log \log n}{n}}\right] \text{ a.s.}$$

$$(4.3.3) \qquad \sup_{\substack{0 \le s < \infty \\ 0 \le t < \infty}} |\tilde{L}_n(s,t) - \tilde{L}(s,t)| = 0\left[\sqrt{\frac{\log \log n}{n}}\right] \text{ a.s.}$$

Proof:

From the multi-dimensional law of the iterated logarithm for empirical distributions of Kiefer, we have

$$P_F\left\{\lim_{n\to\infty} \sup_{\substack{0 \le s < \infty \\ 0 \le t < \infty}} n^{1/2}|H_n(s,t) - H(s,t)|/\left[2^{-1} \log \log n\right]^{1/2} = 1\right\} = 1.$$

Hence, result (4.3.1) follows immediately:

$$\sup_{\substack{0 \le s < \infty \\ 0 \le t < \infty}} |H_n(s,t) - H(s,t)| = 0\left[\sqrt{\frac{\log \log n}{n}}\right] \text{ a.s.}$$

To prove (4.3.2) it is enough to observe that

$$(4.3.4) \qquad \begin{aligned} \tilde{K}(s,t) &= P(X^0 \le s,\ X^0 \le C,\ Y > t) \\ &= P(X^0 \le s,\ X^0 - C \le 0) - P(X^0 \le s,\ X^0 - C \le 0,\ Y \le t). \end{aligned}$$

Therefore $\tilde{K}_n(s,t)$ can be considered as the difference of two empirical distributions:

$$\tilde{K}_n(s,t) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i^0 \le s,\ X_i^0 - C \le 0\}} - \frac{1}{n} \sum_{j=1}^{n} I_{\{X_j^0 \le s,\ X_j^0 - C_j \le 0,\ Y_j \le t\}}.$$

This means that applying again Kiefer's result (once in two –, once in three-dimensions) (4.3.2) is obtained. A similar argument proves (4.3.3).

4.4 CONSISTENCY

The estimator $\hat{F}_{1n}$ was constructed as a product of two one-dimensional Kaplan-Meier estimators. It therefore follows that $\hat{F}_{1n}$ is pointwise consistent.

The pointwise consistency remains true in the case of not necessarily continuous functions F and G, as one can develop, using the same projecting argument the corresponding bivariate estimator as the product of two one-dimensional Kaplan-Meier estimators. The continuity conditions on F and G are not required for this pointwise consistency in that the Kaplan-Meier estimate is consistent.

The estimators developed in this section are fairly simple to compute in the sense that they depend on relatively few estimable functions. However, they are path dependent and may fail to be survival functions. The technique employed in their development can be generalized from two dimensions to higher dimensions. The obvious difference is that the number of possible paths (and hence the estimators) increases from 2 to 2^{k-1} for the k-dimensional analog.

CHAPTER V

ESTIMATION OF A BIVARIATE DISTRIBUTION UNDER RANDOM CENSORSHIP VIA A SUBDISTRIBUTION

5.0 INTRODUCTION

In the previous chapter, we considered two estimators by Campbell and Földes (1982). Both estimators were shown to be uniformly consistent at a rate of convergence equal to that of the empirical distribution function. However, it was pointed out that both estimators need not be survival functions. That is, they do not satisfy the monotonicity requirements of a distribution function. In Section 1, we suitably modify the Campbell and Földes (1982) estimators to satisfy the important monotonicity requirements and to achieve their desirable rate of consistency as was presented in a paper by Burke (1988). In Section 2, a hazard gradient approach identical to that of Chapter IV is further employed to arrive at additional estimators. Section 3 contains a discussion of the multidimensional case.

5.1 THE SUBDISTRIBUTION FUNCTION AND THE ESTIMATION PROBLEM

We will adhere to the notation established in the previous chapters unless otherwise stated. $F^d(x,y)$ denotes the distribution function of (X^0, Y^0).

$$\text{i.e.}\quad F^d(x,y) = P(X^0 \le x,\ Y^0 \le y).$$

Let $G(x,y) = P(C > x, D > y)$ denote the survival function of the censoring variables. Then, the vectors from the random sample (X_i, Y_i) $(i = 1,\dots,n)$ which are uncensored in both coordinates have subdistribution function

$$\tilde{F}^d(x,y) = P(X_i \le x,\ Y_i \le y,\ \delta_{1i} = \delta_{2i} = 1)$$

$$= \int_{-\infty}^{y} \int_{-\infty}^{x} G(u,v)\ d\ F^d(u,v),$$

which implies

$$\frac{d\ \tilde{F}^d(x,y)}{d\ F^d(x,y)} = G(x,y),$$

$$d\ F^d(x,y) = \frac{1}{G(x,y)}\ d\ \tilde{F}^d(x,y)$$

and

$$F^d(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} \frac{1}{G(u,v)}\ d\ \tilde{F}^d(u,v).$$

The subdistribution function $\tilde{F}^d$ can be estimated by

$$\tilde{F}^d_n(x,y) = n^{-1} \sum_{i=1}^{n} \delta_{1i}\ \delta_{2i}\ I(X_i \le x,\ Y_i \le y),$$

where $I(A)$ denotes the indicator function of the event A.

We then estimate the survival function G of the censoring vectors by either the Campbell-Földes estimators, $\hat{G}_{1n}$ or $\hat{G}_{2n}$ as defined by (5.1.2) and (5.2.1) below and arrive at two estimators of F^d.

(5.1.1)
$$\hat{F}^d_{jn}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} \{\hat{G}_{jn}(u,v)\}^{-1}\, d\,\tilde{F}^d_n(u,v)$$
$$= n^{-1}\sum_{i=1}^{n} \xi_{1i}\,\xi_{2i}\,\{\hat{G}_{jn}(X_i,Y_i)\}^{-1}\, I(X_i \le x,\ Y_i \le y).$$

Even though $\hat{F}^d_{jn}$ is simple to compute, in the case of heavy censoring, it will have fewer support points than estimator (6.3.2). $\hat{F}^d_{jn}$ is clearly monotone nondecreasing in both variables because it can be expressed as the integral of a positive function with respect to a monotone nondecreasing one. In particular, over the rectangle $R = (x_1,x_2] \times (y_1,y_2]$, where $(x_1,y_1) \le (x_2,y_2)$

$$\hat{F}^d_{jn}(R) = \int_{y_1}^{y_2}\int_{x_1}^{x_2} \{\hat{G}_{jn}(u,v)\}^{-1}\, d\,\tilde{F}^d_n(u,v),$$

which is nonnegative and hence $\hat{F}_{jn}$ has the monotonicity requirements of a distribution function.

Let

$$\alpha_i(s,t) = I(X_i \le s,\ Y_i > t,\ \xi_{1i} = 0),$$

$$\beta_j(s,t) = I(X_j > s,\ Y_j \le t,\ \xi_{2j} = 0),$$

$$\tau_{1n} = \max\{X_i\},$$

$$\tau_{2n}(s) = \max\{Y_i:\ \ X_i > s\} \quad \text{over} \quad i = 1,\ldots,n$$

where $\qquad i,j = 1,\ldots,n.$

The first Campbell-Földes estimator is

$$\hat{G}_{1n}(x,y) = \hat{G}_n'(x,0)\ \hat{G}_n'(y|x) \tag{5.1.2}$$

where

$$\hat{G}_n'(x,0) = \begin{cases} \displaystyle\prod_{i=1}^{n} \left\{\frac{N(X_i,0)}{N(X_i,0)+1}\right\}^{\alpha_i(x,0)} & (x \le \tau_{1n}) \\ 0, & \text{otherwise} \end{cases}$$

is the univariate Kaplan and Meier (1958) estimator, and

$$\hat{G}'_n(y|x) = \begin{cases} \prod_{i=1}^{n} \left\{ \frac{N(x,Y_i)}{N(x,Y_j)+1} \right\}^{\beta_j(x,y)} & (y \le \tau_{2n}(x)) \\ \\ 0, & \text{otherwise.} \end{cases}$$

Here $N(s,t) = \sum_{i=1}^{n} I(X_i > s,\ Y_i > t)$.

Substituting the above into (5.1.1) we arrive at an estimator for F^d,

$$\text{(5.1.3)} \quad \hat{F}^d_{1n}\ (x,y) = n^{-1} \sum_{i=1}^{n} \xi_{1i} \xi_{2i} \{\hat{G}'_{1n}(X_i 0)\ \hat{G}'_{1n}(Y_i|X_i)\}^{-1}\ I(X_i \le x,\ Y_i \le y).$$

5.2 THE HAZARD GRADIENT APPROACH

The second Campbell-Földes (1982) estimator is given by

$$\text{(5.2.1)} \qquad \hat{G}_{2n}(x,y) = \exp\{-\ R_n(x,y)\},$$

where,

$$R_n(x,y) = \begin{cases} \sum_{i=1}^{n} \frac{\alpha_i(x,0)}{N(x_i,0)} + \sum_{j=1}^{n} \frac{\beta_j(x,y)}{N(x,y_j)}, & (N(x,y) > 0) \\ \\ \infty, & \text{otherwise.} \end{cases}$$

Substituting the above into (5.1.1) we arrive at another estimator for F^d:

(5.2.2) $$\hat{F}^{d}_{2n}(x,y) = n^{-1} \sum_{i=1}^{n} \xi_{1i}\xi_{2i}\{\exp[-R_n(X_i,Y_i)]\}^{-1}\, I(X_i \le x,\ Y_i \le y).$$

5.3 CONVERSION TO SURVIVAL FUNCTION AND THE MULTIDIMENSIONAL CASE

Estimators for the survival function $F(s,t)$, can be defined as

(5.3.1) $$\begin{aligned}\hat{F}_{jn}(s,t) &= P(X^0 > s,\ Y^0 > t) \\ &= 1 - \hat{F}^{d}_{jn}(s,\infty) - \hat{F}^{d}_{jn}(\infty,t) + \hat{F}^{d}_{jn}(s,t), \qquad (j = 1,2).\end{aligned}$$

These estimators inherit the corresponding properties of $\hat{F}^{d}_{jn}$.

While one may define an estimator $\hat{F}^{*}_{jn}$ of F as

$$\hat{F}^{*}_{jn}(s,t) = \int_t^{\infty}\int_s^{\infty} \{\hat{G}_{jn}(u,v)\}^{-1}\, d\, \tilde{F}^{d}_{n}(u,v)$$

and obtain similar consistency results, its behaviour for large (s,t) is not as satisfactory as (5.3.1). In particular, if our largest observations were of the form (X^0,D), (C,Y^0) or (C,D), $F^{*}_{jn}(s,t) = 0$ at these points, which neglects the fact that (X^0,Y^0) is larger. The estimator $\hat{F}_{jn}$ of (5.3.1) would not be zero at these points.

The multidimensional case is straightforward since the distribution function of the survival times of interest, $(X^0_{i1},\ldots,X^0_{ik})$ for $i = 1,\ldots,n$, can be written as

$$F^d(x) = \int_{D_x} \{G(u)\}^{-1} \, d\, \tilde{F}^d(u) \qquad (x \in R^k),$$

where

$$x = (x_1,\ldots,x_k) \in R^k,$$

$$D_x = \{u = (u_1,\ldots,u_k) : u_j \le x_j\},$$

$G(u) = G(u_1,\ldots,u_k)$ is the survival function of the censoring variables $(C_{i1},\ldots,C_{ik})$, for $i = 1,\ldots,n$, and

$$\tilde{F}^d(u) = P_r\left\{\min(X^o_{ij}, C_{ij}) \le u,\ X^o_{ij} \le C_{ij};\quad j = 1,\ldots,k\right\}.$$

As mentioned earlier, the weakness of the estimators developed in this chapter is that in the case of heavy censoring, they have fewer support points than their other competitors. However, $\hat{F}_{jn}$ is simple to compute since $\hat{G}_{jn}$ need only be computed at the uncensored points (X_i, Y_i). The advantage they possess over the other estimators encountered in this study is embodied in the fact that their computer alogorithms require the least cpu time.

CHAPTER VI

A DECOMPOSITION OF THE BIVARIATE SURVIVAL FUNCTION IN TERMS OF ESTIMABLE FUNCTIONS

6.0 INTRODUCTION

In this chapter, we present a new class of estimators of the bivariate survival function as was presented in a paper by Tsai, Leurgans and Crowley (1986). This new class of estimators is kernel and bandwidth dependent but not path dependent. Section 1 contains a decomposition of the bivariate survival function in terms of estimable functions. Based on this decomposition, a new class of estimators is presented in Section 2.

Throughout this chapter, $X \wedge Y$ denotes $\min(X,Y)$ and $[A]$ denotes the indicator of the event A. With this added notation, an equivalent expression for the Kaplan-Meier (1958) PL estimator (2.0.1) is given by

$$\hat{F}(s) = \prod_{X_i \le s} \left[1 - \frac{\xi_i}{\sum_j [X_j \ge X_i]}\right]. \tag{6.0.1}$$

6.1 DECOMPOSITION OF BIVARIATE SURVIVAL FUNCTIONS

Throughout the rest of this chapter, we give formulas for $s \ge t$. Definitions for $s < t$ are obtained by reversing the coordinates. We will use two assumptions, (A1) and (A2) which are well known to us by now to derive the decomposition.

(A1) The vectors (X^0, Y^0) and (C,D) are mutually independent.

(A2) the functions F^0 and G are absolutely continuous with respect to Lebesque measure on R^2.

G above denotes the bivariate survival function of the pair of censoring variables (C,D). The true pair of survival times is denoted by (X^0, Y^0) and its bivariate survival function is

$$F^0(t_1,t_2) = P(X^0 > t_1,\ Y^0 > t_2).$$

The decomposition will be expressed in terms of the following functions and sets:

$$\begin{aligned}
&F(t_1,t_2) = P(X > t_1,\ Y > t_2)\\
&F_3^0(t_2) \quad = F^0(t_2,t_2)\\
&F_2(t_1,t_2) = P(X > t_1,\ Y > t_2,\ \xi_2 = 1)\\
(6.1.1)\qquad &F_{12}(t_1,t_2) = P(X > t_1,\ Y > t_2,\ \xi_1 = 0,\ \ \xi_2 = 1)\\
&F^0(t_1|t_2) \quad = P(X^0 > t_1|Y^0 = t_2)\\
&R(s,t) \qquad = \{(t_1,t_2)|t_1 > s \geq t_2 \geq t\},\\
&\Delta(s,t) \qquad = \{(t_1,t_2)|s \geq t_1 \geq t_2 > t\}.
\end{aligned}$$

Thus, F, F_2 and F_{12} are the observable bivariate (sub) survival functions, $F^0(\cdot|\cdot)$ is the conditional survival function and F_3^0 is the probability that neither event has occured.

Lemma 6.1.1

Assume conditions (A1) and (A2) hold and let $s > t > 0$ be such that $F(s,s) > 0$. Then

$$F^0(s,t) = F_3^0(s) + \int\int_{R(s,t)} F_3^0(t_2)/F(t_2,t_2)\ D\ F_2(t_1,t_2)$$

(6.1.2)

$$+ \int\int_{\Delta(s,t)} \frac{F_3^0(t_2)}{F(t_2,t_2)} \frac{F^0(S|t_2)}{F^0(t_1|t_2)}\ D\ F_{12}(t_1,t_2)$$

where $0/0 = 0$, the integrals are Riemann-Stieltjes integrals, well defined since both F_2 and F_{12} have bounded variation, and D is the differential operator.

We will use Figure 6.1.1 below to throw more light on Lemma 6.1.1 above.

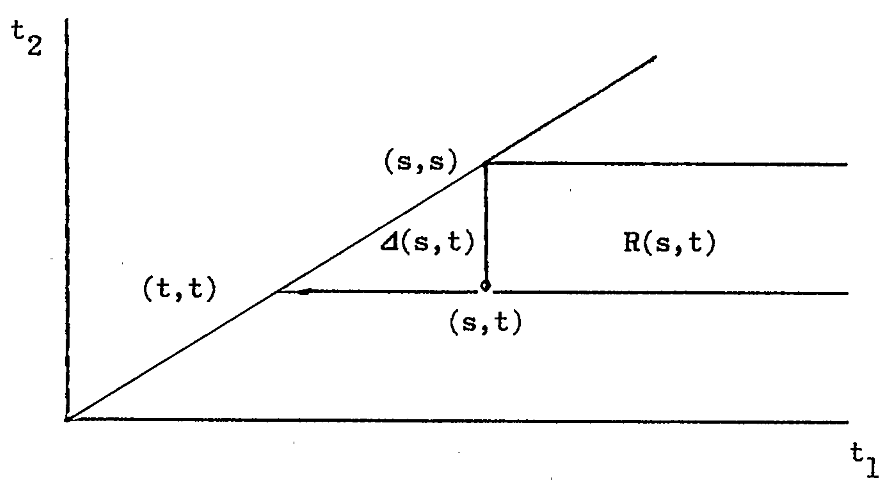

Figure 6.1.1

The probability F^0 assigned to the rectangle $R(s,t)$ can be split according to whether the first coordinate is censored before s and is written as the sum of $P\{(X^0,Y^0) \in R(s,t),\ \xi_1 = 0,\ X \le s\}$ and $P\{(X^0,Y^0) \in R(s,t);\ X > s\}$.

Thus,

$P\{(X^0,Y^0) \in R(s,t),\ \xi_1 = 0,\ X \le s\}$ is absolutely continuous with respect to the identifiable subdistribution F_{12} on the triangle $\Delta(s,t)$ and is displayed as an integral against that measure in Lemma 6.1.1. Similarly, $P\{(X^0,Y^0) \in R(s,t);\ X > s\}$ is written as an integral over $R(s,t)$ with respect to the identifiable subdistribution F_2. The factor $F_3^0(t_2)/F(t_2,t_2)$ is the reciprocal of the probability that $C \wedge D$ is greater than t_2; the factor $F^0(S|t_1)/F^0(t_1|t_2)$ in the former integral is the conditional probability that X^0 is in the rectangle given that the second coordinate is t_2, which forces (X^0,Y^0) to be in $\Delta(s,t) \cup R(s,t)$.

Proof:

We prove this lemma by rewriting each of the double integrals as integrals with respect to $D\,F^0(t_1,t_2)$

$$\int\int_{R(s,t)} \frac{F_3^0(t_2)}{F(t_2,t_2)}\, D\, F_2(t_1,t_2) = \int_{t_2=t}^{s}\int_{t_1=s}^{\infty} \frac{F_3^0(t_2)}{F(t_2,t_2)}\, D\, F_2(t_1,t_2)$$

$$(6.1.3) \qquad = - \int_{t_2=t}^{s} \frac{F_3^0(t_2)}{F(t_2,t_2)} D_{t_2} F_2(s,t_2).$$

But the definitions and (A1) imply that $F(t_2,t_2) = F_3^0(t_2)\, G(t_2,t_2)$ and $D_{t_2} F_2(s,t_2) = G(s,t_2)\, D_{t_2} F^0(s,t_2)$.

Therefore (6.1.3) above reduces to:

$$- \int_{t_2=t}^{s} \frac{F_3^0(t_2)}{F(t_2,t_2)} D_{t_2} F_2(s,t_2) = - \int_{t_2=t}^{s} \frac{G(s,t_2)}{G(t_2,t_2)} D_{t_2} F^0(s,t_2)$$

$$= \int_{t_2=t}^{s} \int_{t_1=s}^{\infty} \frac{G(s,t_2)}{G(t_2,t_2)} D\, F^0(t_1,t_2)$$

$$(6.1.4) \qquad = \iint_{R(s,t)} \frac{G(s,t_2)}{G(t_2,t_2)} D\, F^0(t_1,t_2).$$

Similarly, $D\, F_{12}(t_1,t_2) = D_{t_1} G(t_1,t_2)\, D_{t_2} F^0(t_1,t_2)$ and

$$F^0(t_1|t_2) = P(X^0 > t_1 | Y^0 = t_2)$$

$$= P(X^0 > t_1,\ Y^0 = t_2)/P(Y^0 = t_2)$$

$$= \frac{D_{t_2} F^0(t_1,t_2)}{D_{t_2} F^0(-\infty,t_2)} .$$

Therefore, $\displaystyle\int\int_{\Delta(s,t)}^{\infty} \frac{F_3^0(t_2)}{F(t_2,t_2)} \cdot \frac{F^0(s|t_2)}{F^0(t_1|t_2)}\, D\, F_{12}(t_1,t_2)$

$$= \int_{t_2=t,\,t_1=t_2}^{s}\int^{s} \frac{1}{G(t_2,t_2)} \cdot \frac{D_{t_2}\, F^0(s,t_2)}{D_{t_2}\, F^0(t_1,t_2)}\, D_{t_1}\, G(t_1,t_2)\, D_{t_2}\, F^0(t_1,t_2)$$

$$= \int_{t_2=t}^{s} \frac{D_{t_2}\, F^0(s,t_2)}{G(t_2,t_2)} \left[G(s,t_2) - G(t_2,t_2)\right]$$

$$= -\int_{t_2=t}^{s} \frac{G(t_2,t_2) - G(s,t_2)}{G(t_2,t_2)}\, D_{t_2}\, F^0(s,t_2)$$

$$= -\int_{t_2=t}^{s} \left[1 - \frac{G(s,t_2)}{G(t_2,t_2)}\right] D_{t_2}\, F^0(s,t_2)$$

$$(6.1.5) = \int\int_{R(s,t)} \left[1 - \frac{G(s,t_2)}{G(t_2,t_2)}\right] D\, F^0(t_1,t_2).$$

Therefore the sum of (6.1.4) and (6.1.5) gives

$$\int\int_{R(s,t)} D\, F^0(t_1,t_2) = F^0(s,t) - F^0(s,s)$$

$$= F^0(s,t) - F_3^0(s),$$

Q.E.D.

Lemma (6.1.2) below is Beran's (1981) extension of Peterson's (1977) decomposition to allow simultaneous discontinuities in death and censoring times.

The product integral of a function g is defined by

$$[\gamma(g)](t) = \lim_{\max\limits_{1\le k\le r} (u_k-u_{k-1})\to 0} \prod_{i=1}^{r} \left\{1-(g(u_i) - g(u_{i-1}))\right\},$$

where $0 = u_0 < u_1 < \dots < u_\gamma = t$.

For continuous functions g, $[\gamma(g)](t) = \exp(-g(t))$. If g is an empirical cumulative hazard, $[\gamma(g)](t)$ is the corresponding product-limit or Kaplan-Meier estimator.

Lemma 6.1.2

If $F_0(t) = P(X^0 > t)$, $F(t) = P(X^0 \wedge C > t)$ and $F_\mu(t) = P(C > X^0 > t)$, where X^0 and C are independent random variables, then

$$F_0(t) = \left[\gamma\left[-\int_0^{u^+} \frac{D\, F_\mu(x)}{F(x-)}\right]\right](t).$$

The function F_3^0 is the survival function of $T_3^0 = X^0 \wedge Y^0$. Define the corresponding censoring time $C_3 = C \wedge D$, the observed time $T_3 = T_3^0 \wedge C_3$, and the indicator $\xi_3 = [T_3^0 \le C_3]$. The survival function F_3^0 can be decomposed in terms of $F_3(t) = P(T_3 > t)$ and $F_{3\mu}(t) = P(T_3 > t,\ \xi_3 = 1)$. Since $T_3 = X \wedge Y$ and $\xi_3 = [X > Y]\,\xi_2 + [X < Y]\,\xi_1$, (T_3, ξ_3) is a function of (X, Y, ξ_1, ξ_2) and F_3 and $F_{3\mu}$ can be estimated empirically.

We now state the decomposition.

Theorem 6.1.1.

If the conditions (A1) and (A2) are met, then

$$F^0(s,t) = \left[\gamma\left[-\int_0^{u^+} \frac{D\, F_{3\mu}(z)}{F_3(z-)}\right]\right](s)$$

$$+ \iint_{R(s,t)} \left\{\left[\gamma\left[-\int_0^{u^+} \frac{D\, F_{3\mu}(z)}{F_3(z)}\right]\right](t_2-)\right\} / \left[F(t_2-,t_2-)\right]\, D\, F_2(t_1,t_2)$$

$$(6.1.6) \quad + \iint_{\Delta(s,t)} \left\{\left[\gamma\left[-\int_0^{u^+} \frac{D\, F_{3\mu}(z)}{F_3(z-)}\right]\right](t_2-)\right\} / \left[F(t_2-,t_2-)\right]$$

$$\times \left\{\left[\gamma\left[-\int_0^{u^+} \frac{D_z F_\mu(z|t_2-)}{F(z-|t_2-)}\right]\right](s)\right\} \Big/ \left\{\left[\gamma\left[-\int_0^{u^+} \frac{D_z F_\mu(z|t_2-)}{F(z-|t_2-)}\right]\right](t_1-)\right\} DF_{12}(t_1,t_2),$$

where $F_\mu(t|t_2) = P(X > t,\ \xi_1 = 1|Y = t_2,\ \xi_2 = 1)$ and $F(t|t_2) = P(X > t|Y = t_2,\ \xi_2 = 1)$.

Proof:

From the assumption (A1), D is independent of X^0, we observe

$$\begin{aligned} F^0(t|t_2) &= P(X^0 > t|Y^0 = t_2) \\ &= P(X^0 > t|Y^0 = t_2,\ D > t_2) \\ &= P(X^0 > t|Y^0 = t_2,\ \xi_2 = 1). \end{aligned}$$

The theorem follows from applying Lemma 6.1.2 to F_3 and to $P(X^0 > t|Y = t_2,\ \xi_2 = 1)$ and substituting the resulting representations in Lemma 6.1.1.

Under the continuity assumptions of (A2), (6.1.6) reduces to (6.1.2). We prefer (6.1.6) becuase (A2) is not required for Lemma (6.1.1) or Theorem (6.1.1) and because (6.1.6) can be applied to empirical subsurvival functions to obtain estimators.

6.2 ESTIMATORS OF F^0

Suppose the iid random vectors $\{(X_i,\ Y_i,\ \xi_{1i},\ \xi_{2i})\ \ i = 1,\ldots,n\}$ have the same distribution as the random vector $(X,\ Y,\ \xi_1,\ \xi_2)$. In this section we develop an estimator of F^0.

Natural unbiased estimators of the (sub) survival functions in (6.1.6) are defined below in terms of X_i, Y_i. ξ_{1i}, ξ_{2i}, $T_{3i} = X_i \wedge Y_i$ and $\xi_{3i} = [X_i > Y_i]\ \xi_{2i} + [X_i < Y_i]\ \xi_{1i}$:

$$F^e(t_1,t_2) = \frac{1}{n} \sum_i [X_i > t_1,\ Y_i > t_2],$$

$$F^e_{12}(t_1,t_2) = \frac{1}{n} \sum_i [X_i > t_1,\ Y_i > t_2,\ \xi_{1i} = 0,\ \xi_{2i} = 1]$$

$$F^e_2(t_1,t_2) = \frac{1}{n} \sum_i [X_i > t_1,\ Y_i > t_2,\ \xi_{2i} = 1]$$

$$F^e_3(t_1) = \frac{1}{n} \sum_i [T_{3i} > t_1]$$

$$F^e_{3\mu}(t_1) = \frac{1}{n} \sum_i [T_{3i} > t_1,\ \xi_{3i} = 1]$$

substituting F^e_3 and $F^e_{3\mu}$ into Lemma 6.1.2, we have the Kaplan-Meier estimator for $F^0_3(t)$:

(6.2.1) $$F^0_3(t) = \left[\gamma\left[-\int_0^{u^+} \frac{D\ F^e_{3\mu}(t_1)}{F^e_3(t_1-)}\right]\right](t).$$

The functions $F(t_1|t_2)$ and $F_\mu(t_1|t_2)$ are the conditional probabilities given $Y_i = t_2$, $\xi_{2i} = 1$. Since the assumptions of absolute continuity implies that there is (a.s.) at most one Y_i that is equal to t_2 with $\xi_{2i} = 1$, $F(t_1|t_2)$ and $F_\mu(t_1|t_2)$ can not be estimated stably without smoothing. To estimate a conditional survival function given t_2, we apply the nonnegative weight $W_{n_i}(t_2)$ to (X_i, Y_i). The weight $W_{n_i}(t_2)$ depends on the data through the distance between t_2 and Y_i and through the second components $\{(Y_j, \xi_{2j}),\ j = 1,\ldots,n\}$. With the assumption that $\sum_i W_{n_i}(t_2)\, \xi_{2i} = 1$, the following estimators are discrete subsurvival functions:

$$\hat{F}(t_1|t_2) = \sum_i W_{n_i}(t_2)\ [X_i > t_2,\ \xi_{2i} = 1]$$

$$\hat{F}_\mu(t_1|t_2) = \sum_i W_{n_i}(t_2)\ [X_i > t_1,\ \xi_{1i} = \xi_{2i} = 1].$$

Substituting the estimators $\hat{F}(t_1|t_2)$ and $\hat{F}_\mu(t_1|t_2)$ into the equation yields the following natural estimator for $\hat{F}^0(t_1|t_2)$:

$$\hat{F}^0(t_1|t_2) = \left[\gamma\left[-\int_0^{u^+} \frac{D_t \hat{F}_\mu(t|t_2)}{\hat{F}(t|t_2)}\right]\right](t_1).$$

If $\hat{F}(t_1|t_2) > 0$, and if the jumps points of $\hat{F}_\mu(\cdot|t_2)$ are $t_{11},\ldots,t_{1m}$, then

$$\hat{F}^{o}(t_1|t_2) = \prod_{t_{1i} \le t_1} \left[1 - \frac{\hat{F}_{\mu}(t_{1i}^{-}|t_2) - \hat{F}_{\mu}(t_{1i}|t_2)}{\hat{F}(t_{1i}^{-}|t_2)} \right].$$

Substituting $\hat{F}_3^o$, F_2^e, F_{12}^e, F^e and $\hat{F}^o(t_1|t_2)$ into (6.1.2), we obtain the following estimator of F^o:

(6.2.2) $$\hat{F}^o(s,t) = \hat{F}_3^o(s) + \int\int_{R(s,t)} \hat{F}_3^o(t_2^-)/F^e(t_2^-,t_2^-)\ D\ F_2^e(t_1,t_2)$$

$$+ \int\int_{\Delta(s,t)} \hat{F}_3^o(t_2^-)\ \hat{F}^o(S|t_2^-)/(F^e(t_2^-,t_2^-)\ \hat{F}^o(t_1^-|t_2^-) D\ F_{12}^e(t_1,t_2).$$

Choice of an estimator within this class requires the specification of the weight functions $W_{n_i}(t_2)$. Here we focus on Kernel weights. These weights are constructed by selecting a nonnegative function $k(\cdot)$ of bounded variation on the real line and a sequence of bandwidths $\{h(n),\ n \ge 1\}$ converging to zero. the probability weights are then

(6.2.3) $$W_{n,i}(t_2) = k((Y_i - t_2)/h(n))\ \xi_{2i} \Big/ (\sum_j k((Y_j - t_2)/h(n))\xi_{2j}).$$

That is, we give positive weight only to those observations with an observed failure in the second component near t_2.

In our study, we chose $k(y) = \frac{1}{2}$ on the segment $|y| \leq 1$ and 0 otherwise. $h(n) = n^{-1/5}$, where n is the sample size, was chosen as the bandwidth. This choice of k and h yielded a particularly simple expression for (6.2.2).

It was realized that at most sample points, the contribution made by $\Delta(s,t)$ to the probability being estimated was very small if not zero. The fact that the estimator proposed in this chapter is fairly complicated is a factor worth considering. Further, although it is kernel and bandwidth dependent, this arbitrariness might not matter as much as the path dependence of the estimators mentioned in Chapter IV. Although it was believed at first that a heavy price in loss of efficiency would have to be paid for this arbitrariness, the results of this study indicate otherwise.

CHAPTER VII

NUMERICAL RESULTS

7.0 INTRODUCTION

In this Chapter, we give a simple scheme for generating random numbers from a specific bivariate survival distribution in the presence of censoring. Various procedures for generating independent random numbers from specific univariate distributions are well known to us, but few results are available for non-normal multivariate distributions. Random samples of various sizes were simulated from a given bivariate distribution. There were three different samples corresponding to each sample size:

(i) 10% censoring,

(ii) 40% censoring,

(iii) 50% censoring.

Computer algorithms were written only for the numerical solutions of estimators (4.1.8), (4.2.19), (5.1.3), (5.2.2), (6.3.2) and (2.2.3). The performance of these estimators were compared using the average square error and contour plots.

7.1 PROCEDURE

Cherian (1941) constructed a bivariate gamma distribution as follows. Let X_1, X_2, X_3 be independent and have gamma distributions with index parameters P_1, P_2 and P_3 respectively. Then $X^0 = X_1 + X_3$ and $Y^0 = X_2 + X_3$ have a bivariate gamma distribution. The joint p.d.f. of X^0 and Y^0 is given by

$$(7.1.1)\quad h(x^0,y^0) = \frac{e^{-(x^0+y^0)}}{\prod_{i=1}^{3} \Gamma(P_i)} \int_0^{\min(x^0,y^0)} z^{P_3-1}(x^0-z)^{P_1-1}(y^0-z)^{P_2-1} e^z dz.$$

For $P_1 = P_2 = P_3 = 1$, (7.1.1) reduces to

$$(7.1.2)\qquad h(x^0,y^0) = \begin{cases} e^{-y^0}(1-e^{-x^0}), & x^0 < y^0 \\ e^{-x^0}(1-e^{-y^0}), & x^0 > y^0. \end{cases}$$

Using (7.1.2) it can be shown that

$$F(s,t) = P(X^0 > s,\ Y^0 > t)$$

$$(7.1.3)\qquad = \begin{cases} (t-s+e^{-t}-e^{-s})\ e^{-t} + 2(e^{-t} - \frac{1}{2}\ e^{-2t}), & s < t \\ (s-t+e^{-s}-e^{-t})\ e^{-s} + 2(e^{-s} - \frac{1}{2}\ e^{-2s}), & s \geq t. \end{cases}$$

(7.1.2) represents the p.d.f. of the true survival times (x^0, y^0). The censoring variable (C,D) were generated from (7.1.1) using (i) (2, 3, 0.5), (ii) (2, 3, 1) and (iii) (2.5, 3.5, 2) as values for the vector (P_1, P_2, P_3). These gave rise to 50%, 40% and 10% censoring respectively. In the case of 40% censoring, for each type of estimate and sample size, 25 different samples were taken and the survival estimates formed at the sample points. The sample sizes chosen were {10, 30, 50, ..., 170}. The same was done in the case of 50% and 10% censoring except that the chosen sample sizes were {10, 30, ..., 90}. It is worthwhile to note that the variables X^0, Y^0, C, D were generated by calling the GGAMR package of IMSL (International Mathematical and Statistical Library).

As a measure of goodness of performance of an estimator, many authors use the mean integrated square error, abbreviated as M.I.S.E. If f is the true survival function and $\hat{f}$ is the estimator, the M.I.S.E. is computed by

$$E \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[f(x,y) - \hat{f}(x,y)\right]^2 w(x,y)\, dxdy.$$

The function w is referred to as weight function and is often taken to be identically 1. The integrated square error however, is some what difficult to obtain numerically so we decided to form the average square error,

$$\left[\frac{1}{n}\right] \sum_{i=1}^{n} \left[\hat{f}_n(x_i,y_i) - f(x_i,y_i)\right]^2$$

at the sample points (observations).

For each sample size, the average value of the average square error was computed and are found in Tables 7.2.1, 7.2.4 and 7.2.7.

In the case of estimator (2.2.3) and 40% censoring, the average square errors were only computed up to a maximum sample size of 90. This was a consequence of time sharing vis-a-vis the long computation time required.

The median value of the average square errors are also given in Tables 7.2.2, 7.2.5 and 7.2.8. The motivation behind this is based on the simple realization that the median is affected less than the mean by an occasional large error. As a measure of dispersion, the standard deviation of the average square errors were also formed as given in Tables 7.2.3, 7.2.6 and 7.2.9.

For each censoring scheme and the estimators for which computer algorithms were written, further analysis was carried out by generating the contour plots of the function

$$C(s,t) = F(s,t) - \hat{F}(s,t),$$

for a sample size of 150. $F(s,t)$ and $\hat{F}(s,t)$ are the real and estimated survival functions respectively. Such contour plots are shown in Figures 7.2.1 through 7.2.18.

7.2 OBSERVATIONS AND CONCLUSIONS

From Tables 7.2.3, 7.2.6 and 7.2.9, with increasing sample size the dispersion of the average square errors tends to zero. As expected, it tends to zero quite rapidly with reduced censoring. This does not seem to be the case for estimator (2.2.3). An explanation for this could be found in the fact that in deriving the self-consistent estimator, the weight of the censored observations was spread on all points beyond the censored point. The simulation results thus indicate that estimator (2.2.3) gives average square errors that are quite erratic.

The study also showed that estimator (4.1.8) performed as well as estimator (4.2.19). This was a confirmation of what was expected from theory. The same observations were also made about estimators (5.1.3) and (5.2.2). The contour plots attached attest to the above. In short, it does not seem to matter whether one used the hazard gradient approach or the Kaplan-Meier product limit estimators.

Estimators (5.1.3) and (5.2.2) have fewer support points than the other estimators. Their support points are the observed uncensored points (X_i, Y_i). Thus the survival function G need only be computed at such support points, and hence their estimates were fairly simple to compute. Their computations required the least computation time. As expected, their average square errors reduced with reduced censoring.

Estimators (4.1.8) and (4.2.19) gave the least average square errors. Such performance may be due to the fact that they do not depend on other estimable functions but rely heavily on the data. The study also indicated that the average square errors of estimators (4.1.8) and (4.2.19) on one hand are comparable to those of estimator (6.3.2) as shown in Tables 7.2.1, 7.2.4 and 7.2.7. Further, the remarkable similarity in the contour plots of these estimators is compatible with the above observation.

The fact that estimators (4.1.8) and (4.2.19) are path dependent and may fail to be survival functions leaves us with no alternative but to accept estimator (6.3.2) as the one with the best mean integrated square error (M.I.S.E.) property.

This thesis is limited in scope in the sense that only one distribution was considered in the generation of the random variables. The comparisons made among the various estimators were based upon computations made at the sample points. The observations made are therefore only valid within the framework of the study done.

The experimental data as well as the computer programs used in this study are available in the files of the Mathematics Department of the University of Calgary.

TABLE 7.2.1

THE TABLE BELOW GIVES THE AVERAGE OF THE AVERAGE SQUARE ERRORS FOR THE VARIOUS ESTIMATORS UNDER VARRYING SAMPLE SIZES WITH 10% CENSORING.

n	4.1.8	4.2.19	5.1.3	5.2.2	6.3.2	2.2.3
10	0.0179	0.0205	0.0380	0.0393	0.0177	0.0483
30	0.0054	0.0059	0.0087	0.0084	0.0053	0.0690
50	0.0035	0.0035	0.0069	0.0069	0.0033	0.0577
70	0.0021	0.0023	0.0036	0.0040	0.0022	0.0525
90	0.0016	0.0017	0.0020	0.0021	0.0016	0.0752

n IS THE SAMPLE SIZE

TABLE 7.2.2

THE TABLE BELOW GIVES THE MEDIAN OF THE AVERAGE SQUARE ERRORS FOR THE VARIOUS ESTIMATORS UNDER VARRYING SAMPLE SIZES WITH 10% CENSORING.

n	4.1.8	4.2.19	5.1.3	5.2.2	6.3.2	2.2.3
10	0.0126	0.0171	0.0285	0.0307	0.0125	0.0294
30	0.0039	0.0050	0.0068	0.0066	0.0043	0.0440
50	0.0026	0.0027	0.0053	0.0054	0.0029	0.0475
70	0.0013	0.0014	0.0026	0.0031	0.0015	0.0382
90	0.0013	0.0014	0.0017	0.0020	0.0014	0.0608

n IS THE SAMPLE SIZE

TABLE 7.2.3

THE TABLE BELOW GIVES THE STANDARD DEVIATION OF THE AVERAGE SQUARE ERRORS FOR THE VARIOUS ESTIMATORS UNDER VARRYING SAMPLE SIZE WITH 10% CENSORING.

n	4.1.8	4.2.19	5.1.3	5.2.2	6.3.2	2.2.3
10	0.0001	0.0002	0.0010	0.0010	0.0001	0.0015
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0023
50	0.0000	0.0000	0.0000	0.0000	0.0000	0.0009
70	0.0000	0.0000	0.0000	0.0000	0.0000	0.0009
90	0.0000	0.0000	0.0000	0.0000	0.0000	0.0024

n IS THE SAMPLE SIZE

TABLE 7.2.4

THE TABLE BELOW GIVES THE AVERAGE OF THE AVERAGE SQUARE ERRORS FOR THE VARIOUS ESTIMATORS WITH VARRYING SAMPLE SIZES UNDER 40% CENSORING.

n	4.1.8	4.2.19	5.1.3	5.2.2	6.3.2	2.2.3
10	0.0239	0.0269	0.0541	0.0513	0.0241	0.0581
30	0.0078	0.0086	0.0156	0.0151	0.0086	0.0619
50	0.0044	0.0044	0.0112	0.0113	0.0045	0.0718
70	0.0028	0.0030	0.0086	0.0086	0.0028	0.0639
90	0.0019	0.0020	0.0054	0.0060	0.0020	0.0773
110	0.0014	0.0015	0.0027	0.0030	0.0014	
130	0.0018	0.0019	0.0037	0.0041	0.0019	
150	0.0014	0.0014	0.0032	0.0036	0.0017	
170	0.0013	0.0013	0.0030	0.0032	0.0013	

n IS THE SAMPLE SIZE

TABLE 7.2.5

THE TABLE BELOW GIVES THE MEDIAN OF THE AVERAGE SQUARE ERRORS FOR THE VARIOUS ESTIMATORS UNDER VARRYING SAMPLE SIZES WITH 40% CENSORING.

n	4.1.8	4.2.19	5.1.3	5.2.2	6.2.3	2.2.3
10	0.0209	0.0220	0.0400	0.0419	0.0202	0.0349
30	0.0059	0.0073	0.0119	0.0106	0.0068	0.0509
50	0.0038	0.0035	0.0076	0.0081	0.0039	0.0540
70	0.0021	0.0019	0.0071	0.0062	0.0022	0.0436
90	0.0016	0.0019	0.0042	0.0044	0.0018	0.0629
110	0.0013	0.0014	0.0023	0.0023	0.0010	
130	0.0013	0.0013	0.0025	0.0023	0.0015	
150	0.0011	0.0011	0.0033	0.0029	0.0013	
170	0.0009	0.0008	0.0023	0.0024	0.0008	

n IS THE SAMPLE SIZE

TABLE 7.2.6

THE TABLE BELOW GIVES THE STANDARD DEVIATION OF THE AVERAGE SQUARE ERRORS FOR THE VARIOUS ESTIMATORS UNDER VARRYING SAMPLE SIZES WITH 40% CENSORING

n	4.1.8	4.2.19	5.1.3	5.2.2	6.3.2	2.2.3
10	0.0002	0.0002	0.0019	0.0018	0.0002	0.0028
30	0.0000	0.0000	0.0001	0.0001	0.0000	0.0017
50	0.0000	0.0000	0.0001	0.0001	0.0000	0.0024
70	0.0000	0.0000	0.0000	0.0000	0.0000	0.0021
90	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026
110	0.0000	0.0000	0.0000	0.0000	0.0000	
130	0.0000	0.0000	0.0000	0.0000	0.0000	
150	0.0000	0.0000	0.0000	0.0000	0.0000	
170	0.0000	0.0000	0.0000	0.0000	0.0000	

n IS THE SAMPLE SIZE

TABLE 7.2.7

THE TABLE BELOW GIVES THE AVERAGE OF THE AVERAGE SQUARE ERRORS FOR THE VARIOUS ESTIMATORS UNDER VARRYING SAMPLE SIZES WITH 50% CENSORING

n	4.1.8	4.2.19	5.1.3	5.2.2	6.3.2	2.2.3
10	0.0292	0.0322	0.0963	0.1050	0.0299	0.0610
30	0.0092	0.0099	0.0272	0.0261	0.0098	0.0585
50	0.0044	0.0043	0.0138	0.0134	0.0051	0.0716
70	0.0038	0.0040	0.0089	0.0093	0.0044	0.0683
90	0.0022	0.0024	0.0075	0.0093	0.0025	0.0714

n IS THE SAMPLE SIZE

TABLE 7.2.8

THE TABLE BELOW GIVES THE MEDIAN OF THE AVERAGE SQUARE ERRORS FOR THE VARIOUS ESTIMATORS UNDER VARRYING SAMPLE SIZES WITH 50% CENSORING

n	4.1.8	4.2.19	5.1.3	5.2.2	6.3.2	2.2.3
10	0.0223	0.0236	0.0582	0.0557	0.0225	0.0410
30	0.0065	0.0072	0.0227	0.0231	0.0075	0.0423
50	0.0041	0.0035	0.0117	0.0121	0.0050	0.0526
70	0.0029	0.0027	0.0068	0.0078	0.0036	0.0378
90	0.0019	0.0020	0.0054	0.0063	0.0023	0.0606

n IS THE SAMPLE SIZE

TABLE 7.2.9

THE TABLE BELOW GIVES THE STANDARD DEVIATION OF THE AVERAGE SQUARE ERRORS FOR THE VARIOUS ESTIMATORS UNDER VARRYING SAMPLE SIZES WITH 50% CENSORING

n	4.1.8	4.2.19	5.1.3	5.2.2	6.3.2	2.2.3
10	0.0005	0.0005	0.0112	0.0127	0.0007	0.0031
30	0.0001	0.0001	0.0006	0.0005	0.0001	0.0018
50	0.0000	0.0000	0.0001	0.0001	0.0000	0.0031
70	0.0000	0.0000	0.0001	0.0000	0.0000	0.0034
90	0.0000	0.0000	0.0000	0.0001	0.0000	0.0019

n IS THE SAMPLE SIZE

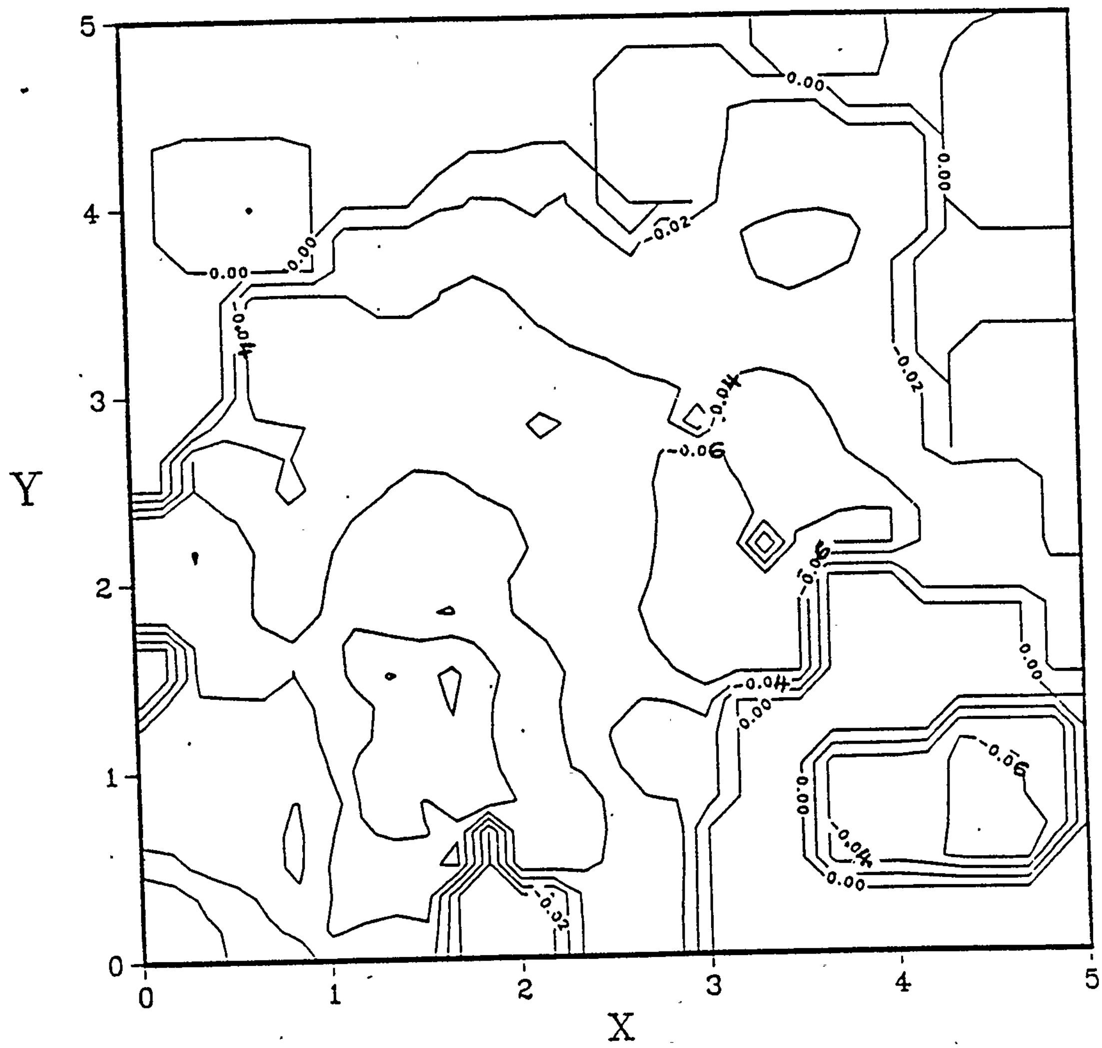

Figure 7.2.1 Difference Map For Estimator (4.1.8)
Sample Size of 150, 10% censoring

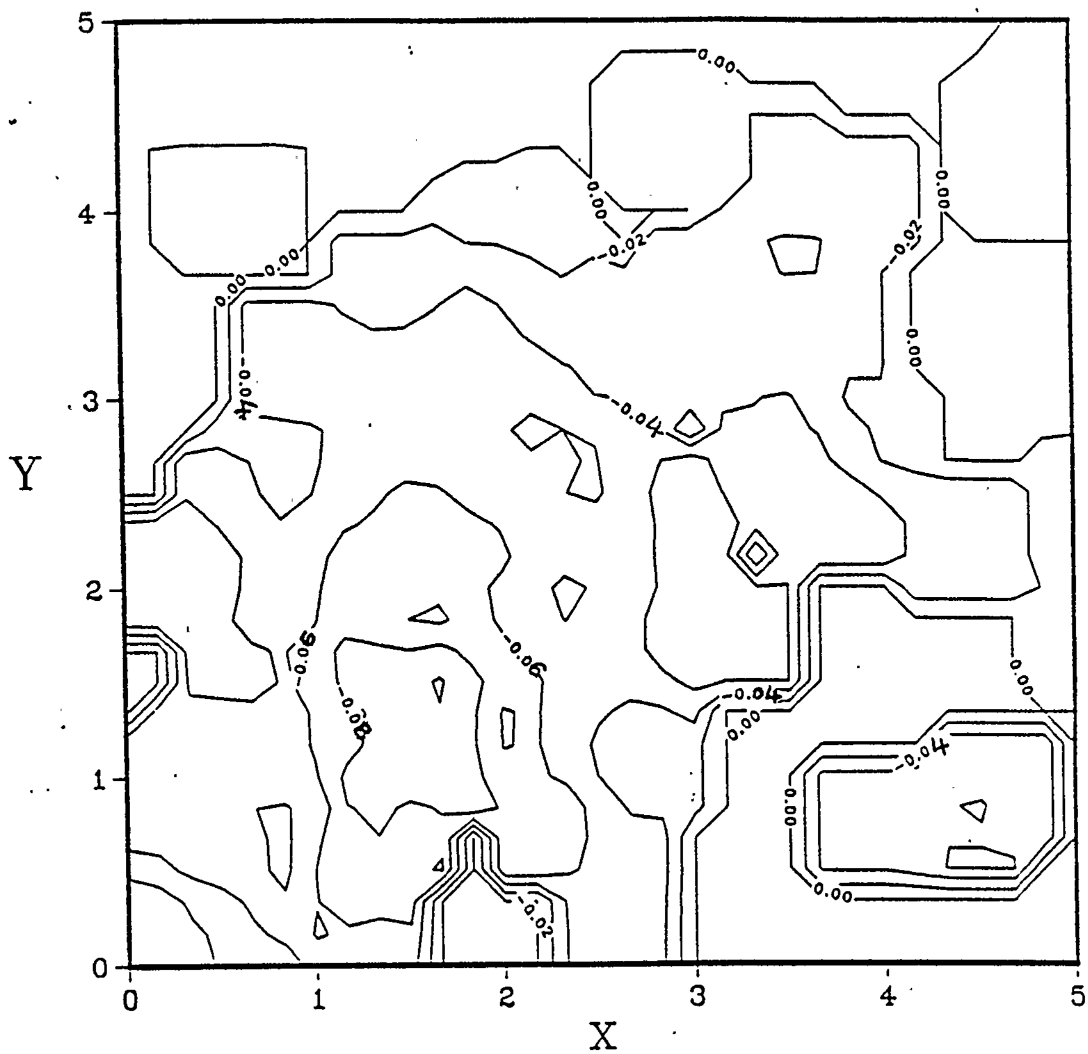

Figure 7.2.2 Difference Map for Estimator (4.2.19)
Sample Size Of 150, 10% censoring

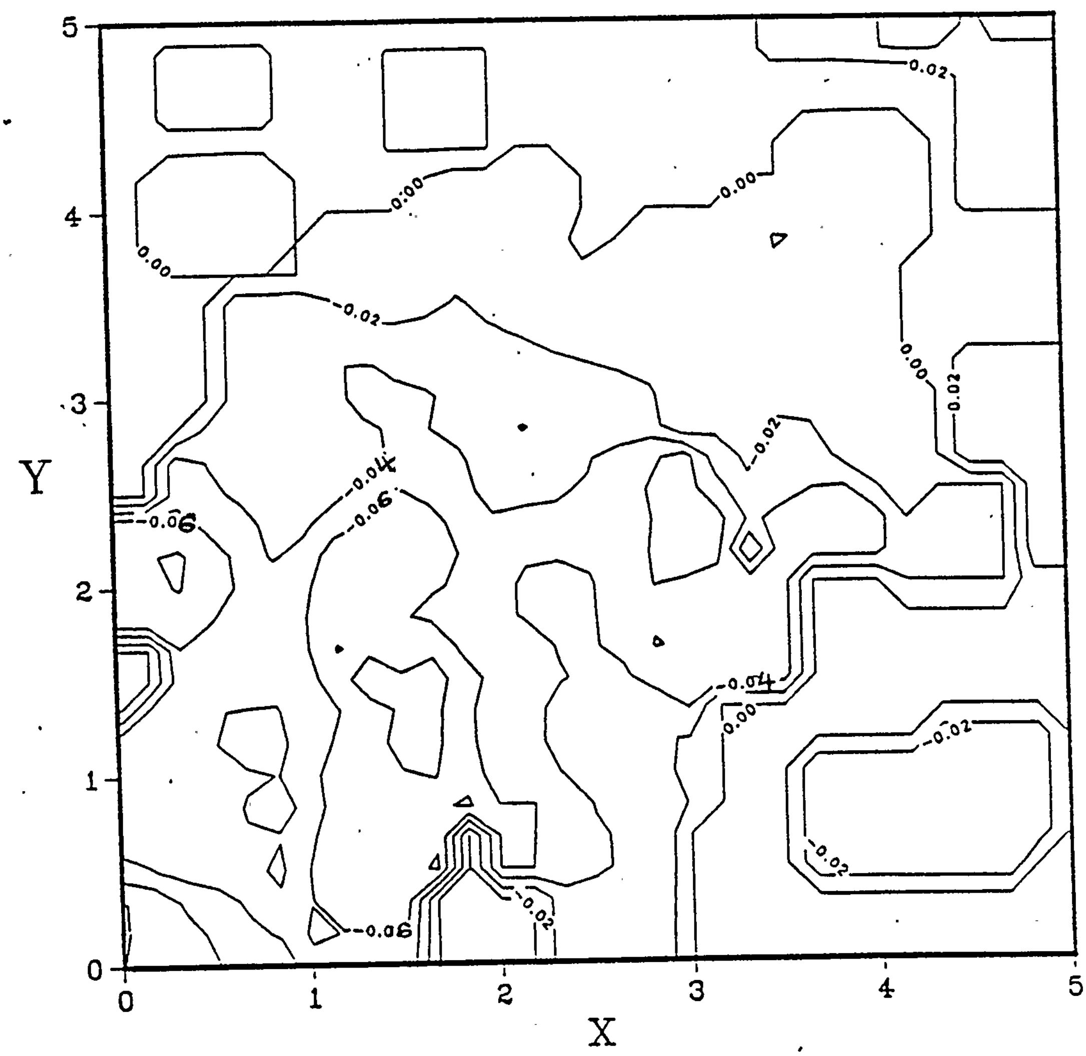

Figure 7.2.3 Difference Map For Estimator (5.1.3)
Sample Size Of 150, 10% censoring

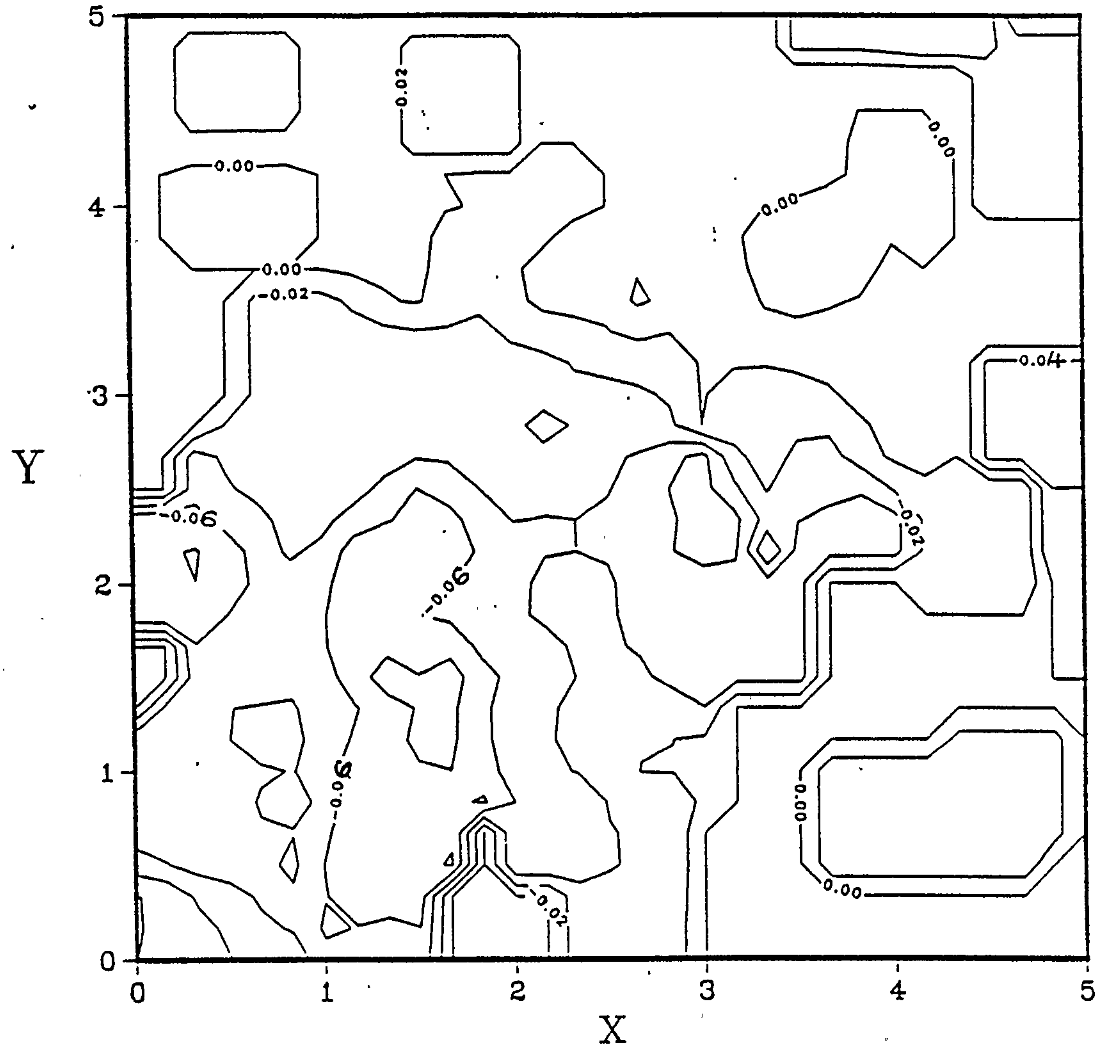

Figure 7.2.4 Difference Map For Estimator (5.2.2)
Sample Size Of 150, 10% censoring

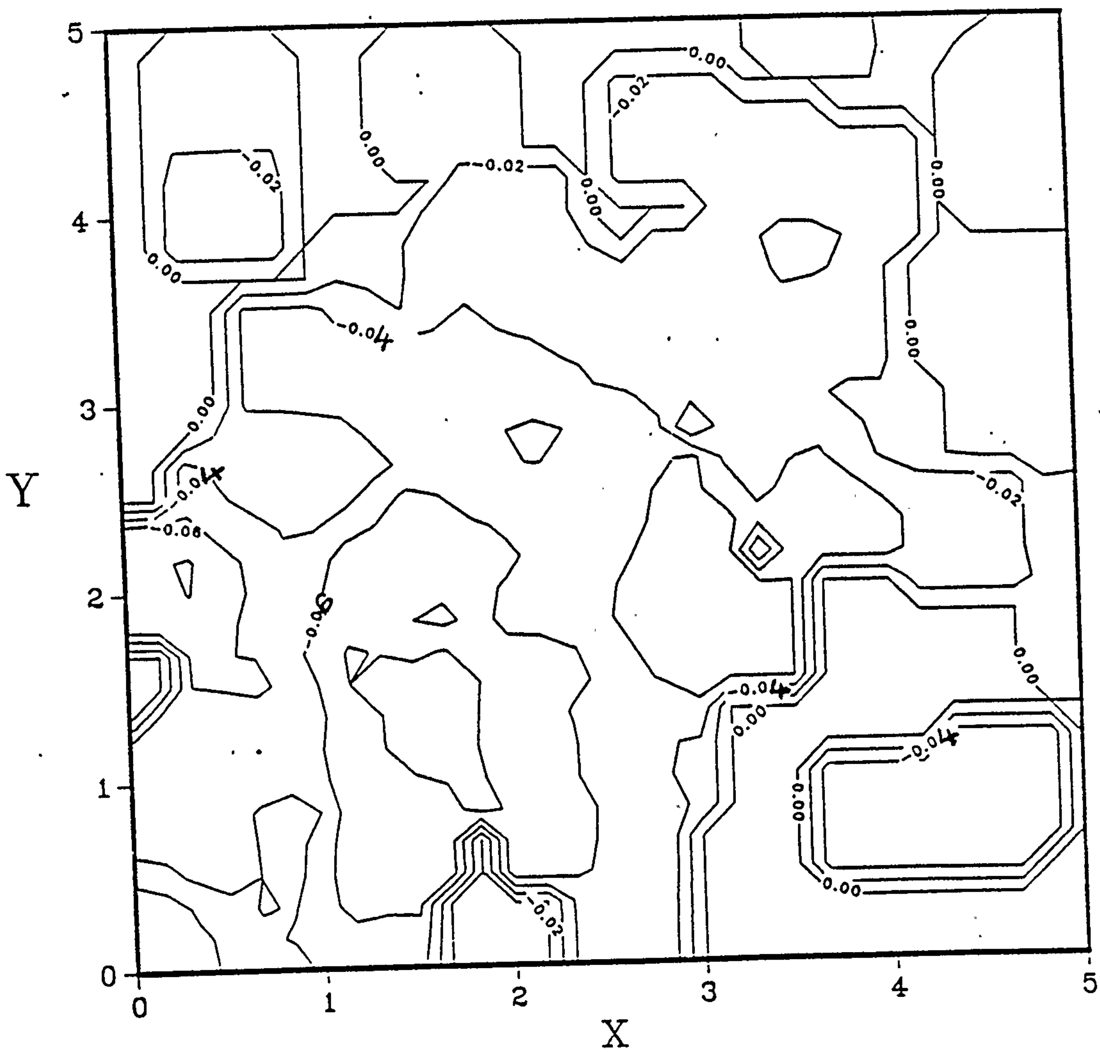

Figure 7.2.5 Difference Map for Estimator (6.3.2)
Sample Size Of 150, 10% censoring

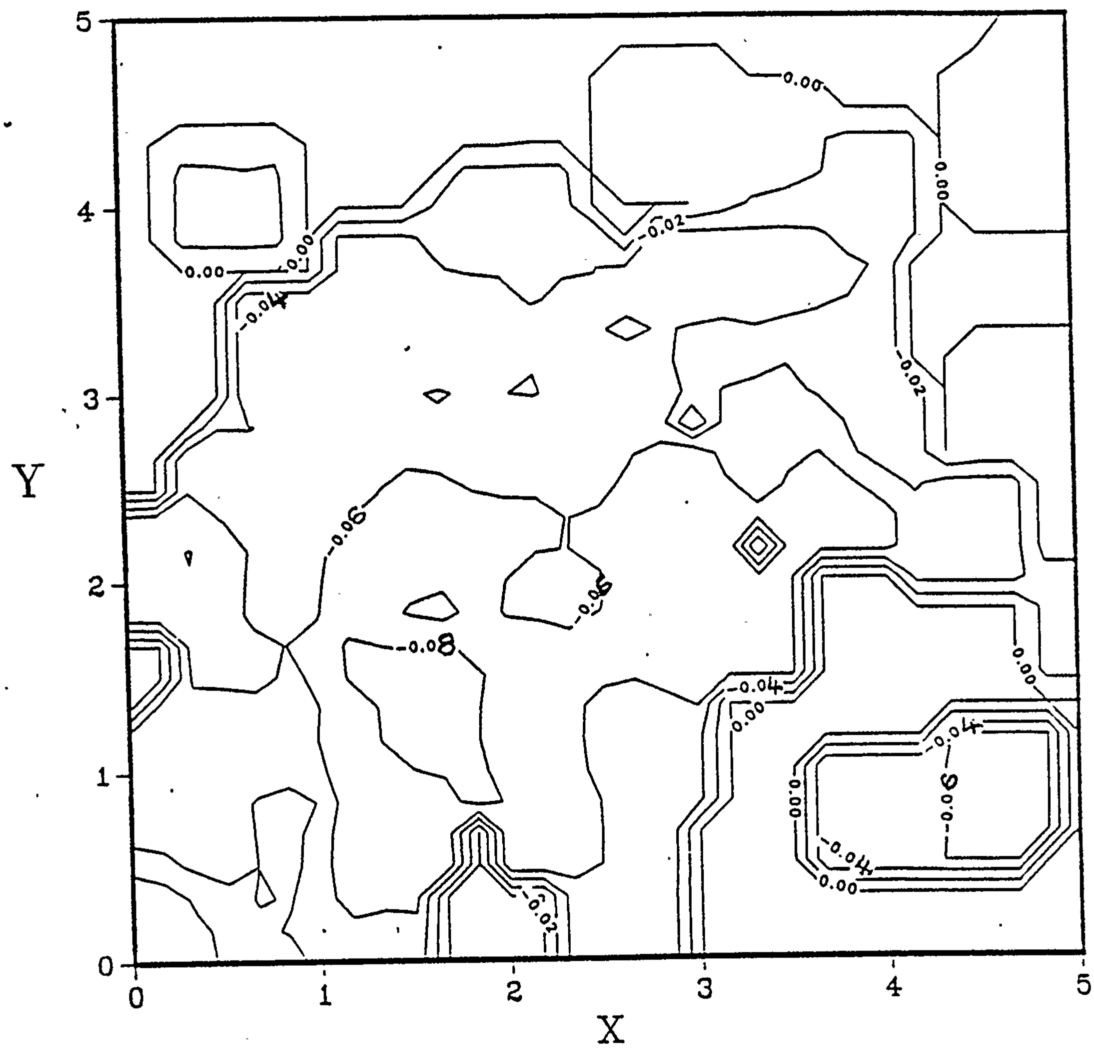

Figure 7.2.6 Difference Map For Estimator (2.2.3)
Sample Size Of 150, 10% censoring

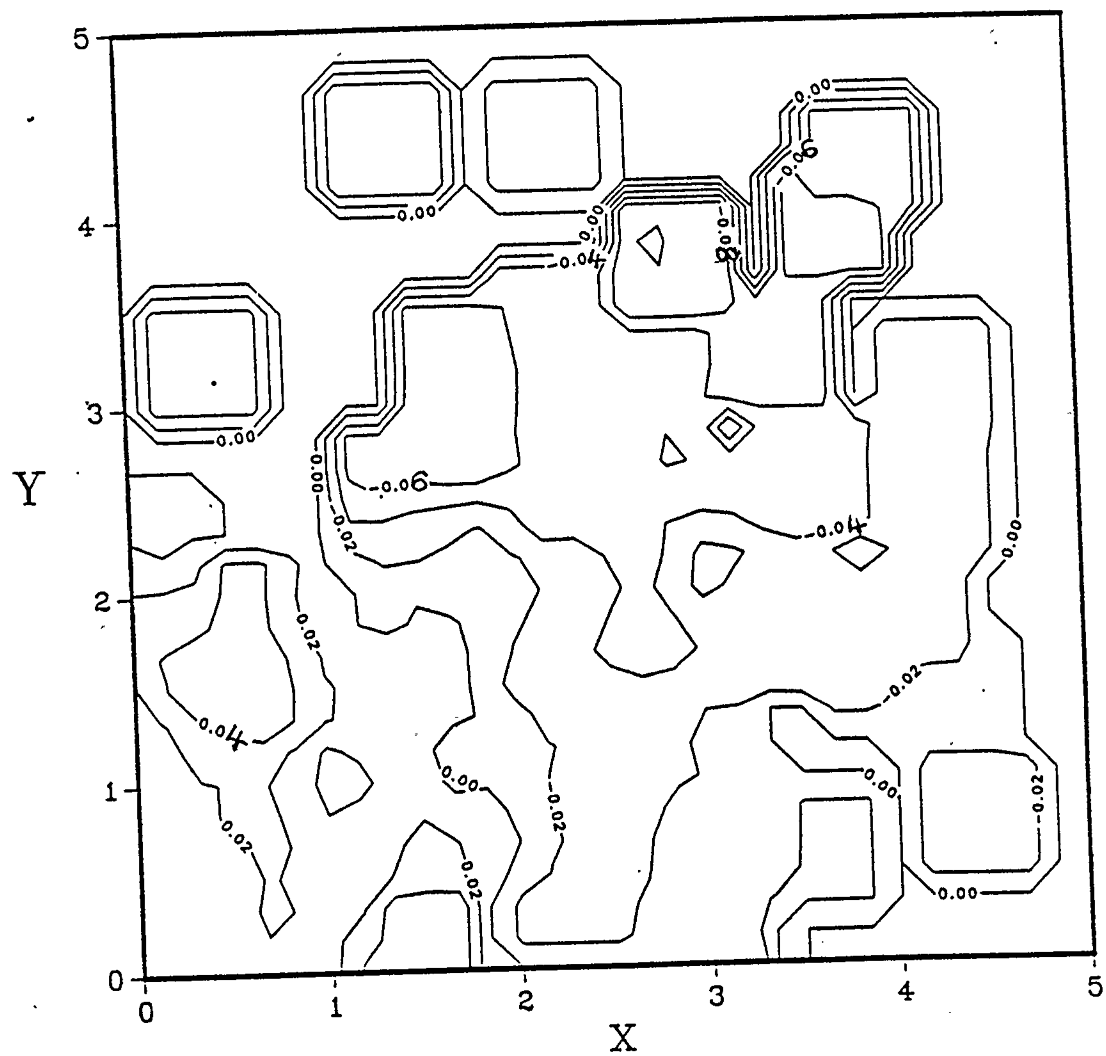

Figure 7.2.7 Difference Map For Estimator (4.1.8)
Sample Size Of 150, 40% censoring

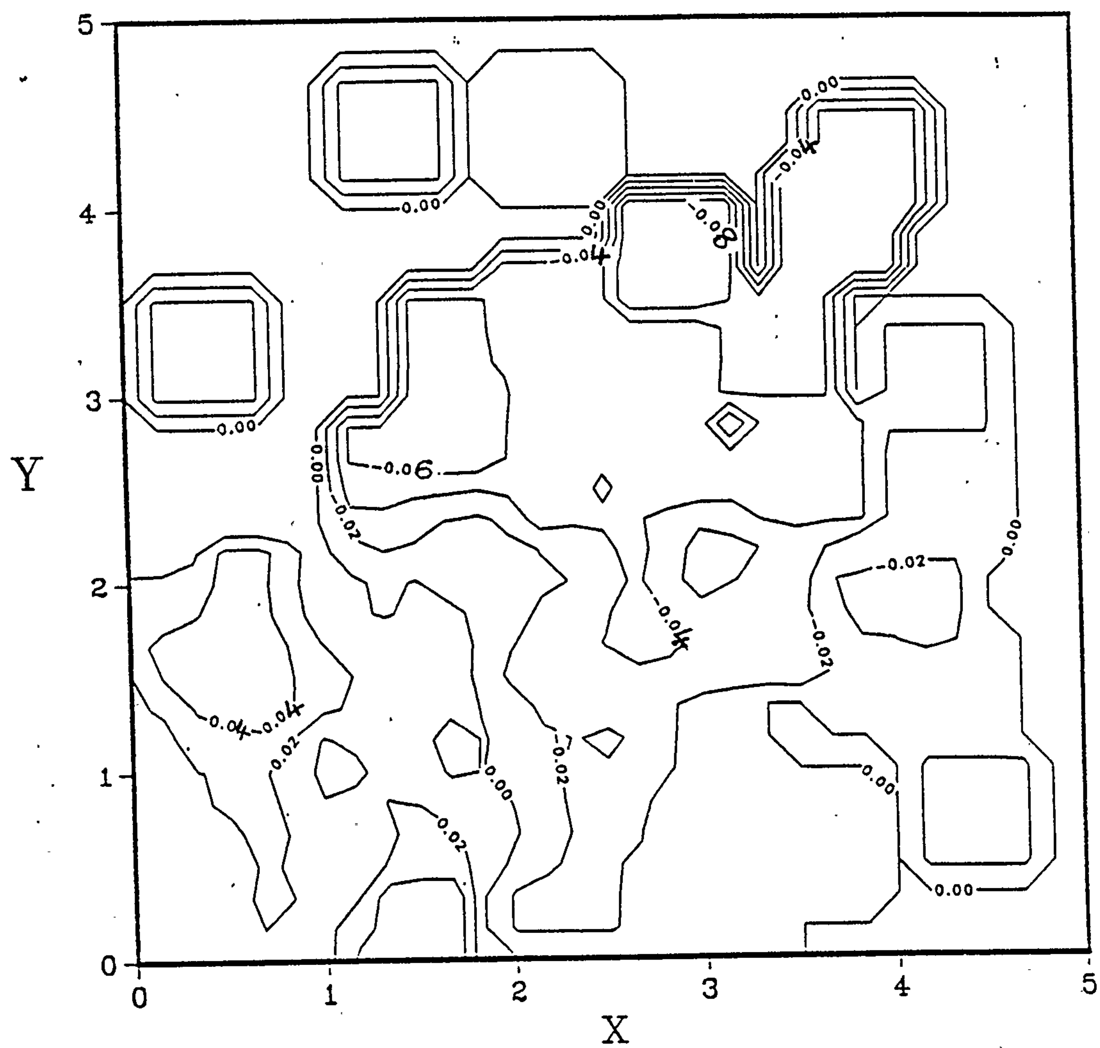

Figure 7.2.8 Difference Map for Estimator (4.2.19)
Sample Size Of 150, 40% censoring

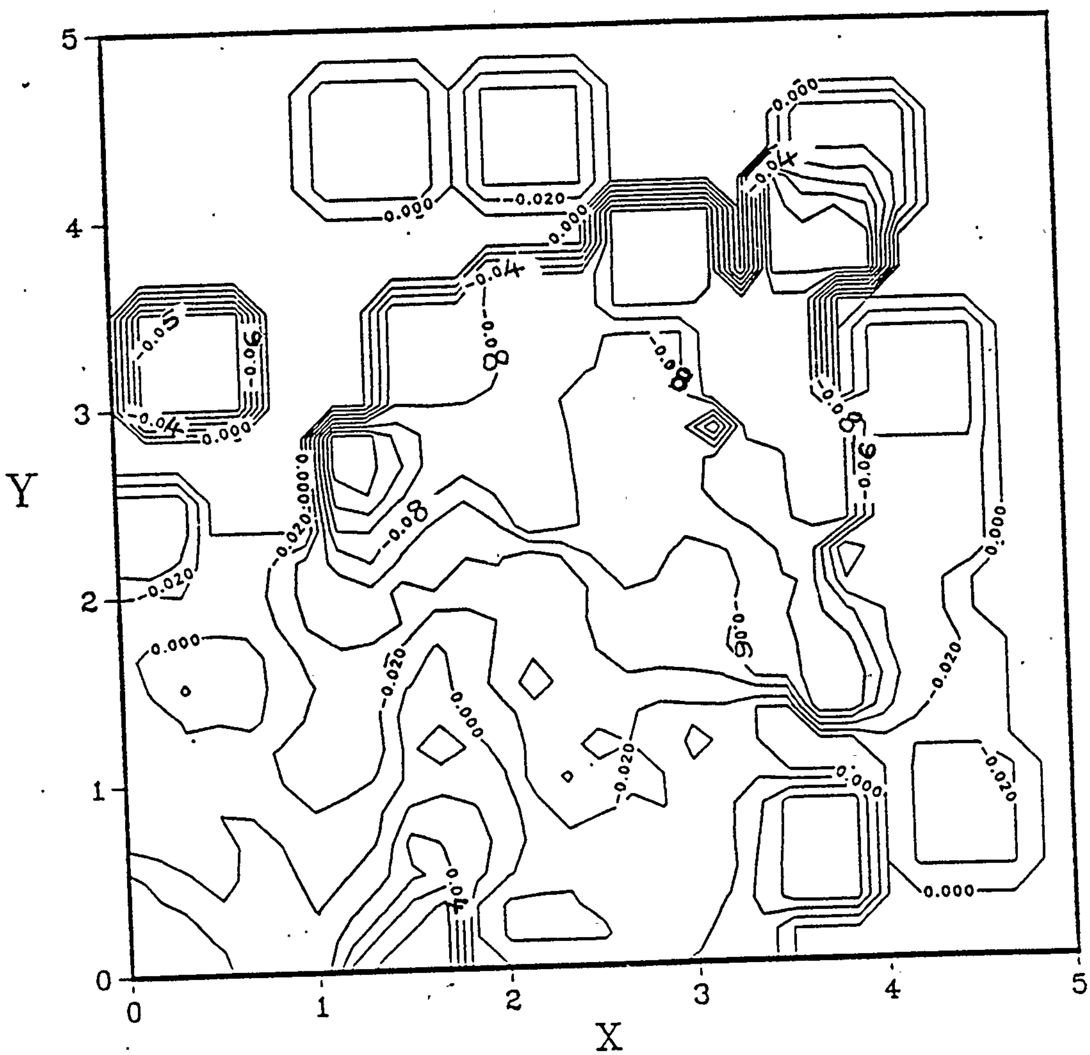

Figure 7.2.9 Difference Map For Estimator (5.1.3)
Sample Size Of 150, 40% censoring

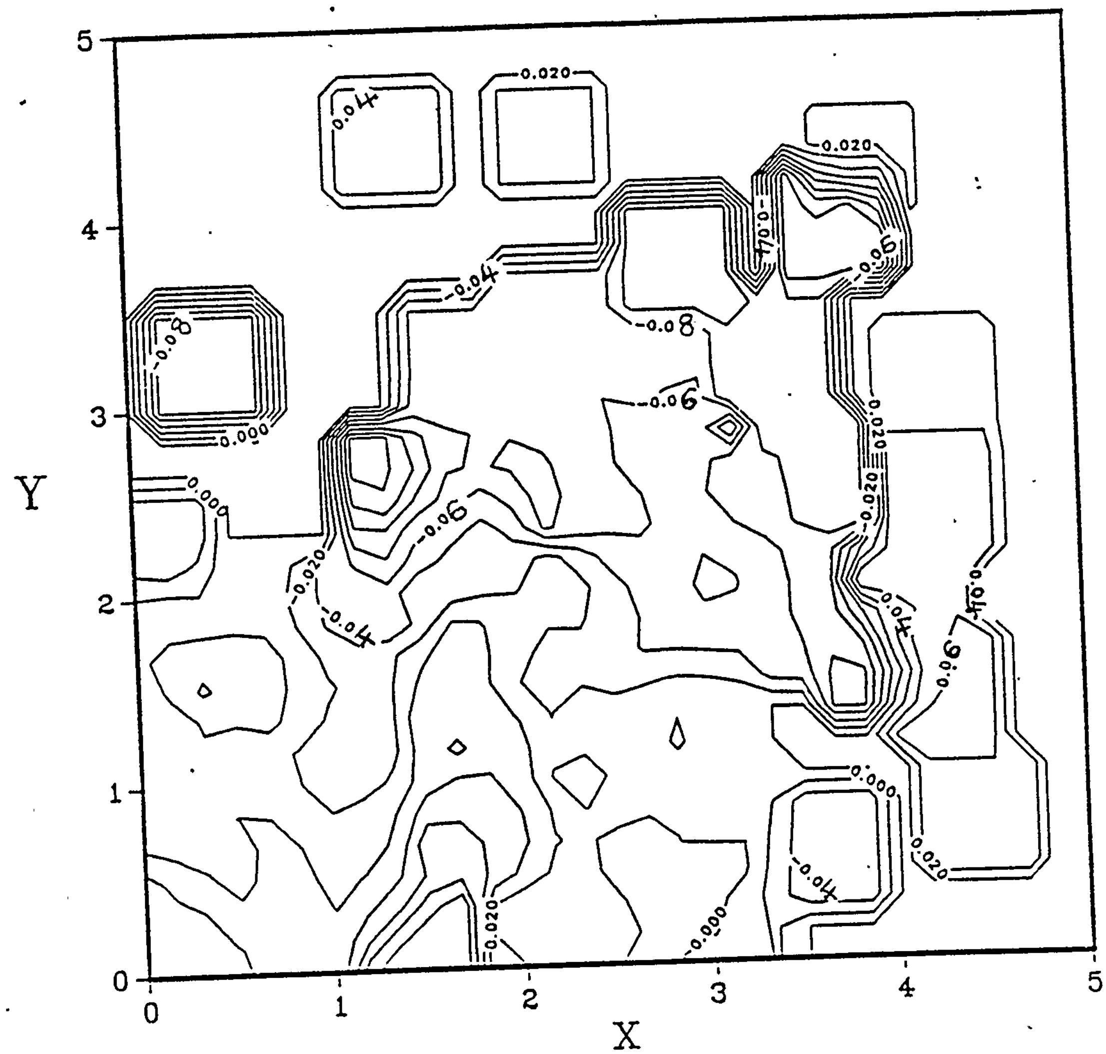

Figure 7.2.10 Difference Map For Estimator (5.2.2)
Sample Size Of 150, 40% censoring

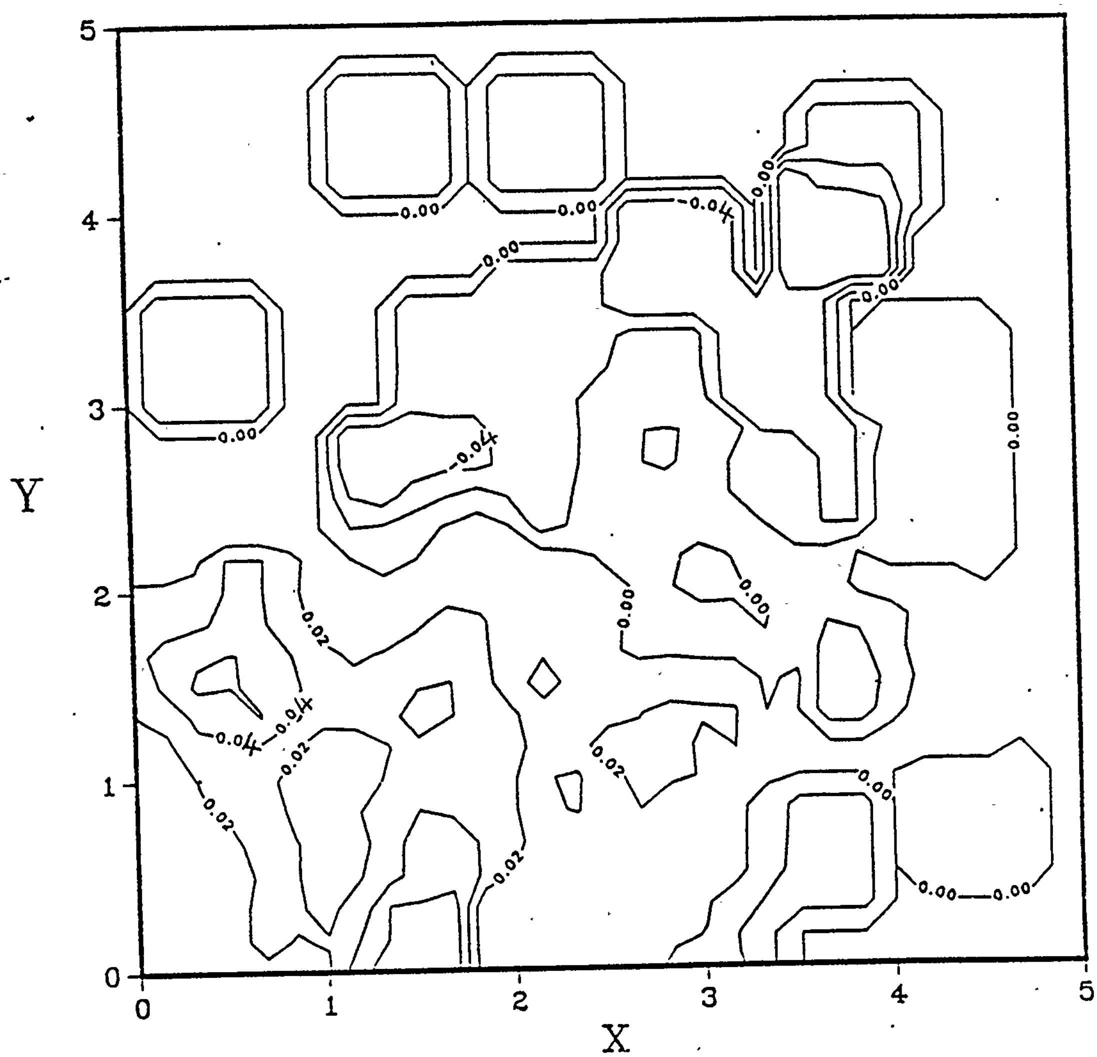

Figure 7.2.11 Difference Map For Estimator (6.3.2)
Sample Size Of 150, 40% censoring

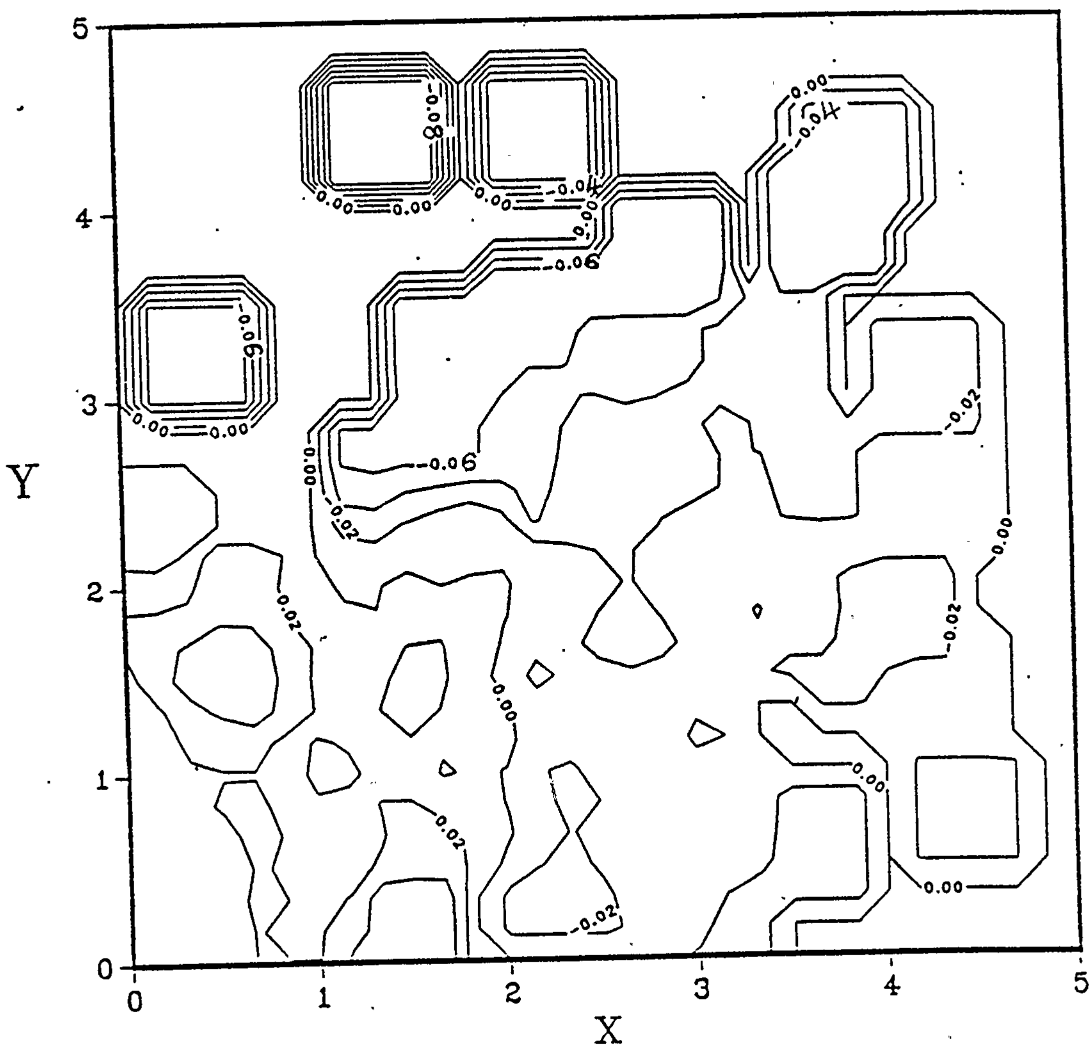

Figure 7.2.12 Difference Map For Estimator (2.2.3)
Sample Size of 150, 40% censoring

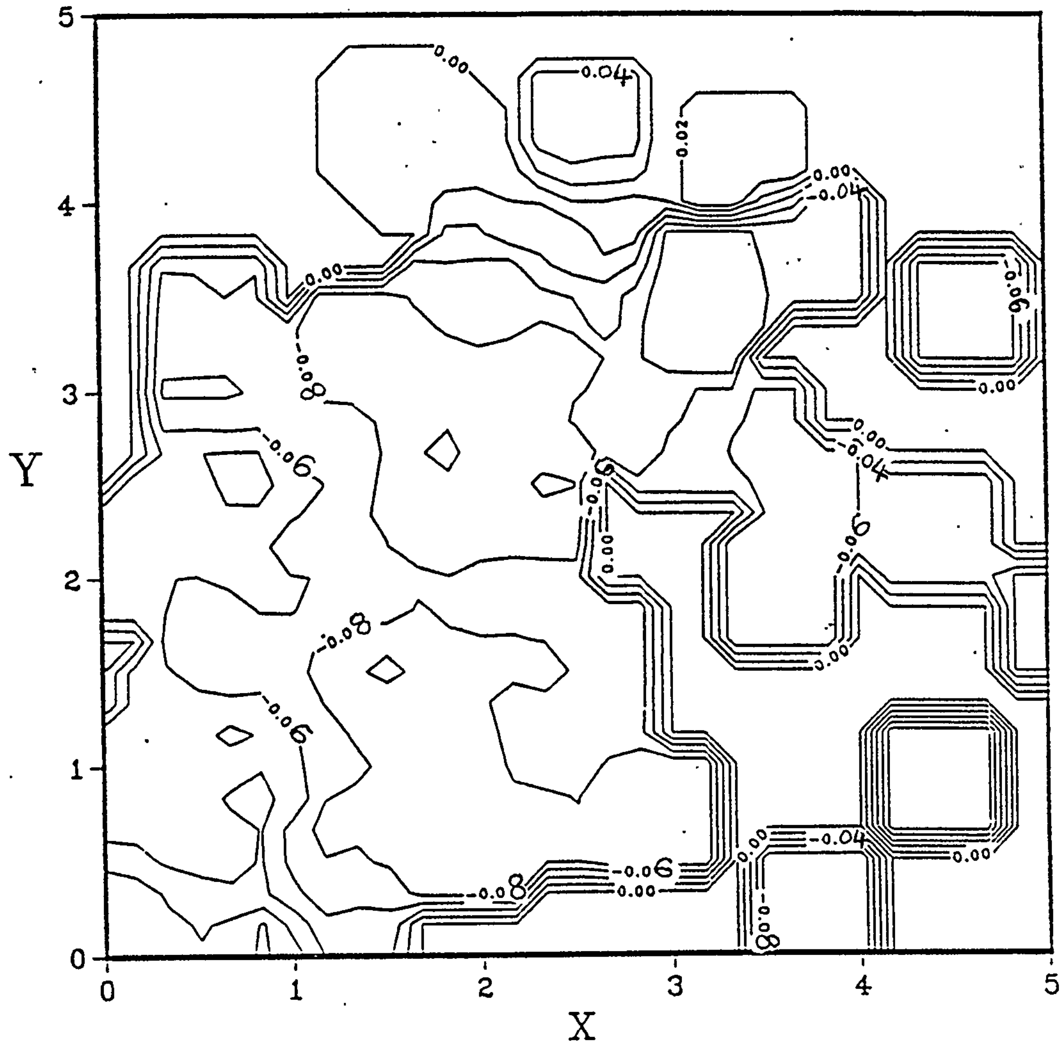

Figure 7.2.13 Difference Map For Estimator (4.1.8)
Sample Size Of 150, 50% censoring

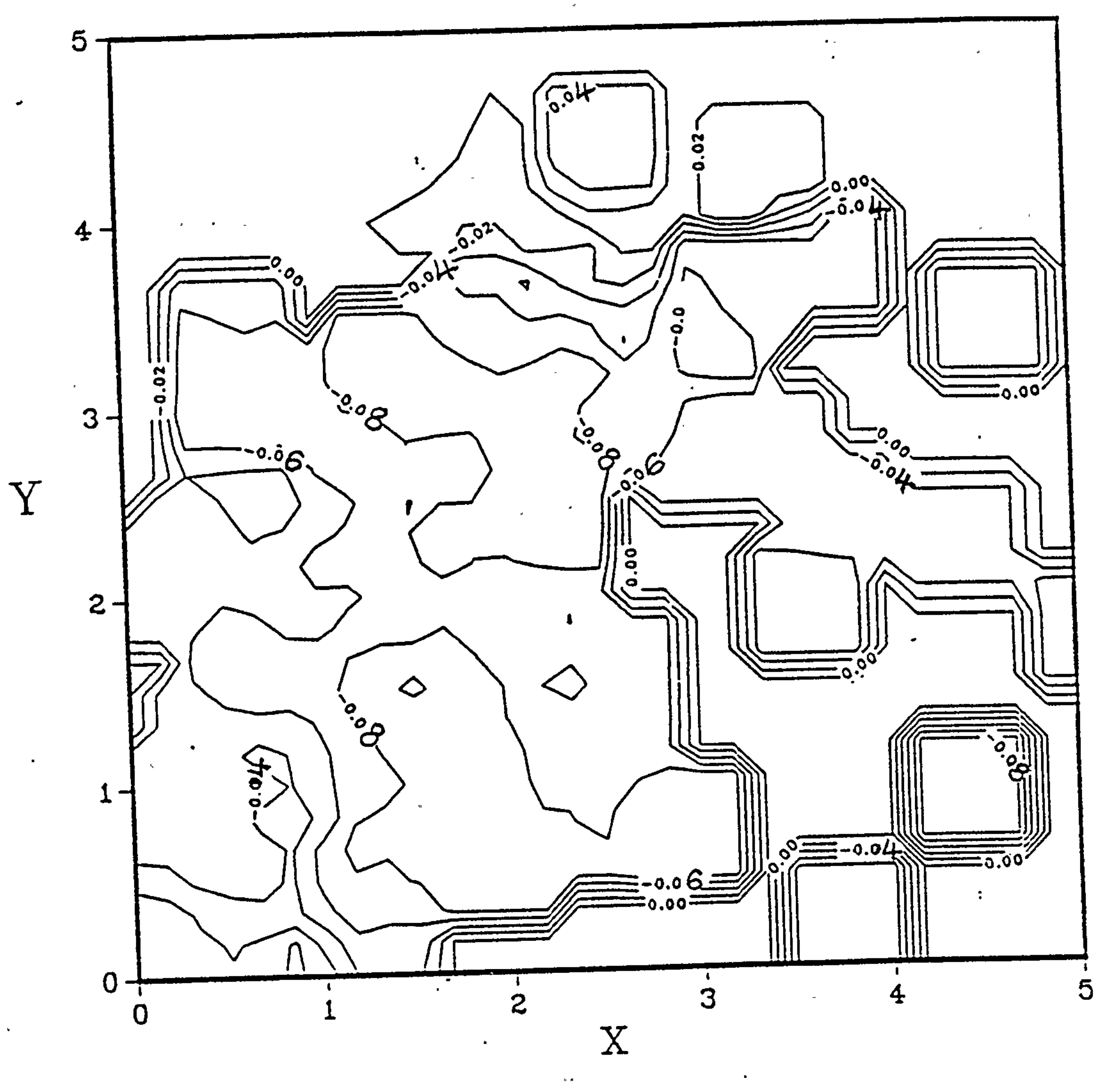

Figure 7.2.14 Difference Map For Estimator (4.2.19)
Sample Size Of 150, 50% censoring

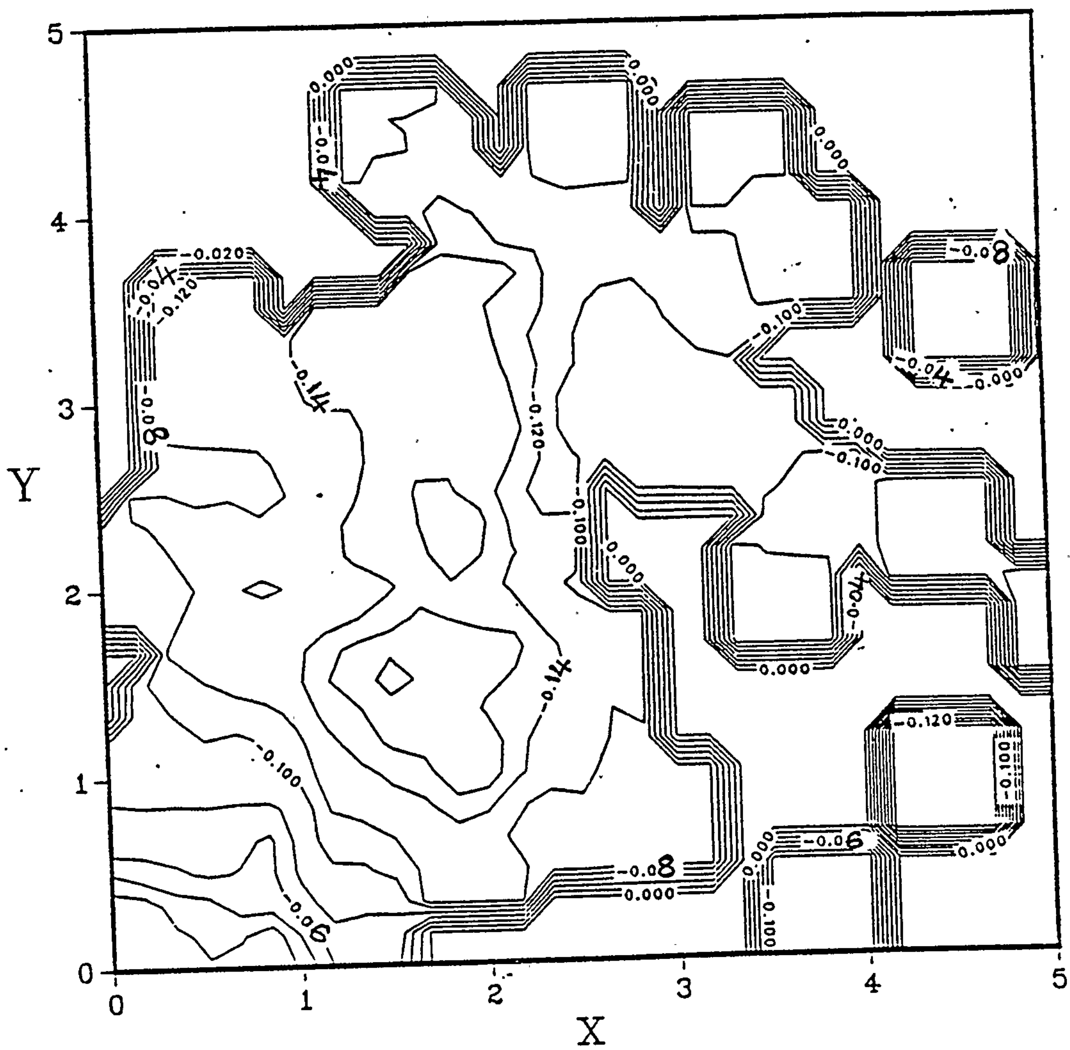

Figure 7.2.15 Difference Map For Estimator (5.1.3)
Sample Size Of 150, 50% censoring

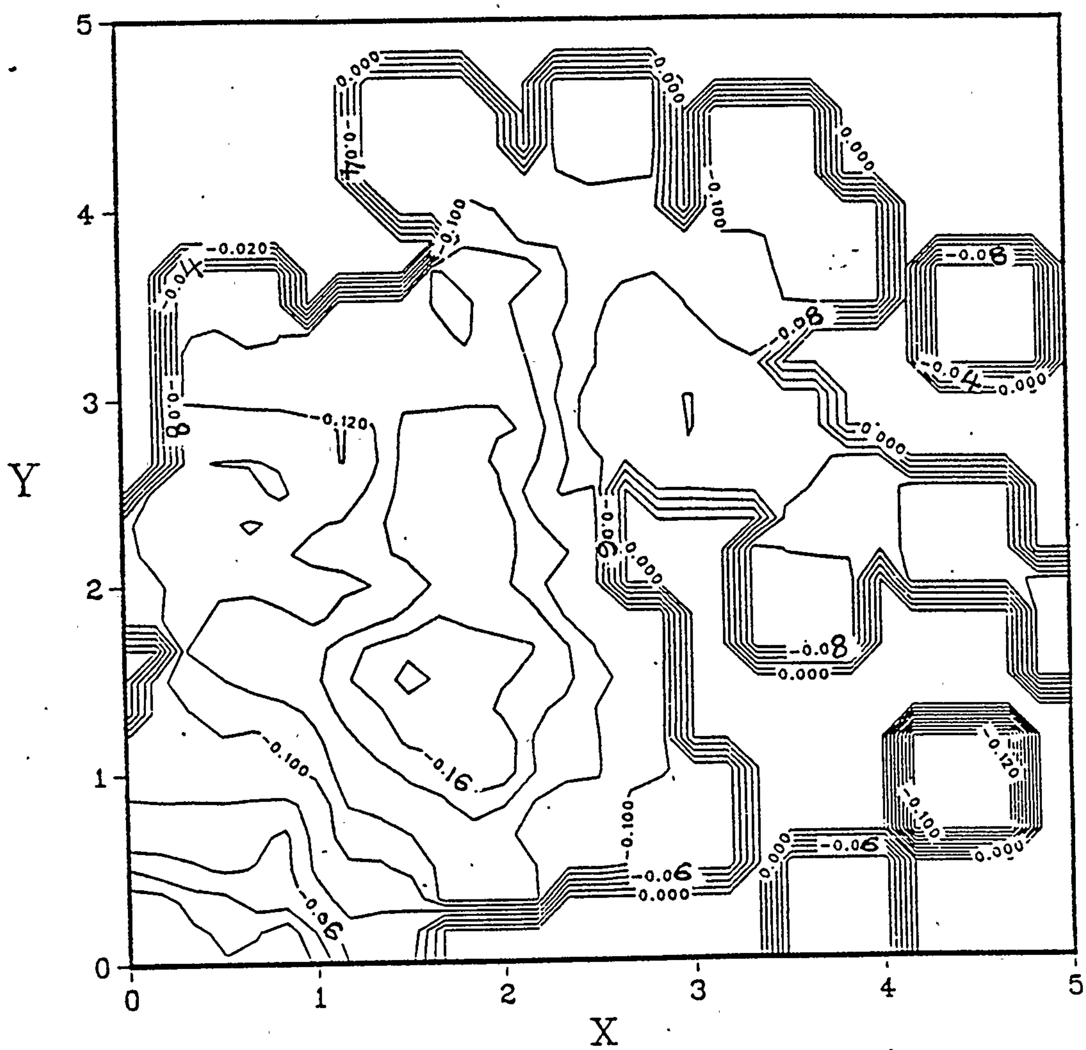

Figure 7.2.16 Difference Map For Estimator (5.2.2)
Sample Size Of 150, 50% censoring

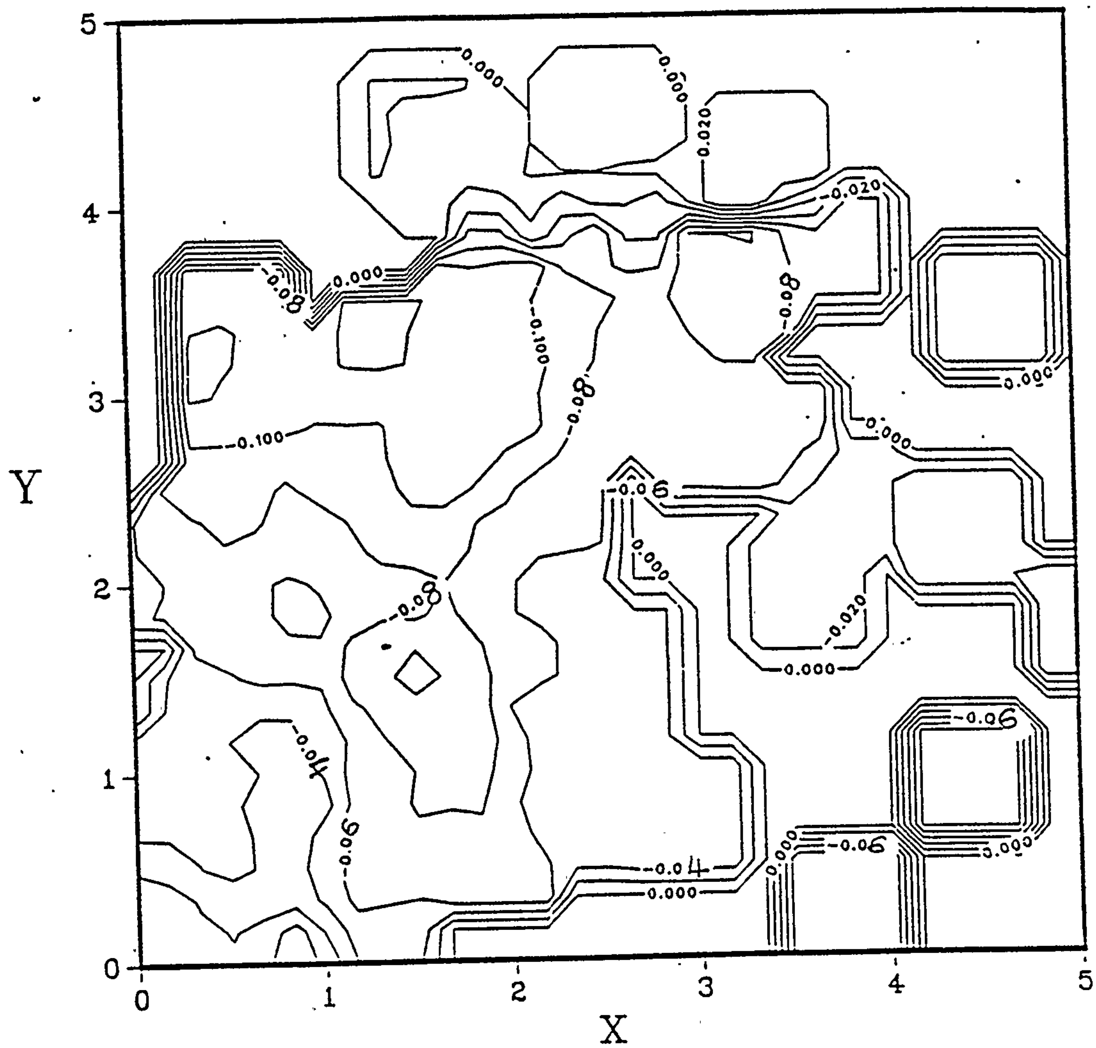

Figure 7.2.17 Difference Map For Estimator (6.3.2)
Sample Size Of 150, 50% censoring

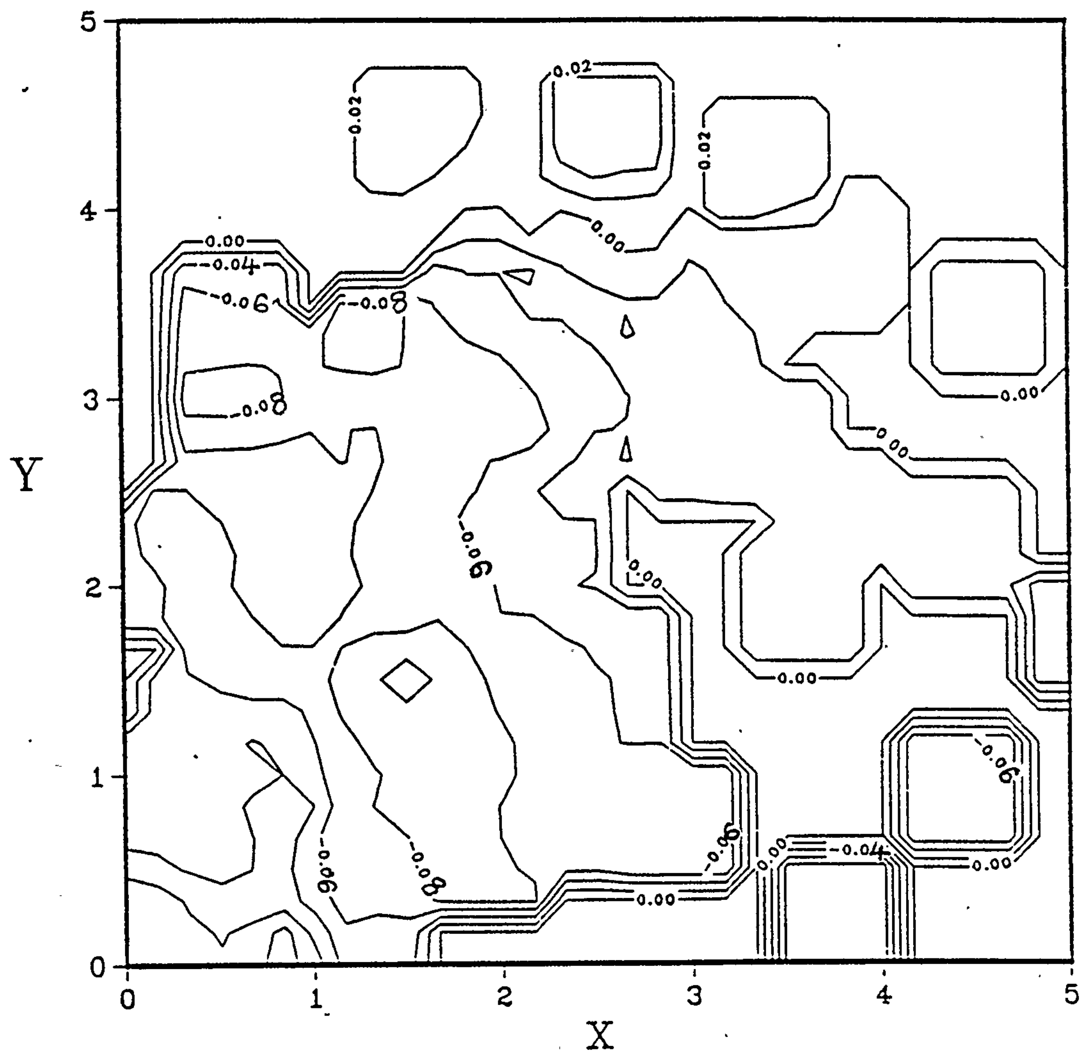

Figure 7.2.18 Difference Map For Estimator (2.2.3)
Sample Size Of 150, 50% censoring

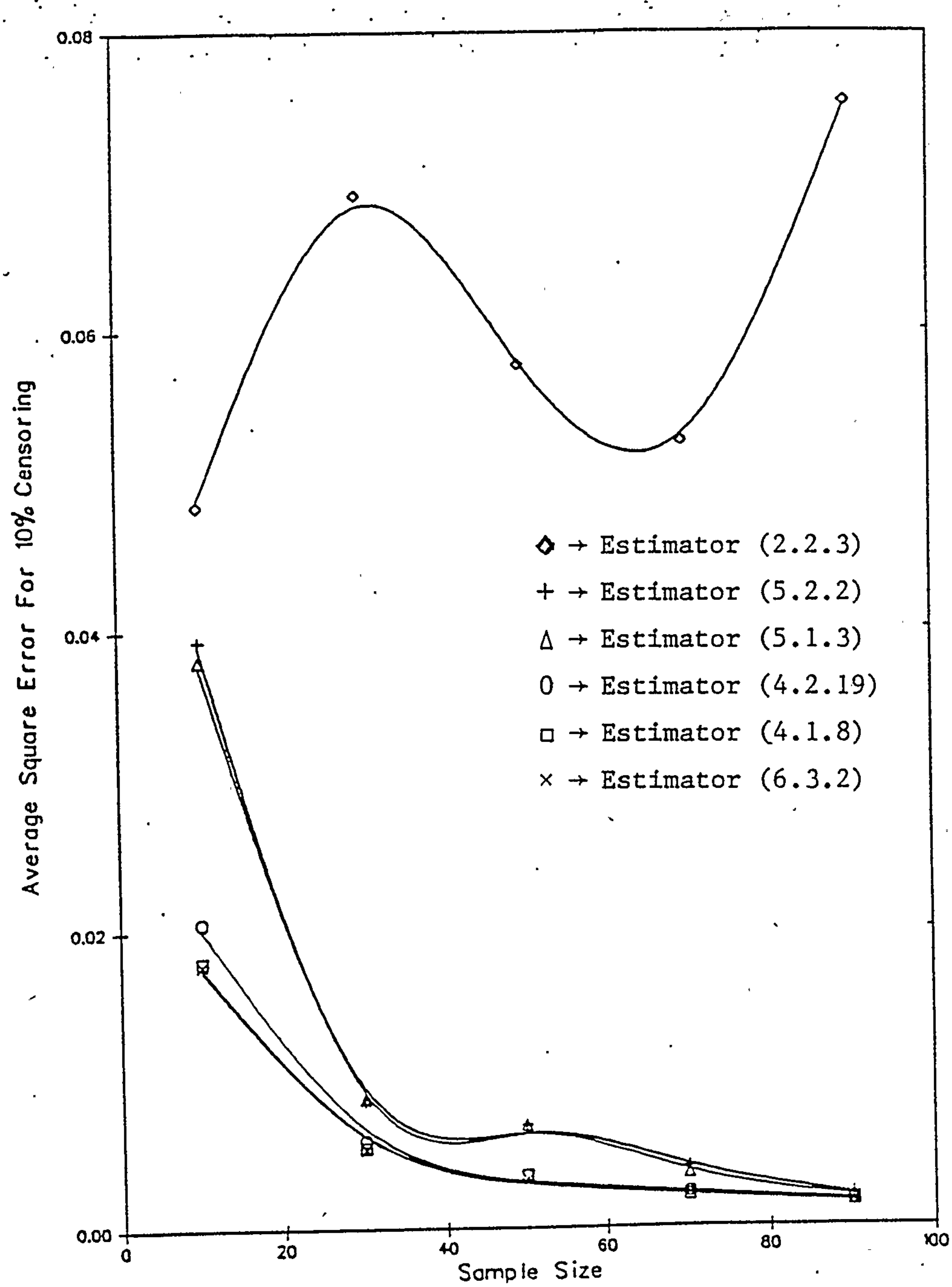

Figure 7.2.19 A Plot of the Data Given in Table 7.2.1

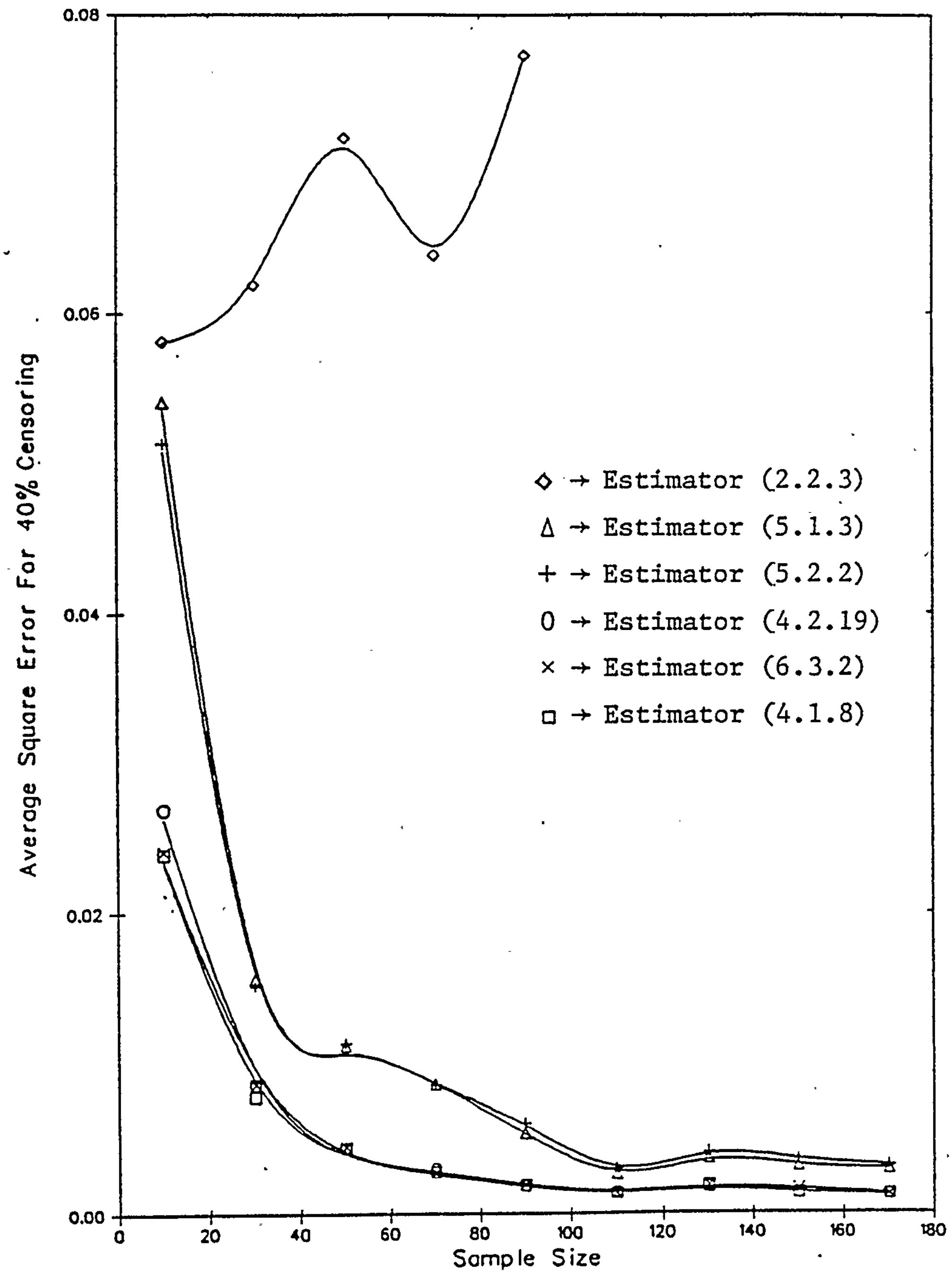

Figure 7.2.20 A Plot of the Data Given in Table 7.2.4

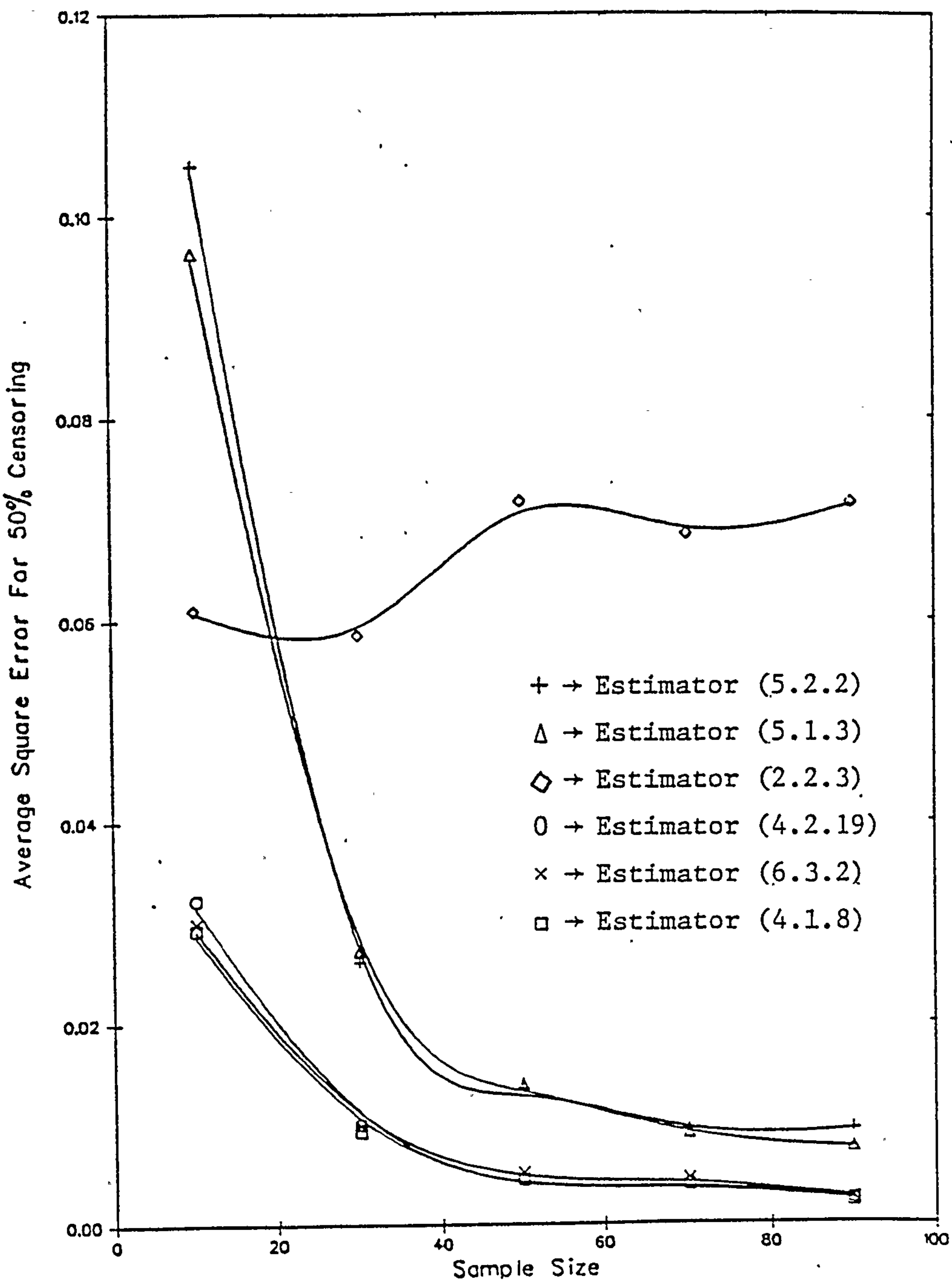

Figure 7.2.21 A Plot of the Data Given in Table 7.2.7

BIBLIOGRAPHY

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Unpublished manuscript, University of California, Berkeley.

Breslow, N., and Crowley, J. (1974). A large sample study of life table and product limit estimates under random censorship. Annals of Statistics, 2, 432-453.

Burke, M.D. (1988). Estimation of a bivariate distribution function under random censorship. Biometrika, 75, 2, 379-382.

Campbell, G., and Földes, A. (1982). Large-Sample properties of nonparametric bivariate estimators with censored data. Colloquia Mathematica Societatis János Bolyai 32, pp.103-21.

Campbell, G. (1981). Nonparametric bivariate estimation with randomly censored data. Biometrika, 68, 2, 417-22.

Cherian, K.c. (1941). A bivariate correlated gamma-type distribution function. Sankhya, 5, 141-54.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the E.M. algorithm. Journal of the Royal Statist. Society, Series B 39, 1-38.

Efron, B. (1967). The two sample problem with censored data. Proc. Fifth Berkeley Symp. Math. Statist. Prob., 4, 881-53.

Földes, A., Rejto, L., and Winter, B.B. (1979). Strong consistency properties of nonparametric estimators for randomly censored data. Periodica Math. Hung. 11 (3), 233-50.

Freund, J.E. (1961). A bivariate extension of the exponential distribution. Journal of the American Statistical Assoc., 56, 971-77.

Gumbel, E.J. (1960). Bivariate exponential distribution. Journal of Amer. Statist. Assoc. 55, 698-707.

Hanley, J.A., and Parnes, M.N. (1983). Nonparametric estimation of a multivariate distribution in the presence of censoring. Biometrics 39, 129-139.

Johnson, N.L., and Kotz, S. (1972). Distributions in Statistics : Continuous multivariate distribution. Wiley, New York.

Kalbfleisch, J.D., and Prentice, R.L. (1980). The statistical analysis of failure time data. Wiley, New York.

Kaplan, E.L., and Meier, P. (1958). Nonparametric estimation from incomplete observations. American Statistical Assoc. Journal, June 1958, 457-481.

Kiefer, J. (1961). On large deviations of the empirical d.f. of vector chance variables and a law of the iterated logarithm. Pacific J. of Math., 11 (1961), 649-660.

Korwar, R.M., and Dahiya, R.C. (1982). Estimation of bivariate distribution function from incomplete observations. Commun. Statist. - Theor. Meth., 11(8), 887-897.

Mardia, K.V. (1970). Families of bivariate distributions. Griffin's Statistical Monographs and Courses.

Marshall, A.W., and Olkin, I. (1967). A multivariate exponential distribution. J. Am. Statist. Assoc. 62, 30-44.

Marshall, A.W. (1975). Some comments on the hazard gradient. Stochastic processes and their applications, 3, 293-300.

Munoz, A. (1980). Nonparametric estimation from censored bivariate observations. Technical Report No. 60, Division of Biostatistics, Stanford University.

Peterson, A.V. (1977). Expressing the Kaplan-Meier estimate as a function of empirical subsurvial functions. Journal of the American Statisticval Association 72, 854-858.

Rosenblatt, M. (1956). "Remarks on some nonparametric estimates of a density function". Ann. Math. Statist., 27, 832-837.

Tapia, R.A. and Thompson, J.R. (1978). Nonparametric probability density estimation. John Hopkins Series in the Mathematical Sciences. Number 1.

Tsai, W., Leurgans, S., and Crowley, J. (1986). Nonparametric estimation in the presence of censoring. The Annals of Statistics, Vol. 14, No. 4, 1351-1365.

Turnbull, B.W. (1976). the emprical distribution function with arbitrarily grouped, censored and truncated data. Journal of the Royal Statistical Society, Series B. 38, 290-295.

Turnbull, B.W., and Weiss, L. (1978). A likelihood ratio statistic for testing goodness of fit with randomly censored data. Biometrics 34, 367-375.

Wegman, E.J. (1972). Nonparametric density estimation. J. Statist. Comput. Simul., Vol.1, 225-245.