AN ABSTRACT OF THE THESIS OF

KENNETH BRUCE McRAE for the DOCTOR OF PHILOSOPHY
(Name) (Degree)

in STATISTICS presented on October 26, 1971
(Major) (Date)

Title: INFERENCE PROCEDURES FOR PAIRS OF DISTRIBUTIONS WITH PROPORTIONAL FAILURE RATE FUNCTIONS

Redacted for privacy

Abstract approved: ______________________
David R. Thomas

Some nonparametric maximum likelihood estimation procedures are developed for the class of pairs of distributions which have proportional failure rate functions. Special consideration is given to the case in which the shape of the failure rate functions are assumed to be either increasing or decreasing. Estimators of the proportionality constant, of the reliability functions, and of the failure rate functions are derived.

A Monte Carlo study using Weibull distributions provides a basis for comparing the various estimators. An estimator of the proportionality constant, based on the distribution of the rank order statistics, is found to be "best" on the basis of minimum MSE. For a given estimate of the proportionality constant, observations from both samples can be combined to estimate either reliability function. Such combined-sample estimators are shown to have smaller MSE

than the appropriate single-sample empirical estimator.

Another Monte Carlo study, using Weibull distributions, provides a basis for comparing several statistics for testing the adequacy of the simple proportionality model. Some test statistics, based on a large sample procedure proposed by Professor David Cox, are found to have good small sample properties.

Inference Procedures for Pairs of Distributions with Proportional Failure Rate Functions

by

Kenneth Bruce McRae

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

June 1972

APPROVED:

Redacted for privacy

Assistant Professor of Statistics

in charge of major

Redacted for privacy

Acting Chairman of Department of Statistics

Redacted for privacy

Dean of Graduate School

Date thesis is presented October 26, 1971

Typed by Clover Redfern for Kenneth Bruce McRae

ACKNOWLEDGMENT

The author appreciates the advice and assistance given freely by Dr. David Thomas during the preparation of this dissertation. Dr. Thomas gave a major contribution to the author's graduate studies by means of nine good courses and hundreds of informal discussions on various aspects of statistical theory.

Also, the author has benefited from advice given by all faculty members of the Department of Statistics at Oregon State University. The problem to which the dissertation addresses itself stems from some work by Dr. Don Pierce and Dr. Thomas, and their subsequent discussion with Professor Ronald Pyke. Many faculty members have contributed during the development of the thesis; in particular, Dr. Joseph Bowman's suggestion that part of the work be placed in a nonlinear programming framework is appreciated, and also, the comments by other members of the author's committee: Doctors Brunk, Mullooly, and Petersen are well received. The project could not have been undertaken without the generous financial support provided by the Computer Center at Oregon State University. To all those who have given assistance, criticism, and support the author is extremely grateful.

Dedication

This dissertation is dedicated to those who taught me to value education: Murdina, my family, my friends and my teachers.

TABLE OF CONTENTS

LIST OF TABLES

Table		Page

INFERENCE PROCEDURES FOR PAIRS OF DISTRIBUTIONS WITH PROPORTIONAL FAILURE RATE FUNCTIONS

1. INTRODUCTION

The intent of this paper is to establish some statistical procedures associated with the class of pairs of distributions $\mathcal{L}_e = \{(F, G): G = 1 - (1-F)^{\theta},\ \theta > 0\}$, which is closely related to the nonparametric alternative to the nonparametric hypothesis $H_0: F = G$, proposed by Lehmann in 1953. $\mathcal{L}_e$ is one of the one-parameter family of nonparametric classes known as the Lehmann alternatives. Although Lehmann establishes the use of nonparametric classes of alternatives to compare the power of nonparametric tests as early as 1953, the principal use of the Lehmann alternatives continues to be for the analysis of nonparametric tests. Other properties of distributions $(F, G) \in \mathcal{L}_e$, however, may be useful to a statistician considering nonparametric methods of estimation. We review the established properties of $\mathcal{L}_e$ by following Shorack (1967), who gives a good summary of the nature of this class.

(i) An interesting interpretation is obtained when θ is a positive integer, say k, in which case $G(x) = 1 - (1-F(x))^k$ will be the distribution function of the minimum of a random sample of size k from distribution F.

(ii) The class $\mathcal{L}_e$ contains a host of pairs of parametric distributions, since F may be any parametric distribution and G is the distribution defined by the relationship $G(x) = 1 - (1-F(x))^k$. In most cases, the function $G(x)$ would be mathematically complex and would not belong to the particular parametric family as F; however, if F is the Weibull distribution defined by $F(x) = 1 - e^{-\beta(x-\xi)^\alpha}$, $x \geq \xi$, $\alpha > 0$, $\beta > 0$, then $G(x) = 1 - e^{-\theta\beta(x-\xi)^\alpha}$ is also a Weibull distribution. Another case in which F and G belong to the same parametric family is given by the modified extreme value distribution $F(x) = 1 - e^{-\alpha(e^{\gamma x}-1)}$, $\alpha > 0$, $x \geq 0$, $\gamma > 0$. Shorack notes that other less well known pairs $(F, G) \in \mathcal{L}_e$ can be found in Dubey (1965a, 1965b).

(iii) If $(F, G) \in \mathcal{L}_e$ then for $\theta > 1$, $(0 < \theta < 1)$ the distribution G is stochastically greater (less) than the distribution F.

(iv) The following characterization theorem is given by Shorack (1967): For random variables X and Y with absolutely continuous distribution functions F and G, then $G = 1 - (1-F)^\theta$, $\theta > 0$ if and only if there exists a strictly increasing continuous transformation $g(\cdot)$ and constants θ_1 and θ_2, $\theta_1/\theta_2 = \theta$, such that $X' = g(X)$ and $Y' = g(Y)$ are exponentially distributed with parameters θ_1 and θ_2.

(v) Let F be an absolutely continuous distribution with density f. The failure rate function (or hazard function) is defined by $r_F(x) = f(x)/(1-F(x))$ for $F(x) < 1$. It is proved in Allen (1963), Shorack (1967), and Nádas (1970) that if densities f and g exist, then a necessary and sufficient condition for $(F, G) \in \mathcal{L}_e$ is that $r_G(x) = \theta r_F(x)$ on the support of the distributions.

Nádas uses the following counter-example to show that this result does not hold for discrete distributions. Let $0 < p < 1$ and $0 < kp < 1$. Let $P(X=x) = p(1-p)^x$ and $P(Y=x) = kp(1-kp)^x$, $x = 0, 1, 2, \ldots$. Then $kp = r_G(x) = kr_F(x)$, but $P(Y \geq x) = (1-kp)^x \neq (1-p)^{kx} = (P(X > x))^k$ when $x = 1, 2, \ldots$ unless $k = 1$.

(vi) The following result appears in two different forms. Allen (1963) and Thomas (1969) independently prove minor variations of the following proposition: Let X and Y be independent random variables having continuous distributions with common support, then $P(X < Y | X < t)$ is independent of t if and only if $r_G(t) = \theta r_F(t)$. Allen's result is for n independent random variables, $n \geq 2$.

Nádas proves the related result: Let X and Y be independent random variables with distributions F and G being either purely discrete or absolutely continuous. Then $P(t \leq X | X \leq Y) = P(t \leq Y | Y \leq X)$, $-\infty < t < \infty$, if and only if $r_G(t) = \theta r_F(t)$. Nádas describes this proposition as: The minimum of independent random variables X and Y is independent of the event $X < Y$ if and only if the hazard functions of X and Y are proportional. The relationship between the two forms may be established by an elementary probability argument.

(vii) Van der Laan (1970) carries out an extensive analytical and numerical study of the behavior of moments and densities of $F^k(x)$ and $1 - (1-F(x))^k$ for various classes of distribution functions $F(x)$.

Property (v) is the genesis of this paper. A frequent criticism of nonparametric estimation is that, since they allow for every contingency, the estimators tend to be too conservative in comparison to parametric estimators. In certain cases, however, one may be able to use a priori information to reduce the admissible class of estimators. Such is the case in the single sample problem of estimating the failure rate function, subject to the assumption that the failure rate is monotone. Marshall and Proschan (1965) show that the monotone maximum likelihood estimator is closer to the failure rate function, with respect to a certain metric, than the unconstrained maximum likelihood estimator. It is reasonable to expect other nonparametric estimators to be improved, in some sense, by using assumed properties of the failure rate function or distribution function. For the case of two samples, one is able to use a priori information about the functional relationship between the failure rate functions to construct a combined-sample estimator of each failure rate function. This study considers some inference procedures based on the assumption that the failure rate functions are proportional.

We expect that in many experimental situations it would be natural to assume the proportionality of failure rate functions. An example is the situation where one is considering mortality-type data in which two random samples are drawn from the same population, one of which is used as a placebo and the other being subjected to a treatment. The simplest time-dependent-relationship between the two failure rate functions is proportionality; provided that this assumption is valid, the constant of proportionality (θ) will reflect the effect of the treatment, being greater than one if the treatment increases the failure rate, and less than one if the treatment decreases the failure rate. In such an experimental situation, one can obtain estimates of θ as well as nonparametric estimators of each failure rate function based on the combined sample, and consequently, nonparametric estimators of each reliability and distribution function.

Cox (1971) generalizes the work of Kaplan and Meier (1958) by incorporating regression-like arguments into life-table analysis. Each individual on trial is assumed to have available observations on one or more explanatory variables. The failure rate function is taken to be a function of the explanatory variables and unknown regression coefficients multiplied by an arbitrary and unknown function of time. By conditioning upon the risk set (the individuals on test) at the failure times, Cox gives the conditional probability that individual j fails at time t_j. Conditioning on these probabilities, a likelihood function is obtained, from which inference procedures about the unknown regression coefficients are given. Good use is made of likelihood ratio procedures in the inference section.

Cox's paper is closely related to this dissertation; Cox gives a general framework for the study of distributions having structurally related failure rate functions. The case of censored data is considered. Methods of estimation and test procedures stem naturally from the conditional likelihood function. For complete data, Cox's estimator of the proportionality constant (θ) coincides with our estimator $\hat{\theta}_{MI}$. We use Cox's methods in Chapter 4 where we consider tests of proportionality. Our work concentrates on the method of maximum likelihood for the simple proportionality model. Many of the potential generalizations mentioned by Cox are contained in our work.

We consider only the method of maximum likelihood and restrict our attention to the case in which the distributions are absolutely continuous. In Chapter 2 we derive the nonparametric maximum likelihood estimators of θ, the failure rate functions, and the reliability functions. A Monte Carlo study of the estimators is evaluated in Chapter 3.

In Chapter 4 we consider the problem of testing the assumption that the failure rate functions are proportional. Only one new test is described but an exposition of three other proposed tests is given.

Graphical prodecures are also discussed. A Monte Carlo study of the test statistics is evaluated in Chapter 5.

A summary is given in Chapter 6.

2. DEVELOPMENT OF ESTIMATORS

2.1. Introduction to Nonparametric Maximum Likelihood Estimation

Given two random samples $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ from respective distribution functions F and G belonging to the parametric family $\mathcal{L}_e = \{(F,G): G = 1 - (1-F)^\theta,\ \theta > 0\}$, we derive nonparametric maximum likelihood estimators for θ, $\bar{F}$, $\bar{G}$ (the reliability functions associated with F and G) and for the failure rate functions r_F and r_G. We consider four classes which are associated with $\mathcal{L}_e$. Let $\mathcal{L} = \{(F,G)$: F and G are absolutely continuous c.d.f.'s $\}$, $\mathcal{L}_C = \{(F,G): (F,G) \in \mathcal{L}$ and $G = 1 - (1-F)^\theta$, $\theta > 0\}$, $\mathcal{L}_{IFR} = \{(F,G): (F,G) \in \mathcal{L}_C$ and F and G have increasing failure rate$\}$, and $\mathcal{L}_{DFR} = \{(F,G): (F,G) \in \mathcal{L}_C$ and F and G have decreasing failure rate$\}$. Clearly $\mathcal{L}_{IFR} \subset \mathcal{L}_C \subset \mathcal{L}$, $\mathcal{L}_{DFR} \subset \mathcal{L}_C \subset \mathcal{L}$ and $\mathcal{L}_C \subset \mathcal{L}_e$.

The classes $\mathcal{L}$, $\mathcal{L}_C$, $\mathcal{L}_{IFR}$, $\mathcal{L}_{DFR}$ provide the first of two underlying concepts for the discussion of nonparametric maximum likelihood estimation of θ. Each of these classes provide a different parameterization of the likelihood function and the resulting estimators are generally different. It is necessary then, to refer to a nonparametric MLE of θ with respect to a particular class or parameterization; for example, $\hat{\theta}_{IFR}$ and $\hat{\theta}_{DFR}$ refer to MLE of θ with respect to $\mathcal{L}_{IFR}$ and $\mathcal{L}_{DFR}$, respectively.

The second underlying concept is the meaning of nonparametric maximum likelihood estimation. Since the distributions F and G are both unknown, the value of $F(\cdot)$ and of $G(\cdot)$ at the order statistics are considered nuisance parameters when estimating θ. The invariance procedure used in Section 2.2 removes the difficulty of having to estimate these functions at order statistics. However, in other procedures we do estimate the distribution function F at the order statistics.

For the single sample problem, Kaplan and Meier (1958) say of nonparametric estimation:

> Most general methods of estimation, such as maximum likelihood or minimum chi-square, may be interpreted as prodecures for selecting from an admissible class of distributions one which, in a specified sense, best fits the observations. To estimate a characteristic (or parameter) of the true distribution one uses the value that the characteristic has for this best fitting distribution function. It seems reasonable to call an estimation procedure nonparametric when the class of admissible distributions from which the best fitting one is to be chosen is the class of all distributions.

There are two difficulties involved in applying this definition of nonparametric estimation. If the class of all distributions is considered for maximum likelihood estimation, then there is no sigma-finite measure relative to which all distributions are absolutely continuous. Kiefer and Wolfowitz (1956, p. 893) propose a generalization of the maximum likelihood concept which is used by Barlow (1968) whose definition we use. For some class of distributions $\mathcal{F}$, let $F_1, F_2 \in \mathcal{F}$, and let $f(\cdot\,; F_1, F_2)$ denote the Radon-Nikodym derivative of F_1 with respect to the measure induced by $F_1 + F_2$. $\hat{F}$ is called the maximum likelihood estimate relative to $\mathcal{F}$ if $\hat{F}$ satisfies

$$\sup_{F \in \mathcal{F}} \Pi_{i=1}^{n} [f(X_i; F, \hat{F})\{1 - f(X_i; F, \hat{F})\}^{-1}] = 1,$$

where $X = (X_1, X_2, \ldots, X_n)$ is a random sample. Barlow notes that this definition coincides with the usual definition when the family is dominated by a sigma-finite measure. For example, if $\mathcal{F}$ is the class of discrete distributions, then consider $\tilde{F}$ which places mass $\tilde{p}_i$ at each of the order statistics X_i, $i = 1, \ldots, n$, where $\Sigma_{i=1}^{n} \tilde{p}_i = 1$. Now consider an arbitrary distribution $F \in \mathcal{F}$ which places mass p_i at each X_i, $i = 1, \ldots, n$, $\Sigma_{i=1}^{n} p_i \leq 1$. Then

$$f(X_i; F, \tilde{F}) = p_i/(p_i + \tilde{p}_i)$$

so that

$$f(X_i; F, \tilde{F})/\{1 - f(X_i; F, \tilde{F})\} = p_i/\tilde{p}_i.$$

For $\tilde{F}$ to be the maximum likelihood estimator w.r.t. $\mathcal{F}$, $\tilde{F}$ must be such that

$$\sup_{F \in \mathcal{F}} \Pi_{i=1}^{n} (p_i/\tilde{p}_i) = 1,$$

or equivalently,

$$\Pi_{i=1}^{n}\tilde{p}_i = \sup_{F \in \mathcal{F}} \Pi_{i=1}^{n} p_i .$$

Thus the estimator of the generalized definition coincides with the usual MLE.

As an application of Kaplan and Meier's definition, suppose we use the method of maximum likelihood to select a distribution from the admissible class $\mathcal{F}_{NS}$, where $\mathcal{F}_{NS}$ is the class of distribution functions which can be decomposed into a discrete component and an absolutely continuous component. Then for a given random sample $X_1, \ldots, X_n$ we consider distributions

$$F(t) = \int_{-\infty}^{t} f(x)dx + \Sigma_{i=1}^{n} p_i I_{(-\infty, t]}(X_i),$$

where $f(x)$ is the density function associated with the absolutely continuous component and $I_{(-\infty, t]}(\cdot)$ is the indicator function. The maximum likelihood estimator, of course, places mass $\hat{p}_i$ at each order statistic X_i and $\Sigma_{i=1}^{n}\hat{p}_i = 1$. Then for an ϵ-neighborhood about X_i we have the relationship,

$$2\epsilon f(X_i) + p_i \doteq \int_{X_i-\epsilon}^{X_i+\epsilon} f(t;F,\hat{F})d(F+\hat{F})$$

$$= \int_{X_i-\epsilon}^{X_i+\epsilon} f(t;F,\hat{F})f(t)dt + (p_i+\hat{p}_i)f(X_i;F,\hat{F})$$

if the Radon-Nikodyn derivative of F w.r.t. the measure induced by $F + \hat{F}$ is defined by

$$f(t; F, \hat{F}) = 1 \quad \text{if} \quad t \neq X_i$$

$$= \frac{p_i}{p_i + \hat{p}_i} \quad \text{if} \quad t = X_i .$$

That is, the right hand side equals $\int_{X_i - \epsilon}^{X_i + \epsilon} f(t)dt + p_i$.

For $\hat{F}$ to be the MLE with respect to $\mathcal{F}_{NS}$, the vector $\hat{\underline{p}}$ must satisfy

$$\Pi_{i=1}^{n} \hat{p}_i = \sup_{F \in \mathcal{F}_{NS}} \Pi_{i=1}^{n} p_i .$$

Kaplan and Meier prove their estimator of F does achieve this supremum, and therefore, in the sense of the generalized definition of Keifer and Wolfowitz, is called the maximum likelihood estimator of F with respect to $\mathcal{F}_{NS}$.

We also adopt Kiefer and Wolfowitz's generalized definition of maximum likelihood estimators (MLE) and shall refer to nonparametric likelihood functions of the form $L(\underline{X} | F) = \Pi_{i=1}^{n} \{F(X_i) - F(X_i + 0)\}$ as likelihood functions, even if there does not exist a dominating sigma-finite measure.

The second difficulty is also involved with using the class of all distributions as the admissible class. There is no objection with the definition for the one sample problem since it conveniently allows the use of a discrete distribution to estimate a continuous distribution. However, for the two-sample problem one wants to restrict the class of admissible distributions to a certain family of distributions, such as $\mathcal{L}_e$ for example. By using such a restriction, one is able to construct a likelihood function for a particular parameterization. Thus we use $\mathcal{L}_{IFR}$ or $\mathcal{L}_{DFR}$ as the admissible class when we assume $(F, G) \in \mathcal{L}_{IFR}$ or $\mathcal{L}_{DFR}$. When we assume (F, G) to belong to $\mathcal{L}_C$ we use $\mathcal{L}_e$ to be the admissible class to allow $\bar{F}$ to be estimated by a discrete distribution.

For the two sample problem, we use an analogue of Kiefer and Wolfowitz's definition. For some class of distributions $\mathcal{F}$, let $F_1, F_2, G_1, G_2 \in \mathcal{F}$ and let $f(\cdot; F_1, F_2)$ and $g(\cdot; G_1, G_2)$ denote the Radon-Nikodym derivatives of F_1 and G_1 w.r.t. the measure induced by $F_1 + F_2$ and $G_1 + G_2$, respectively. Let $\mathcal{P}$ be some class of pairs of distributions (F, G) such that $F \in \mathcal{F}$, $G \in \mathcal{F}$, and G is a function of F.

<u>Definition 1.</u> $(\hat{F}, \hat{G})$ are called the maximum likelihood estimators relative to $\mathcal{P}$ if $(F, G) \in \mathcal{P}$ and satisfy

$$\sup_{(F,G) \in \mathcal{P}} \Pi_{i=1}^{m} \frac{f(X_i; F, \hat{F})}{1 - f(X_i; F, \hat{F})} \; \Pi_{j=1}^{n} \frac{g(Y_j; G, \hat{G})}{1 - g(Y_j; G, \hat{G})} = 1,$$

where $X_1, \ldots, X_m$ is a random sample from F and $Y_1, \ldots, Y_n$ is a random sample from G.

<u>Example:</u> Let $\mathcal{F}_{NS}$ be the class of distributions defined previously, and consider the class of pairs of distributions $\mathcal{L}_e$. For $(\hat{F}, \hat{G})$ to be MLE w.r.t. to $\mathcal{L}_e$ the following relationship must be satisfied,

$$\sup_{(F,G) \in \mathcal{L}_e} \Pi_{i=1}^{m} \frac{f(X_i;F,\hat{F})}{1-f(X_i;F,\hat{F})} \Pi_{j=1}^{n} \frac{g(Y_j;G,\hat{G})}{1-g(Y_j;G,\hat{G})} = 1.$$

For the one-sample case we have

$$\begin{aligned} f(X_i;F,\hat{F}) &= 1 \quad \text{if} \quad t \neq X_i \\ &= p_i/(p_i+\hat{p}_i) \end{aligned}$$

$$\begin{aligned} g(Y_j;G,\hat{G}) &= 1 \quad \text{if} \quad t \neq Y_j \\ &= q_j/(q_j+\hat{q}_j) \end{aligned}$$

so that $(\hat{\underline{p}}, \hat{\underline{q}})$ must satisfy

$$\sup_{(F,G) \in \mathcal{L}_e} \Pi_{i=1}^{m} \frac{p_i}{\hat{p}_i} \Pi_{j=1}^{n} \frac{q_j}{\hat{q}_j} = 1,$$

or equivalently,

$$L(\hat{\underline{p}}, \hat{\underline{q}}) = \Pi_{i=1}^{m} \hat{p}_i \, \Pi_{j=1}^{n} \hat{q}_j = \sup_{(F,G) \in \mathcal{L}_e} \Pi_{i=1}^{m} p_i \Pi_{j=1}^{n} q_j$$

or in terms of the reliability functions, which are assumed to be left continuous,

$$L(\hat{\underline{p}}, \hat{\underline{q}}) = \sup_{(F,G) \in \mathcal{L}_e} \Pi_{i=1}^{m} \{\bar{F}(X_i) - \bar{F}(X_i+0)\} \Pi_{j=1}^{n} \{\bar{G}(Y_j) - \bar{G}(Y_j+0)\}$$

$$= \sup_{F \in \mathcal{F}_{NS}, \theta \in (0,\infty)} \Pi_{i=1}^{m+n} \{\bar{F}(t_i) - \bar{F}(t_i+0)\}^{\delta_i}$$

$$\times \{\bar{F}(t_i)^{\theta} - \bar{F}(t_i+0)^{\theta}\}^{1-\delta_i},$$

where $t_1, \ldots, t_{m+n}$ are the ordered observations with corresponding identification vector $(\delta_1, \ldots, \delta_{m+n})$. This is the maximization problem which is solved in Section 2.3.1, so that the resulting estimators are MLE w.r.t. $\mathcal{L}_e$ in the sense of Definition 1.

In a similar manner, the MLE of F and G w.r.t. $\mathcal{L}_{IFR}$ and w.r.t. $\mathcal{L}_{DFR}$ can be found by means of constrained maximization methods.

We believe that these modifications to Kaplan and Meier's definition give sufficient meaning to the term "nonparametric maximum likelihood estimation." Henceforth we delete the word "nonparametric." We will denote a nonparametric maximum likelihood estimator as MLE with respect to the particular class to which we assume (F,G) belongs.

In Section 2.2 we find the MLE of θ with respect to the distribution of a maximal invariant--the ranks of the ordered X observations in the combined sample. We denote this invariant MLE of θ by $\hat{\theta}_{MI}$. Following the same procedure as Kaplan and Meier (1956) we express, in Section 2.3, the likelihood function for the combined sample in terms of F and θ to obtain $\hat{F}_C$ and $\hat{\theta}_C$. A different approach is used in Section 2.4 where the MLE $\hat{\theta}_U$ is a function of F and G by way of the U-statistic, $\hat{P}(X > Y)$. When the failure rate is monotone, an appropriate likelihood function can be written in terms of the failure rate $r_F(t)$ and θ where $r_G(t) = \theta r_F(t)$. Both IFR and DFR cases are considered in Section 2.5.

2.2. MLE of θ With Respect to the Distribution of Ranks

For the first procedure, we use the principles of invariance to find a maximal invariant and its distribution. Following Ferguson (1967), if ϕ is the group of transformations $g_\phi(x_1, \ldots, x_m, y_1, \ldots, y_n) = (\phi(x_1), \ldots, \phi(x_m), \phi(y_1), \ldots, \phi(y_n))$ where ϕ is a continuous increasing function from the real line onto the real line, then a maximal invariant is the set of ranks of the X order statistics in the combined sample, $R = (r_1, \ldots, r_m)$. Once the distribution $P(R|\theta)$ is known, we can maximize the distribution function with respect to θ.

Lehmann (1953) derives $P(R|G = F^\theta, \theta > 0)$ and Shorack (1967) notes that it is possible to show by an analogous method to Lehmann (1953), or by a combinatorial argument, that

$P(R|G = F^{\theta}) = P(R'|G = 1 - (1-F)^{\theta})$, where R' is the decreasing ranks of the X sample. An alternate derivation of $P(R|G = 1 - (1-F)^{\theta})$ is given for completeness of this section.

Let $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ be samples from densities f and g respectively, and let R denote the vector of ranks of the X's in the total sample, and let the rank order statistic vector be denoted by $\underline{\delta} = (\delta_1, \ldots, \delta_{m+n})$ where $\delta_i = 1(0)$ if the ith order statistic in the combined sample is an $X(Y)$. Then by a special case of Hoeffding's theorem as given in Ferguson (1967), provided $g(x) = 0$ implies $f(x) = 0$,

$$P(R|\theta) = \binom{N}{m}^{-1} E\, \Pi_{i=1}^{N} \left(\frac{f(V_i)}{g(V_i)}\right)^{\delta_i} ,$$

where V_i are the order statistics from a sample of size $m+n$ with density g. Since the distribution of R depends on f and g only through the function $\psi(F)$, where $G(F(x); \theta) = \psi(F(x)) = 1 - (1-F(x))^{\theta}$ and $g(x; \theta) = \psi'(F(x))f(x)$, we may choose any f and g so related. A natural choice is to take $f(v) = e^{-v}$ and $g(v) = \theta e^{-\theta v}$ so that,

$$f(v)/g(v) = \theta^{-1} e^{(\theta-1)v},$$

and,

$$P(R|\theta) = \binom{N}{m}^{-1} E\{\exp \Sigma_{i=1}^{N} \delta_i(\theta-1)V_i - \ln \theta\}$$

$$= \binom{N}{m}^{-1} \theta^{-m} E\{\exp \Sigma_{i=1}^{N} \delta_i(\theta-1)V_i\}.$$

It is well known, see for example Ferguson (1967, p. 255), that by defining new variables through $V_i = \Sigma_{j=1}^{i} Y_j$, $i = 1, \ldots, N$, that the Y's are independently distributed with

$$Y_j \sim (N+1-j)\theta \exp\{-(N+1-j)\theta y_j\}, \quad j = 1, \ldots, N.$$

Thus

$$E[\exp\{\Sigma_{i=1}^{N}(\theta-1)\delta_i V_i\}] = \Pi_{i=1}^{N} E[\exp\{(\theta-1)(\Sigma_{j=1}^{N}\delta_i)Y_i\}]$$

$$= \Pi_{i=1}^{N}\{(N+1-i)\theta\}/\{(N-i+1)\theta+(1-\theta)\Sigma_{j=i}^{N}\delta_j\}$$

and

$$P(R|\theta) = m!n!\theta^n/\Pi_{i=1}^{N}\{\Sigma_{j=i}^{N}(\delta_j+\theta(1-\delta_j))\}. \tag{2.1}$$

Theorem 2.1. For the distribution $P(R|\theta)$, there exists a unique maximum with respect to $\hat{\theta} \in (0,\infty)$ provided $\Sigma_{i=1}^{m}\delta_i \neq m$ and $\Sigma_{i=1}^{n}(1-\delta_i) \neq n$.

Proof: Since neither $\hat{\theta} = 0$ nor $\hat{\theta} = \infty$ can maximize $P(R|\theta)$ it is necessary that $\frac{d}{d\theta}P(R|\theta)|_{\hat{\theta}} = 0$ for $\hat{\theta}$ to be a MLE of θ. By differentiation we obtain

$$n/\hat{\theta} - \Sigma_{j=1}^{N} n_j/(m_j+\hat{\theta}n_j) = 0, \qquad (2.2a)$$

where $n_j = \Sigma_{i=j}^{N}(1-\delta_i)$ and $m_j = \Sigma_{i=j}^{N}\delta_i$. Every solution to (2.2a) is also a solution to

$$S(\hat{\theta}) = \Sigma_{j=1}^{N} \hat{\theta}n_j/(m_j+\hat{\theta}n_j) = n. \qquad (2.2b)$$

Consider first the trivial solutions to (2.2b). If $\Sigma_{i=1}^{m}\delta_i = m$ then (2.2b) becomes $\hat{\theta}n/(\hat{\theta}n+m) + \ldots + \hat{\theta}n/(\hat{\theta}n+1) + n\hat{\theta}/n\hat{\theta} + \ldots + \hat{\theta}/\hat{\theta} = n$. A solution is approached as $\hat{\theta} \to 0$. If $\Sigma_{i=1}^{n}(1-\delta_i) = n$ then (2.2b) becomes $\hat{\theta}n/(\hat{\theta}n+m) + \ldots + \hat{\theta}/(\hat{\theta}+m) + 0 + \ldots + 0 = n$. In this trivial case a solution is approaches as $\hat{\theta} \to \infty$.

To prove that (2.2b) has a unique positive solution we make the following observations. The first terms in the sum are of the form $\hat{\theta}n_j/(\hat{\theta}n_j+m_j)$ with $n_j \geq 1$, $m_j \geq 1$, which are nonnegative, strictly increasing functions of $\hat{\theta} \in (0,\infty)$ with range $(0, 1)$, and latter terms which are either all 0 or all 1. If the latter terms are 0 then by the assumption $\Sigma_{i=1}^{n}(1-\delta_i) \neq n$, $S(\hat{\theta})$ has $(0, n+1)$ in its range. If the latter terms are 1, then by assumption $\Sigma_{i=1}^{m}\delta_i \neq m$ there are fewer than $n-1$ terms identically 0, so that $S(\hat{\theta})$ has $(n-1, N)$ in its range. Thus $S(\hat{\theta})$ is a strictly increasing function of $\hat{\theta} \in (0,\infty)$ with n in its range, and the result follows.

The following method was used to solve Equation (2.2b).

If either of the trivial cases occur then $\hat{\theta}$ could be taken to be 0 or ∞ respectively. In other cases the Newton-Raphson method can be used to find the root of

$$f(\hat{\theta}) = \Sigma_{j=1}^{N} n_j\hat{\theta}/(m_j+n_j\hat{\theta}) - n = 0,$$

where

$$f'(\hat{\theta}) = \Sigma_{j=1}^{N} m_j n_j /(m_j+n_j\hat{\theta})^2 > 0.$$

Then given initial ${}^1\theta$, ${}^{i+1}\theta = {}^i\theta - f({}^i\theta)/f'({}^i\theta)$ can be sequentially found until $f({}^i\theta) < \epsilon_1$, or until ${}^i\theta - {}^{i+1}\theta/({}^{i+1}\theta) < \epsilon_2$. Since $g({}^i\theta) = {}^i\theta - f({}^i\theta)/f'({}^i\theta)$ is analytic for $\theta \epsilon (0,\infty)$ and $g'(\theta^*) = 0$, where θ^* is a solution to $f(\theta) = 0$, we have sufficient conditions, by a theorem in Macon (1963, p. 30), for the sequence ${}^1\theta, {}^2\theta, \ldots$ to converge in an ϵ-neighborhood of θ^*. Moreover, $f''(\theta) = -2\Sigma_{j=1}^{N} m_j n_j^2 (m_j+n_j\theta)^{-3} < 0$ so that $f(\theta)$ is a strictly increasing concave function of $\theta > 0$ and the sequence ${}^1\theta, {}^2\theta, \ldots$ will converge for any ${}^1\theta$, where $0 < {}^1\theta < \infty$.

2.3. Product-Limit Parameterization of the Likelihood Function

2.3.1. Derivation of the MLE of θ

In their classic paper, Kaplan and Meier (1956) discuss non-parametric estimation of a reliability function from incomplete data.

One section of their work relevant to this paper is the maximum likelihood derivation of the product-limit estimate, since we use the same parameterization of the likelihood function. Efron (1965) constructs a "self-consistent" estimate of $\bar{F}(x) = 1 - F(x)$ which coincides with Kaplan and Meier's product-limit estimate. For the two sample problem, he uses these MLE of $\bar{F}$ and $\bar{G}$ to show that $\hat{W} = -\int \hat{\bar{F}}(s)d\hat{\bar{G}}(s)$ is the maximum likelihood estimate of $P(X > Y)$. Provided $\bar{F}^{\theta} = \bar{G}$, we can use $\hat{P}(X > Y)$ to obtain a MLE of θ with respect to $\hat{W}$. In this section however, we assume $(F, G) \in \mathcal{L}_C$ to write a likelihood function in terms of $\bar{F}$ and θ. In comparison to the former method, this approach has the disadvantage of requiring an iterative solution but has a distinct advantage by obtaining a combined sample MLE of $\bar{F}$, denoted by $\hat{\bar{F}}_C$. Both methods are suited for arbitrarily censored data.

Let $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ be random samples from their respective distributions F and G belonging to $\mathcal{L}_C$. Assume that all ordered observations are distinct and complete, each being denoted by $t_1, \ldots, t_n$ with associated indicator variables δ_i having values $1(0)$ if t_i is an $F(G)$ observation.

By expressing $\bar{G}(t)$ as $\bar{F}(t)^{\theta}$ and letting $\bar{F}(t_i+0)$ represent the limit of $\bar{F}(t_i)$ from the right, we define the (nonparametric) likelihood function

$$L(\underline{t}, \underline{\delta} | F, \theta) = \Pi_{i=1}^{N} \{\bar{F}(t_i) - \bar{F}(t_i+0)\}^{\delta_i} \{\bar{F}(t_i)^{\theta} - \bar{F}(t_i+0)^{\theta}\}^{1-\delta_i}.$$

This likelihood function will be maximized by making $\bar{F}(t_i)$ as large as possible and making $\bar{F}(t_i+0)$ as small as possible while remaining consistent with nonincreasing property of reliability functions. This requires that $\bar{F}(t_i+0) = \bar{F}(t_{i+1})$, $i = 1, \ldots N-1$, $\bar{F}(t_1) = 1$ and $\bar{F}(t_N+0) = 0$. We change parameters to express $\bar{F}(t_i)$ in terms of $p_i = P(X \geq t_{i+1} | X \geq t_i)$ $i = 1, \ldots, N-1$, so that $F(t_i) = \Pi_{k=1}^{i-1} p_k$. Then $\bar{F}(t_i) - \bar{F}(t_{i+1}) = \{\Pi_{k=1}^{i-1} p_k\}(1-p_i)$ and $\bar{F}(t_i)^{\theta} - \bar{F}(t_{i+1})^{\theta} = \{\Pi_{k=1}^{i-1} p_k^{\theta}\}(1-p_i^{\theta})$. Then,

$$L(\underline{t}, \underline{\delta} | \underline{p}, \theta) = \Pi_{i=1}^{N} [\{(\Pi_{k=1}^{i-1} p_k)(1-p_i)\}^{\delta_i} \{(\Pi_{k=1}^{i-1} p_k)^{\theta} (1-p_i^{\theta})\}^{1-\delta_i}]$$

$$= \Pi_{i=1}^{N} p_i^{\Sigma_{k=i+1}^{N} \{\delta_k + \theta(1-\delta_k)\}} (1-p_i)^{\delta_i} (1-p_i^{\theta})^{1-\delta_i}$$

$$= \Pi_{i=1}^{N} p_i^{N_i(\theta) - \delta_i - \theta(1-\delta_i)} (1-p_i)^{\delta_i} (1-p_i^{\theta})^{1-\delta_i},$$

where

$$N_i(\theta) = m_i + \theta n_i = \Sigma_{k=i}^{N} \{\delta_k + \theta(1-\delta_k)\}.$$

Before maximizing the likelihood function, we observe that $0 < p_i < 1$ for the likelihood to be positive. Let $L(\underline{p}, \theta) = \ln L(\underline{t}, \underline{\delta} | \underline{p}, \theta)$. By setting the gradient $\nabla \ln L(\underline{p}, \theta) = \underline{0}$ we obtain

$$\frac{\partial \ln L}{\partial p_i}\Bigg|_{\hat{p}_i,\hat{\theta}} = \{N_i(\hat{\theta})-\delta_i-\hat{\theta}(1-\delta_i)\}/\hat{p}_i-\delta_i/(1-\hat{p}_i) - (1-\delta_i)\,\frac{\hat{\theta}\hat{p}_i^{\hat{\theta}-1}}{1-\hat{p}_i^{\hat{\theta}}} = 0$$

which implies,

$$\{\hat{p}_i-N_i(\hat{\theta})+\delta_i/N_i(\hat{\theta})\}\;\{\hat{p}_i^{\hat{\theta}}-N_i(\hat{\theta})+(1-\delta_i)\hat{\theta}/N_i(\hat{\theta})\} = 0$$

so that,

$$\hat{p}_i = \{N_i(\hat{\theta})-\delta_i\}/N_i(\hat{\theta}) \quad \text{if} \quad \delta_i = 1 \tag{2.3}$$

$$\hat{p}_i^{\hat{\theta}} = \{N_i(\hat{\theta})-(1-\delta_i)\hat{\theta}\}/N_i(\hat{\theta}) \quad \text{if} \quad \delta_i = 0 \tag{2.4}$$

and

$$\frac{\partial \ln L}{\partial \theta}\Bigg|_{\underline{p},\hat{\theta}} = \Sigma_{i=1}^{N}[\{n_i-(1-\delta_i)\}\ln\hat{p}_i - (1-\delta_i)\hat{p}_i^{\hat{\theta}}\ln\hat{p}_i/(1-\hat{p}_i^{\hat{\theta}})] = 0$$

to imply

$$\Sigma_{i=1}^{N}\{n_i-(1-\delta_i)/(1-\hat{p}_i^{\hat{\theta}})\}\ln\hat{p}_i = 0.$$

By substituting $\hat{p}_i$ appropriate to δ_i into the last equation, we obtain

$$\Sigma_{\delta=1}\{n_i\ln(1-\delta_i/N_i(\hat{\theta}))\} = \Sigma_{\delta=0}[m_i\ln\{1-(1-\delta_i)\hat{\theta}\}/N_i(\hat{\theta})]/\hat{\theta}^2, \tag{2.5}$$

where the sets $\{i: \delta_i=1\}$ and $\{i: \delta_i=0\}$ are denoted by $\delta = 1$ and $\delta = 0$ respectively.

We need the following lemma to prove the uniqueness of the solution to (2.5).

<u>Lemma. 2.1.</u> If $x \geq 1$, $y \geq 1$ and $\theta > 0$ then

$h(\theta) = -y\theta \ln(\frac{x-1+y\theta}{x+y\theta})$ is a strictly increasing function of θ.

Proof: Let $\lambda(\gamma) = -\gamma \ln(1-1/(x+\gamma))$, $\gamma > 0$ so that $\lambda'(\gamma) = -\ln(1-1/(x+\gamma)) - \gamma/(x+\gamma)(x+\gamma-1) > [1/(x+\gamma)][1-\gamma/(x+\gamma-1)] \geq 0$, since $-\ln(1-y) > y$ for $y < 1$.

Corollary. If $x \geq 1$, $y \geq 1$ and $\theta \geq 0$ then $k(\theta) = -x/\theta \ln[(x/\theta+y-1)/(x/\theta+y)]$ is a strictly decreasing function in θ.

Theorem 2.2. The solution to the maximum likelihood equation (2.5) is unique provided $\Sigma_{i=1}^{m}\delta_i \neq m$ and $\Sigma_{i=1}^{n}(1-\delta_i) \neq n$, where $N = m+n$.

Proof: $L(\underline{p}, \theta) > 0$ only if $0 < p_i < 1$, $i = 1, \ldots, N$ and $0 < \theta < \infty$, so that a solution cannot exist on the boundary of the parameter set. Then for $(\hat{\underline{p}}, \hat{\theta})$ to maximize $L(\underline{p}, \theta)$ it is necessary for $\nabla L(\hat{\underline{p}}, \hat{\theta}) = \underline{0}$. The solution vector $\underline{p}$ is a unique function of θ, so that if there is a unique $\hat{\theta}$ satisfying (2.5), then $(\hat{\underline{p}}, \hat{\theta})$ is unique. Let $f(\theta) = \Sigma_{\delta=1} n_i \ln(1-\delta_i/N_i(\theta)) \leq 0$, $0 \leq \theta < \infty$, w.e.h. iff $\Sigma_{i=1}^{n}(1-\delta_i) \neq n$, let $g(\theta) = \Sigma_{\delta=0}^{m} m_i \ln[1-(1-\delta_i)\theta/N_i(\theta)] \leq 0$, $0 < \theta \leq \infty$, w.e.h. iff $\Sigma_{i=1}^{m}\delta_i \neq m$ and let $\phi(\theta) = g(\theta)/(\theta^2 f(\theta))$.

We adopt the convention that if either $m_i = 0$ or $n_i = 0$ we define $m_i \ln[1-\theta/N_i(\theta)] = 0$ or $n_i \ln(1-1/N_i(\theta)) = 0$, respectively, for all $\theta \in [0, \infty]$. Thus we need only consider non-zero terms in $f(\theta)$ and $g(\theta)$ and can therefore assume that $m_i \geq 1$ and $n_i \geq 1$ for each term being summed. Provided the hypothesis is satisfied, then for $\epsilon > 0$, $\theta > 0$ we have

$$\{g(\theta+\epsilon)/[(\theta+\epsilon)^2 f(\theta+\epsilon)]\} \{[\theta^2 f(\theta)\}/g(\theta)\}$$

$$= \frac{\Sigma_{\delta=0} \frac{m_i}{\theta+\epsilon}(1 - \frac{\theta+\epsilon}{m_i+n_i(\theta+\epsilon)})}{\Sigma_{\delta=0} \frac{m_i}{\theta} \ln(1 - \frac{\theta}{m_i+n_i\theta})} \quad \frac{\Sigma_{\delta=1} n_i \theta \ln(1 - \frac{1}{m_i+\theta n_i})}{\Sigma_{\delta=1} n_i(\theta+\epsilon) \ln(1 - \frac{1}{m_i+(\theta+\epsilon)n_i})} \leq 1$$

by application of the corollary and Lemma 2.1 to the first and second terms, respectively. Hence $\phi(\theta)$ is a strictly decreasing function for positive θ.

Before evaluating the limiting values of $\phi(\theta)$, we consider the limiting values of the component functions;

$$\lim_{\theta \to 0} g(\theta) = \lim_{\theta \to 0} \Sigma_{\delta=0} m_i \ln[1 - \theta/(m_i+\theta n_i)] = 0,$$

and

$$\lim_{\theta \to 0} f(\theta) = \lim_{\theta \to 0} \Sigma_{\delta=1} n_i \ln[1 - 1/(m_i+\theta n_i)]$$

$$= \Sigma_{\substack{\delta=1 \\ m_i \neq 1}} n_i \ln[(m_i - 1)/m_i]$$

$$+ \lim_{\theta \to 0} \Sigma_{\substack{\delta=1 \\ m_i = 1}} n_i \ln[1 - 1/(1+\theta n_i)]$$

which equals a negative constant plus a term tending to infinity if $x_m < y_n$. If $x_m > y_n$ then $\lim_{\theta \to 0} \theta^2 f(\theta) = 0$. If $x_m < y_n$ then

$$\lim_{\theta \to 0} \theta^2 f(\theta) = 0 + \lim_{\substack{\theta \to 0 \\ m_i = 1}} \Sigma_{\delta=1} \theta^2 n_i \ln[1 - 1/(1+\theta m_i)]$$

$$= \lim_{\theta \to 0} [n_i^2/(1+\theta n_i)(\theta n_i)]/[-2\theta^{-3}] = 0$$

by use of l'Hopital's rule. Thus the form of $\lim_{\theta \to 0} \phi(\theta)$ is $\frac{0}{0}$.

Now,

$$\lim_{\theta \to 0} \phi(\theta) = \lim_{\theta \to 0} g'(\theta)/[2\theta f(\theta) + \theta^2 f'(\theta)]$$

$$= \lim_{\theta \to 0} \frac{\Sigma_{\delta=0} m_i^2/\{N_i(\theta)[N_i(\theta)-\theta]\}}{2\theta \Sigma_{\delta=1} n_i \ln(1-1/N_i(\theta)) + \theta^2 \Sigma_{\delta=1} n_i^2/\{N_i(\theta)[N_i(\theta)-1]\}}$$

$$= \infty.$$

To find $\lim_{\theta \to \infty} \phi(\theta)$, we write

$$\lim_{\theta \to \infty} \phi(\theta) = \lim_{\theta \to \infty} \frac{g(\theta)/\theta^2}{f(\theta)} \qquad \text{(which is of form } \tfrac{0}{0}\text{)}$$

$$= \lim_{\theta \to \infty} \frac{-2g(\theta)/\theta^3 + g'(\theta)/\theta^2}{f'(\theta)}$$

$$= \lim_{\theta \to \infty} \frac{-2g(\theta)/\theta + g'(\theta)}{\theta^2 f'(\theta)}$$

$$= \lim_{\theta \to \infty} -2/\theta \Sigma_{\delta=0} m_i \ln[1-\theta/N_i(\theta)] - \Sigma_{\delta=0} m_i^2/\{N_i(\theta)[N_i(\theta)-\theta]\} = 0.$$

Hence $\phi(\theta)$ is a strictly decreasing function of θ having range $(0, \infty)$, and therefore, the equation $g(\theta)/\theta^2 = 1$ has a unique positive solution. Since $f(\theta) \neq 0$ on $[0, \infty)$, the equation $f(\theta) = g(\theta)/\theta^2$ has a unique positive solution.

To numerically find a solution to (2.5), one can use the Newton-Raphson method to find a solution to the related equation $h(\theta) = f(\theta) - g(\theta)/\theta^2 = 0$. Care should be exercised to begin the iteration in the interval $(0, \theta')$ where θ' is the solution to $h'(\theta) = 0$. The function $h(\theta)$ has limiting value ∞ at 0^+ and decreases to a negative value at θ', after which, $h(\theta)$ increases asymptotically to 0^- as $\theta \to \infty$. If $h'({}^n\theta) < 0$ for some n, then ${}^n\theta \in (\theta, \theta')$.

2.3.2. Estimation for Arbitrarily Censored Data

The derivation of the full sample likelihood function is easily extended to the case of arbitrarily censored data via the method of Kaplan and Meier (1958). In brief, the extension is as follows: Suppose for the ith item placed on test there is a fixed observation period L_i, such that, if the item fails before L_i then the failure time is recorded, but if the item is effective after time L_i then the censoring time is recorded. An example of such a scheme is given by the use of random arrivals in a medical experiment which has a fixed termination date.

Let $t_1, \ldots, t_k$, $k < N$, denote distinct failure times and let $\lambda_i(\epsilon_i)$ denote the number of $F(G)$ censorings in $[t_i, t_{i+1})$, $i = 0, \ldots, k$, and $t_{k+1} = \infty$. Then by the same argument as in the complete sample case, but conditioning upon the censored observations,

$$L(\underline{m}, \underline{n} | \underline{p}, \theta) = \Pi_{i=1}^{k} p_i^{m_i+\theta n_i-\delta_i-(1-\delta_i)\theta} (1-p_i)^{\delta_i} (1-p_i^{\theta})^{1-\delta_i}$$

where $m_i = \Sigma_{j=i}^{k} \delta_j + \lambda_j$ and $n_j = \Sigma_{j=i}^{k}(1-\delta_j) + \epsilon_j$. The maximum likelihood estimates have the same form as in the complete sample case.

2.3.3. Combined-Sample MLE of the F Reliability Function and Failure Rate Function

By expressing the likelihood function in terms of $\bar{F}$, θ and subsequently $\underline{p}$, θ we have readily available a combined-sample MLE of $\bar{F}$. The conditional probability $p_i = P(X \geq t_{i+1} | X \geq t_i)$ is estimated by $\hat{p}_i = \{1-[\delta_i+(1-\delta_i)\hat{\theta}]/N_i(\hat{\theta})\}^{\delta_i+(1+\delta_i)/\hat{\theta}}$. Hence the corresponding MLE of $\bar{F}(s)$ is given by $\hat{\bar{F}}_C(s) = \Pi_{i=1}^{k-1} \hat{p}_i$ for $s \epsilon (t_{k-1}, t_k]$. Also, $\bar{G}(s)$ may be estimated by $\hat{\bar{G}}_C(s) = \hat{\bar{F}}_C(s)^{\hat{\theta}}$.

The failure rate function for a discrete distribution function with probability mass at x_i, $i = \ldots, -1, 0, 1, 2, \ldots$ is defined to be $r_i = p_i / \Sigma_{j=i}^{\infty} p_j$. Although the t_i are random in our setup, we do have a MLE of $\bar{F}(t_i)$ and consequently of $r(t_i)$ at each t_i. If required to estimate $r(s)$ on $[t_1, t_N)$ then one could, in the absence of smoothing considerations, estimate $r(t_i) = p_i/\bar{F}(t_i)$ for $i = 1, \ldots, N-1$ and then adjust the discrete estimator by spreading the mass over each interval to obtain a continuous estimator:

(1) $\tilde{r}(u) = \tilde{r}(t_i)$ for $t_i \leq u < t_{i+1}$

(2) linear approximation on (t_i, t_{i+1}), $\tilde{r}(u) = \tilde{r}(t_i) + m_i(u-t_i)$ where $m_i = [\tilde{r}(t_{i+1}) - \tilde{r}(t_i)]/(t_{i+1}-t_i)$.

2.4. MLE of θ as a Function of a U-Statistic

Several papers are devoted to the two-sample problem of estimating $P(X > Y)$, two of which are Govindarajulu (1968) and Church and Harris (1970). In general, these estimates are used for either constructing confidence intervals on $P(X > Y)$ or for testing the two-sample problem hypothesis, $H_0: F = G$. Efron (1965, p. 838) notes:

> A desirable property of any test statistic for the two sample problem is that when the null hypothesis is not true, that is when $F^0 \neq G^0$, the statistic estimates some reasonable measure of the difference between the distributions. Usually we are not interested in a simple acceptance or rejection of the null hypothesis but would like to made a quantitative assessment of the treatment differences.

The MLE $\hat{P}(X > Y)$ is such a measure for the two-sample problem. For the case $(F, G) \in \mathcal{L}_C$ there is also such a measure of interest to the statistician. By calculating $P(X > Y)$ given that

$(F, G) \in \mathcal{L}_C$, we have

$$P(X > Y \mid \bar{G} = \bar{F}^{\theta}) = -\int_{-\infty}^{\infty} \bar{F}(s) d\bar{G}(s)$$

$$= -\frac{\theta}{\theta+1} \bar{G}(s)^{(\theta+1)/\theta} \Big|_{-\infty}^{\infty}$$

$$= \theta/(\theta+1).$$

Ghosh (1970) refers to the parameter θ as a measure of the intensity with which X tends to be larger than Y.

If a MLE of $P(X > Y)$ is found within the class of all pairs of distributions, then by the invariance property of MLE under 1-1 transformations, $\hat{\theta} = \hat{P}(X > Y)/[1 - \hat{P}(X > Y)]$ is the MLE of θ based on $(\hat{F}, \hat{G})$ within the subclass $\mathcal{L}_C$. Efron (1965) provides a general MLE of $P(X > Y)$ when he investigates the two sample problem with censored data. He shows that his "self-consistent" estimator of $\bar{F}$ and $\bar{G}$ coincide with Kaplan and Meier's product limit estimator, and hence are MLE. The statistic defined by $\hat{W} = -\int_{-\infty}^{\infty} \hat{\bar{F}}(s) d\hat{\bar{G}}(s)$ is shown to be the MLE of $P(X > Y)$ as well as being asymptotically normally distributed.

We refer to $\hat{\theta}_U = \hat{W}/(1 - \hat{W})$ as the maximum likelihood estimate of θ based on the U-statistic $\hat{W}$. Thus $\hat{\theta}_U$ is the ratio $\hat{P}(X > Y)/\hat{P}(X < Y)$. Another way to view this relationship is the following. If the random walk associated with the empirical distributions

F_m and G_n is plotted in the unit square, the walk beginning at (0, 0) and moving one step of size $1/m$ to the right or one step of size $1/n$ up at each order observation (t_i) if $\delta_i = 1$ or 0 respectively, then $\hat{\theta}_U$ is the ratio of the area below the random walk to the area above the random walk.

2.5. Estimation of θ and $r_F(x)$ Assuming Monotone Failure Rate

2.5.1. Introduction

From either the structure of an experiment or from prior knowledge of items under test, one may be able to assume some property of the failure rate function. When the failure rate is monotone, an appropriate likelihood function can be written. Monotone failure rate has been studied extensively for the single sample problem: Grenander (1956) derived the MLE for distributions with IFR, Marshall and Proschan (1965) extended the MLE to distributions with monotone failure rate for both continuous and discrete distributions, Bray *et al*. (1967) derived the MLE for distributions with U-shaped failure rate functions. Other papers have extended MLE for various censoring schemes.

Let $\mathcal{F} = \{F \mid F(0) = 0$ and $-\log[1-F(x)]$ is convex for $x \geq 0\}$ which defines the class of IFR (increasing failure rate) distributions. It is proved in Barlow and Proschan (1965) that IFR distributions are absolutely continuous on their interval of support but may have a jump at the right-hand end. Barlow (1968) notes that by using the generalized definitions of MLE one need only consider estimators absolutely continuous with respect to Lebesgue measure on $[0, X_n)$ with a jump at X_n. However, in applying the Kuhn-Tucker conditions, we find the bounded approach used by Grenander to be a convenient formulation of the problem. The resultant estimators $(\hat{F}, \hat{G})$ satisfy the generalized definition.

2.5.2. Two-Sample IFR Likelihood Function

Let $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ be random samples from distributions F and G where $(F, G) \in \mathcal{L}_{IFR}$. Let $\underline{t}$ denote the vector of the combined ordered sample with corresponding identification vector $\underline{\delta}$, where $\delta_i = 1(0)$ if t_i is an $F(G)$ observation. By using the relations $r(t) = f(t)/(1-F(t))$ and $1 - F(t) = \exp\{-\int_0^t r(u)du\}$ with corresponding expressions for G, we can write

$$L_N(\underline{t}, \underline{\delta} | \underline{r}, \theta) = \Pi_{i=1}^N r_F(t_i)^{\delta_i} r_G(t_i)^{1-\delta_i} \exp[-\delta_i \int_0^{t_i} r_F(u)du - (1-\delta_i) \int_0^{t_i} r_G(u)du]$$

$$= \Pi_{i=1}^N r(t_i) \theta^{1-\delta_i} \exp\{-\Sigma_{i=1}^N \{\int_{t_{i-1}}^{t_i} r(u)du \, \Sigma_{j=i}^N [\delta_j + \theta(1-\delta_j)]\}\}$$

since $r_G(t) = \theta r_F(t)$, $t \geq t_0 = 0$, and where $r(t) = r_F(t)$. If $L_N(\underline{t}, \underline{\delta} | \underline{r}, \theta)$ is to be maximized subject to the constraints $r(t_0) \leq r(t_1) \leq \cdots \leq r(t_N) \leq M < \infty$ then, as Grenander argues, the maximum likelihood estimator of $r(t)$ must be of the form

$$\tilde{r}(u) = r(t_i), \quad t_i \leq u < t_{i+1}, \quad i = 0, 1, \ldots, N-1$$
$$= M \qquad t_N \leq u.$$

Thus the log likelihood function is

$$\mathcal{L}_N(\underline{t}, \underline{\delta} | \underline{r}, \theta) = \Sigma_{i=1}^N \{\ln r_i + (1-\delta_i) \ln \theta - (a_i + b_i \theta) r_{i-1}\},$$

where $r_i = r(t_i)$, $a_i = \Delta t_{i-1} \Sigma_{j=i}^N \delta_j$ is the exposure time of the F observations in the interval $[t_{i-1}, t_i)$, and $b_i = \Delta t_{i-1} \Sigma_{j=i}^N (1-\delta_j)$, for $i = 1, \ldots, N$.

To carry out the maximization of $\mathcal{L}_N$ subject to $(F, G) \in \mathcal{L}_{IFR}$, we state the problem in a non-linear programming framework:

<u>NLP Problem 1.</u>

$$\max \mathcal{L}_N(\underline{r}, \theta) = \Sigma_{i=1}^N \{\ln r_i + (1-\delta_i) \ln \theta - (a_i + b_i \theta) r_{i-1}\}$$

such that $g_0(\underline{r}, \theta) = r_0 \geq 0$

$$g_i(\underline{r}, \theta) = r_i - r_{i-1} \geq 0, \quad i = 1, \ldots, N$$

$$g_{N+1}(\underline{r}, \theta) = \theta \geq 0$$

$$g_{N+2}(\underline{r}, \theta) = M - r_N = 0.$$

Following Zangwill (1969, p. 42) we state the Kuhn-Tucker (K-T) conditions which are necessary for $(\underline{r}^*, \theta^*)$ to be optimal for this problem:

(1) $(\underline{r}^*, \theta^*)$ is feasible;

there exist multipliers $\lambda_i \geq 0$, $i = 0, \ldots, N+1$, and unconstrained multiplier λ_{N+2}, such that

(2) $\lambda_i g_i(\underline{r}^*, \theta^*) = 0, \quad i = 0, \ldots, N+1$

and

(3) $\nabla \mathcal{L}_N(\underline{r}^*, \theta^*) + \Sigma_{i=0}^{N+1} \lambda_i \nabla g_i(\underline{r}^*, \theta^*) = 0.$

Condition (3) requires

$$\begin{bmatrix} -(a_1+b_1\theta) \\ 1/r_1-(a_2+b_2\theta) \\ 1/r_2-(a_3+b_3\theta) \\ \vdots \\ 1/r_{N-1}-(a_N+b_N\theta) \\ 1/r_N \\ \Sigma_{i=1}^{N}(1-\delta_i)/\theta-\Sigma_{i=0}^{N-1} b_{i+1} r_i \end{bmatrix} + \begin{bmatrix} \lambda_0 \\ \\ \\ \\ \\ \\ \\ \end{bmatrix} + \begin{bmatrix} -\lambda_1 \\ \lambda_1 \\ \\ \\ \\ \\ \\ \end{bmatrix} + \ldots + \begin{bmatrix} \\ \\ \\ \\ -\lambda_N \\ \lambda_N \\ \\ \end{bmatrix} + \begin{bmatrix} \\ \\ \\ \\ \\ \\ \lambda_{N+1} \end{bmatrix} + \begin{bmatrix} \\ \\ \\ \\ \\ -\lambda_{N+2} \\ \\ \end{bmatrix} = \underline{0}$$

or equivalently,

$$\lambda_1 = -(a_1+b_1\theta) + \lambda_0$$

$$\lambda_2 = 1/r_1 - (a_2+b_2\theta) + \lambda_1$$

$$\vdots$$

$$\lambda_N = 1/r_{N-1} - (a_N+b_N\theta) + \lambda_{N-1}$$

$$\lambda_{N+2} = 1/r_N + \lambda_N$$

$$\lambda_{N+1} = \Sigma_{i=0}^{N-1} b_{i+1} r_i - n/\theta.$$

The equations resulting from K-T condition (3) can be reduced by using condition (1) and (2). If $r_0 > 0$ then $\lambda_0 = 0$ which implies $\lambda_1 = -(a_1+b_1\theta) < 0$. Therefore r_0 must equal 0 and letting $\lambda_0 = (a_1+b_1\theta)$ gives $\lambda_1 = 0$. $\theta = 0$ implies $\mathcal{L}_N(\underline{r}, \theta) = -\infty$ which cannot be optimal, so that it is necessary that $\lambda_{N+1} = 0$ to make $\lambda_{N+1} g_{N+1}(\underline{r}, \theta) = 0$. To be feasible, r_N must equal M, and provided M is sufficiently large, $r_{N-1} < M$, to require $\lambda_N = 0$.

We now seek multipliers $\lambda_2, \ldots, \lambda_{N-1}$ such that

(1) $(r_i - r_{i-1}) \geq 0$, $i = 2, \ldots, N-1$;

(2) $\lambda_i(r_i - r_{i-1}) = 0$, $i = 2, \ldots, N-1$;

and

$$\left.\begin{aligned} \lambda_2 &= 1/r_1 - (a_2+b_2\theta) \\ \lambda_3 &= 1/r_2 - (a_3+b_3\theta) + \lambda_2 \\ &\vdots \\ \lambda_{N-1} &= 1/r_{N-2} - (a_{N-1}+b_{N-1}\theta) + \lambda_{N-2} \\ 0 &= 1/r_{N-1} - (a_N+b_N\theta) + \lambda_{N-1} \end{aligned}\right\} \quad (2.6)$$

$$\Sigma_{i=1}^{N-1} b_{i+1} r_i = n/\theta. \tag{2.7}$$

For any fixed finite $\theta > 0$ there is a sequential procedure for obtaining a vector $\underline{\lambda}$ and a feasible vector $\underline{r}$. This procedure is analogous to the sequential algorithms given by Grenander (1956), Brunk (1958), and Marshall and Proschan (1965); the significance of this description is that the algorithm is described within the framework of Kuhn-Tucker theory, a framework which we find to be convenient for solving NLP Problem 1.

Given θ, set $\lambda_i = 0$, $i = 2, \ldots, N-1$, to obtain unconstrained estimates of $r_i = 1/(a_{i+1}+b_{i+1}\theta)$, $i = 1, \ldots, N-1$. Set $j = 1$ and begin with step (a).

(a) $j = j+1$. If $j = N$ then procedure is completed.

(b) If $r_j \geq r_{j-1}$ return to (a).

(c) If $r_j < r_{j-1}$ then constraint $g_j = r_j - r_{j-1}$ is not feasible.

(i) Set $r_j = r_{j-1}$ so that a harmonic average for r_j will be found. (Note that condition (2) appropriate to constraint j, that is $\lambda_j g_j = 0$, does not require $\lambda_j = 0$ since $g_j = r_j - r_{j-1} = 0$.) Suppose that r_{j-1} has been averaged with k prior r's, where $k = 0, 1, \ldots, j-2$, so that $r_{j-1} = \ldots = r_{j-1-k}$. Let λ_{j-k} through λ_j equal the right hand sides of their equations in (2.6) and solve

$$\lambda_{j+1} = 0 = 1/r_{j-1-k} - (a_{j-k}+\theta b_{j-k}) + \ldots + 1/r_j - (a_{j+1}+\theta b_{j+1})$$

to obtain

$$r_{j-1-k} = \ldots = r_j = (k+2)/[\Sigma_{\ell=j-1-k}^{j}(a_\ell+\theta b_\ell)] = 1/(\bar{a}_j+\theta\bar{b}_j).$$

(ii) If $r_{j-1-k} \geq r_{j-2-k}$ return to (a). Otherwise average r_{j-2-k} (and preceding r's if averaged) with r_{j-1-k} through r_j and return to (ii) with k now equalling the number of r's averaged with r_{j-1}.

There are a few notes that should be made about the sequential procedure for finding $(\underline{r}, \underline{\lambda})$:

1. The resultant vector $\underline{r}$ has components $r_1 \leq r_2 \leq \cdots \leq r_{N-1} < r_N = M < \infty$, $r_j = 1/(\bar{a}_{j+1}+\bar{b}_{j+1}\theta)$ where $\bar{a}_{j+1}$ and $\bar{b}_{j+1}$ may be averages of two or more terms.

2. The resultant vector $\underline{\lambda}$ has zero component λ_{k+1} if r_{k+1} is not averaged with r_k, and positive component λ_{j+1} if r_{j+1} is averaged with r_j. The necessity of $\lambda_{j+1} > 0$ follows since if initially $r_j > r_{j+1}$ and $r_{j-1} < r_j$; then $a_{j+1} + b_{j+1}\theta < a_{j+2} + b_{j+2}\theta$ so that, $\bar{r}_j = \bar{r}_{j+1} = [a_{j+1}+a_{j+2}+(b_{j+1}+b_{j+2})\theta]/2 = \bar{a}_{j+2} + \bar{b}_{j+2}\theta$ and

$$\lambda_{j+1} = 1/\bar{r}_{j+1} - (a_{j+1}+b_{j+1}\theta) = [(a_{j+2}-a_{j+1}) + (b_{j+2}-b_{j+1})\theta] > 0$$

and $\lambda_{j+2} = 1/\bar{r}_{j+1} - (a_{j+2}+b_{j+2}\theta) + \lambda_{j+1} = 0.$

Similarly, if r_j is averaged with ℓ previous r's, $\ell = 0, 1, 2, \ldots, j-1$ and $r_j > r_{j+1}$, then the intermediate positive multipliers $\lambda_{j-\ell}, \ldots, \lambda_j$ increase when $\bar{r}_j$ is decreased with the averaging with the smaller r_{j+1}, and consequently the reciprocal of the average r is increased. For example, if we have

$$\begin{aligned}
\lambda_{j-\ell+1} &= 1/\bar{r}_j - (a_{j-\ell+1}+b_{j-\ell+1}\theta) && > 0 \\
\lambda_{j-\ell+2} &= 1/\bar{r}_j - (a_{j-\ell+2}+b_{j-\ell+2}\theta) + \lambda_{j-\ell+1} && \geq 0 \\
&\vdots && \vdots \\
\lambda_j &= 1/\bar{r}_j - (a_j+b_j\theta) + \lambda_{j-1} && > 0 \\
\lambda_{j+1} &= 1/\bar{r}_j - (a_{j+1}+b_{j+1}\theta) + \lambda_j && = 0 \\
\lambda_{j+2} &= 1/r_{j+1} - (a_{j+2}+b_{j+2}\theta) + \lambda_{j+1} && = 0
\end{aligned}$$

and $\bar{r}_j > r_{j+1}$, where the subscript on $\bar{r}$ denotes the highest index of the r's in the average, then the resulting average $\bar{r}_{j+1}$ is less than the previous average $\bar{r}_j$, and $1/\bar{r}_{j+1} > 1/\bar{r}_j$ to increase the value of $\lambda_{j-\ell+1}$ through λ_{j+1}. Consequently, if it is necessary to include r_{j+1} into the average, then the previous ordering will not be changed.

3. The procedure terminates after at most N-2 averagings.

Lemma 2.2. For NLP Problem 1, if $a_{m+1} \neq 0$ and $b_{n+2} \neq 0$ or equivalently, $\Sigma_{j=m+1}^{N} \delta_i > 0$ and $\Sigma_{j=n+2}^{N}(1-\delta_j) > 0$ or equivalently, $\Sigma_{j=1}^{m} \delta_j \neq m$ and $\Sigma_{j=1}^{n+1}(1-\delta_j) \neq n$, then there exist $(\underline{r}, \underline{\lambda}, \theta)$ which satisfies the K-T conditions.

Proof: For any fixed $\theta > 0$, the sequential procedure guarantees a set of ordered r_i (functions of θ) which are feasible, and the vector $(\underline{r}, \underline{\lambda})$ satisifes K-T condition (2) and (3) with the exception of (2.7). But each of the 2^{N-2} possible orderings or averagings of $\underline{r}$ gives a constrained estimate of the form $r_i = 1/(\bar{a}_{i+1}+\bar{b}_{i+1}\theta)$, $i = 1, \ldots, N-1$, so that, if the condition $\Sigma_{i=1}^{N-1} \theta b_{i+1} r_i = n$ is satisfied for each possible ordering, then there exist $(\underline{r}, \underline{\lambda}, \theta)$ satisfying the K-T conditions.

Now

$$\Sigma_{i=1}^{N-1} \theta b_{i+1} r_i = \Sigma_{i=1}^{N-1} \theta b_{i+1}/(\bar{a}_{i+1}+\bar{b}_{i+1}\theta) = \Sigma_{i=1}^{N-1} \theta \bar{b}_{i+1}/(\bar{a}_{i+1}+\bar{b}_{i+1}\theta)$$

in which at least the first $n+1$ terms are not identically zero by assumption $b_{n+2} > 0$, and fewer than n of the last terms in the summation are identically 1 by assumption $a_{m+1} > 0$. The terms not identically zero or one are positive increasing functions of θ with range $(0, 1)$, so that the sum must have the interval $(n-1, n+1)$ within its range.

Theorem 2.3. For NLP Problem 1, if $a_{m+1} \neq 0$ and $b_{n+2} \neq 0$ then the vector $(\hat{\underline{r}}\ \underline{\lambda}\ \hat{\theta})$ satisfying the K-T conditions is unique.

Proof: For any finite $\theta > 0$ the sequential procedure finds a feasible vector $\underline{r}$ in which the components are ordered with the form $r_i = 1/(\bar{a}_{i+1}+\bar{b}_{i+1}\theta)$, $i = 1, \ldots, N-1$, and satisfy the relationship $\Sigma_{i=1}^{N-1}(a_{i+1}+b_{i+1}\theta)r_i = N-1$. Then for $0 < \theta < \infty$ and the corresponding ordered vector $\underline{r}$, the logarithmic likelihood function becomes

$$\mathcal{L}_N(\underline{r}, \theta) = \Sigma_{i=1}^{N-1} \ln r_i + n(\ln \theta) - (N-1)$$

being a function of only θ and N for given $\underline{a}$, $\underline{b}$, and may be denoted by

$$\mathcal{L}_N(\underline{r}(\theta), \theta).$$

The dependence of the ordered r's on θ is as follows. Given the initial constants $\underline{a}$ and $\underline{b}$ then for given θ, $0 < \theta < \infty$, we find unconstrained estimators $r_1, \ldots, r_{N-1}$ where $r_i = 1/(a_{i+1}+b_{i+1}\theta)$, $i = 1, \ldots, N-1$. When the vector $\underline{r}$ is constrained, some of the adjacent components r_j, r_{j+1} will be set equal, which in turn requires the corresponding constants a_j, a_{j+1} and b_j, b_{j+1} to be arithmetically averaged. Since there are $N-1$ components, there are at most 2^{N-2} possible averagings in which

r_j may or may not be averaged harmonically with r_{j+1}. Then viewing the vector $\underline{r}(\theta)$ as a function of θ, as θ varies from 0^+ to ∞, the orderings or averagings of $r(\theta)$ will in general change with θ. But there are at most 2^{N-2} values of θ, say $0 < \theta_1 < \theta_2 < \ldots < \theta_{N_2} < \infty$ where $N_2 \leq 2^{N-2}$, at which the ordering of $\underline{r}(\theta)$ will change.

Such a change in ordering will involve either averaging or "de-averaging" components r_{j-1} and r_j, say, as θ crosses one of the points θ_k. But this necessitates that the average and unaveraged values of r_{j-1} and r_j are equal at θ_k. The crucial points for the proof are that both the upper and the lower limits of the ordered components of $\underline{r}(\theta)$ are equal at each θ_k and that the constants in each constrained $r_j(\theta)$ are a function of only the points $\theta_1, \ldots, \theta_{N_2}$ and initial vectors $\underline{a}$ and $\underline{b}$.

Consider now $\mathcal{L}_N(\tilde{r}(\theta), \theta)$ where $\underline{\tilde{r}}(\theta)$ is any one of the 2^{N-2} possible orderings. The hypothesis $b_{n+2} \neq 0$ implies that there are at least $n+1$ terms in $\Sigma_{i=1}^{N-1} \ln r_i(\theta)$ of the form $-\ln(\bar{a}_{i+1}+\bar{b}_{i+1}\theta)$ with $\bar{b}_{i+1} \neq 0$ so that, $\lim_{\theta \to \infty} \mathcal{L}_N(\underline{\tilde{r}}(\theta), \theta) = -\infty$. Similarly, the hypothesis $a_{m+1} \neq 0$ implies that there are fewer than n terms in $\Sigma_{i=1}^{N-1} \ln \tilde{r}_i(\theta)$ of the form $-\ln(\bar{a}_{i+1}+\bar{b}_{i+1}\theta)$, $\bar{a}_{i+1} = 0$, so that $\lim_{\theta \to 0} \mathcal{L}_N(\underline{\tilde{r}}(\theta), \theta) = -\infty$.

For $\theta \in (0, \infty)$, $\mathcal{L}_N(\underline{r}(\theta), \theta)$ is a differentiable function. Since

$$\mathcal{L}_N(\underline{r}(\theta), \theta) = n \ln \theta - \Sigma_{i=1}^{N-1} \ln(\bar{a}_{i+1} + \bar{b}_{i+1}\theta) + (N-1)$$

and

$$\mathcal{L}'_N(\underline{r}(\theta), \theta) = n/\theta - \Sigma_{i=1}^{N-1} \bar{b}_{i+1}/(\bar{a}_{i+1} + \bar{b}_{i+1}\theta)$$

$$= n/\theta - \Sigma_{i=1}^{N-1} \bar{b}_{i+1} r_i(\theta)$$

$$= n/\theta - \Sigma_{i=1}^{N-1} b_{i+1} r_i(\theta),$$

it is clear that $\mathcal{L}_N$ is differentiable between the points $\theta_1, \theta_2, \ldots, \theta_{N_2}$. But

$$\lim_{\theta \to \theta_k^+} r_i(\theta) = \lim_{\theta \to \theta_k^-} r_i(\theta), \quad \forall\, \theta_k, \quad \forall\, r_i$$

so that

$$\lim_{\theta \to \theta_k^+} \mathcal{L}'_N(\underline{r}(\theta), \theta) = \lim_{\theta \to \theta_k^-} \mathcal{L}'_N(\underline{r}(\theta), \theta) \quad \forall\, \theta_k,\ k = 1, \ldots, N_2,$$

to obtain the result that $\mathcal{L}'_N(\underline{r}(\theta), \theta)$ is differentiable on $(0, \infty)$.

A necessary condition for $\mathcal{L}_N(\underline{r}(\theta), \theta)$ to be maximized is that $\left.\frac{d\mathcal{L}_N}{d\theta}\right|_{\hat{\theta}} = 0$. $n/\hat{\theta} - \Sigma_{i=1}^{N-1} \bar{b}_{i+1}/(\bar{a}_{i+1} + \bar{b}_{i+1}\hat{\theta}) = 0$, $\hat{\theta} \in (0, \infty)$, if and only if

$$h(\hat{\theta}) = \Sigma_{i=1}^{N-1} \{\hat{\theta}\bar{b}_{i+1}/(\bar{a}_{i+1} + \bar{b}_{i+1}\hat{\theta})\} - n = 0. \tag{2.8}$$

$$h'(\hat{\theta}) = \Sigma_{i=1}^{N-1} \bar{a}_{i+1}\bar{b}_{i+1}/(\bar{a}_{i+1} + \bar{b}_{i+1}\hat{\theta})^2 > 0, \quad \hat{\theta} \in (0, \infty), \tag{2.9}$$

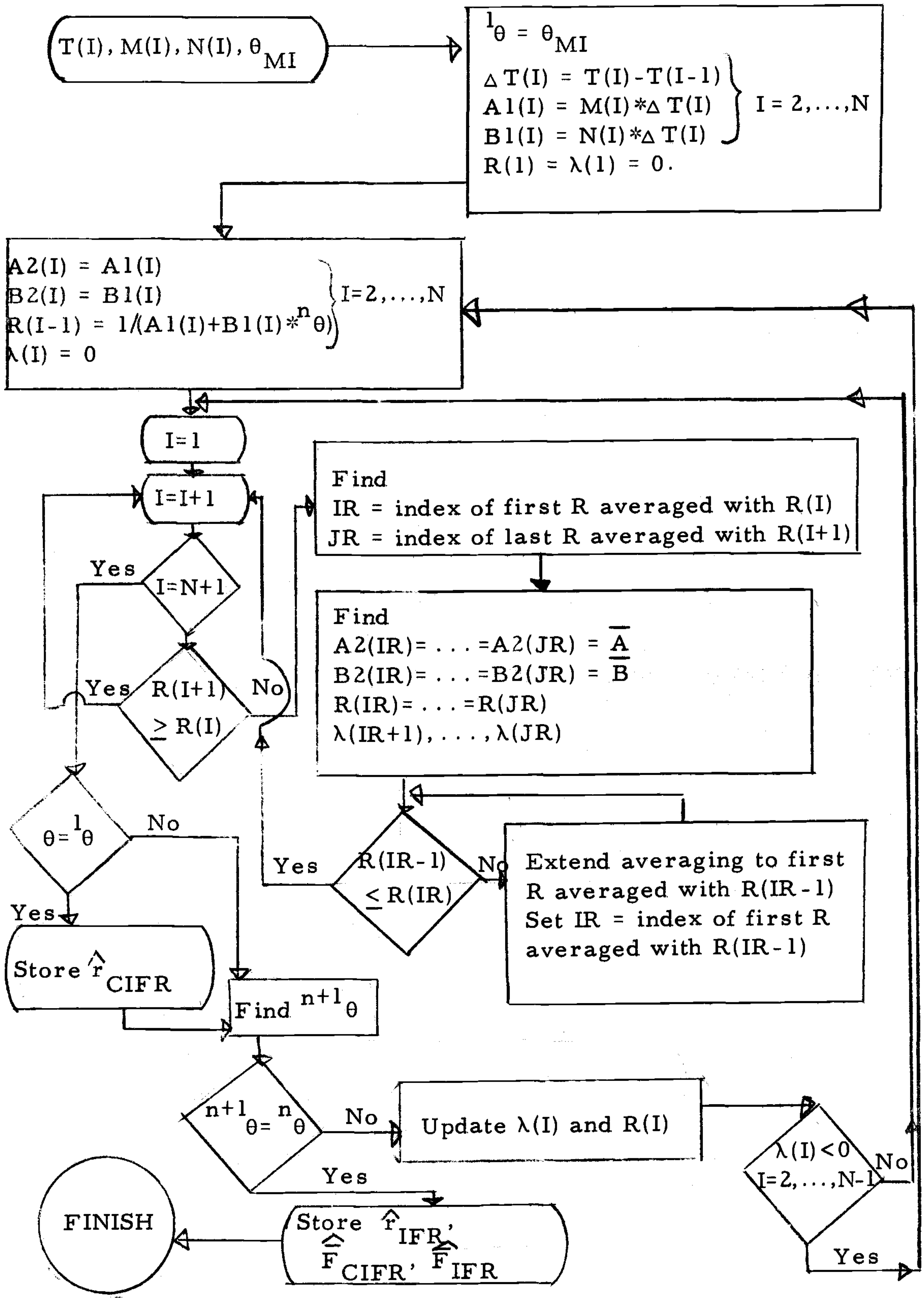
so that by the same argument as in Lemma 2.2 $h(\theta)$ has only one solution. By recognizing that $\left.\frac{d\mathcal{L}_N}{d\theta}\right|_{\underline{r}(\hat{\theta}), \hat{\theta}} = 0$ is an equivalent equation to Equation (2.7) the result follows.

2.5.3. Description of the Algorithm Used for Maximizing IFR Likelihood Function

The algorithm used to find $(\hat{\underline{r}}, \hat{\theta}, \underline{\lambda})$ in the simulations is an adaptation of the sequential procedure for finding $(\underline{r}^*, \underline{\lambda})$ for a given value of θ, which is itself, a two sample analogue of the procedure in Marshall and Proschan (1965). Since the functional form of $\underline{r}(\theta)$ is known, the value of θ may be updated by solving Equation (2.7). Because short computation time is crucial in simulation studies, we maintained the averaged form of $\underline{r}(\theta)$ from one value of θ, say ${}^{n}\theta$, to the updated value ${}^{n+1}\theta$ when we would check to see if future averagings were necessary and if any previous averaging were unnecessary for ${}^{n+1}\theta$. Future averagings were identified by the K-T feasiblity conditions. Unnecessary averagings for ${}^{n+1}\theta$ were identified by negative Lagrangian multipliers λ. For the simple case $r_{i+1} = r_i$ and $\lambda_i = 0$, $\lambda_{i+2} = 0$, $\lambda_{i+1} = 1/\overline{r}_i - (a_{i+1}+b_{i+1}\theta)$ becomes negative if $a_{i+2} + b_{i+2}\,{}^{n+1}\theta < a_{i+1} + b_{i+1}\,{}^{n+1}\theta$. A similar situation occurs for the "multiple averaged" case. When a Lagrangian multiplier became negative (defined to be $< -10^{-10}$) the algorithm maintained the present value of ${}^{n+1}\theta$ and reordered the initial estimators $\underline{r}({}^{n+1}\theta)$, where $r_i = 1/(a_{i+1}+b_{i+1}\,{}^{n+1}\theta)$, $i = 1, \ldots, N-1$.

When the value of θ did not change during a reordering the

Flow Chart for the IFR Algorithm



algorithm terminated. The function $h(\theta)$, whose root the algorithm finds, is of the same form as Equation (2.2b) so that by the same reasoning as in Section 2.2 the Newton-Raphson method will converge.

An alternate algorithm might for each ${}^{n}\theta$ order the initial r's so that the check for $\lambda_i < 0$ would not be necessary. This approach has the advantage of not requiring the vector $\underline{\lambda}$ to be calculated but will usually require more time than the algorithm we used.

2.5.4. Decreasing Failure Rate Distribution

Suppose that $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ are random samples from distributions F and G, let $\underline{t}$ be the combined ordered vector, and let δ_i be an indicator variable having value 1 or 0 according to whether t_i is an F or G observation. Assume also that x and y are positive random variables. Then by use of the same relationships used in the IFR case we can write

$$L_N(\underline{t}, \underline{\delta} | \underline{r}, \theta) = [\Pi_{i=1}^{N} r(t_i)\theta^{1-\delta_i}] \exp\{-\Sigma_{i=1}^{N}[\int_{t_{i-1}}^{t_i} r(u)du \; \Sigma_{j=i}^{N}(\delta_j + \theta(1-\delta_j))]\}.$$

If $L_N(\underline{t}, \underline{\delta} | \underline{r}, \theta)$ is to be maximized subject to the constraint $r(t_0) \geq r(t_1) \geq \cdots \geq r(t_N)$ then $r(u) = r(t_i) = r_i$ for $t_{i-1} < u \leq t_i$ will minimize the exponential terms. Then

$$L_N(\underline{t}, \underline{\delta} | \underline{r}, \theta) = [\Pi_{i=1}^{N} r_i \theta^{1-\delta_i}] \exp[-\Sigma_{i=1}^{N}(a_i + b_i\theta)r_i],$$

where

$$a_i = (t_i - t_{i-1}) \Sigma_{j=i}^{N} \delta_j$$

$$b_i = (t_i - t_{i-1}) \Sigma_{j=i}^{N} (1-\delta_j), \quad i = 1, \ldots, N.$$

Using the same approach as in the IFR case,we express the DFR problem as a NLP problem.

NLP Problem 2.

$$\max \quad \mathcal{L}(\underline{r}, \theta) = \Sigma_{i=1}^{N} [\ln r_i - (a_i + b_i \theta) r_i] + n \ln \theta$$

such that $g_i(\underline{r}, \theta) = r_i - r_{i+1} \geq 0, \quad i = 1, \ldots, N-1,$

$$g_N(\underline{r}, \theta) = r_N \geq 0$$

$$g_{N+1}(\underline{r}, \theta) = \theta \geq 0.$$

Lemma 2.3. For NLP Problem 2, if $a_{n+1} \neq 0$ and $b_{m+1} \neq 0$ then there exist $\underline{\lambda}$ such that $(\underline{r}, \theta)$ satisfies the K-T conditions.

Proof: K-T condition (3) requires

$$\lambda_1 = a_1 + b_1 \theta - 1/r_1$$

$$\lambda_2 = a_2 + b_2 \theta - 1/r_2 + \lambda_1$$

$$\vdots$$

$$\lambda_N = a_N + b_N \theta - 1/r_N + \lambda_{N-1}$$

$$\Sigma_{i=1}^{N} \theta b_i r_i = n$$

after setting $\lambda_{n+1} = 0$. The sequential method of finding $\underline{\lambda}$ is the same as in the IFR case with the obvious difference of having $r_i - r_{i+1} \geq 0$. Existence of ${}^n\theta$ for each of the 2^{N-1} possible orderings is the same as the proof for the existence of $\hat{\theta}_{MI}$ except that the a's and b's may be averaged. There is no bound M. Using a similar argument as in the IFR case we have

<u>Theorem 2.4.</u> For NLP Problem 2, if $a_{n+1} \neq 0$ and $b_{m+1} \neq 0$ then the vector $(\hat{\underline{r}}, \underline{\lambda}, \hat{\theta})$ satisfying the K-T conditions is unique.

2.6. A Worked Example Giving Estimates of θ

Consider the following hypothetical data:

t_i	1	1.5	2.5	2.75
Δt_i	1	.5	1.0	.25
δ_i	1	0	1	0
m_i	2	1	1	0
n_i	2	2	1	1

<u>Combined sample estimate, $\hat{\theta}_C$.</u> Evaluation of

$$\Sigma_{\delta=1} n_i \ln(1 - 1/N_i(\hat{\theta})) = \Sigma_{\delta=0} m_i \ln[1 - \hat{\theta}/N_i(\hat{\theta})]/\hat{\theta}^2$$

gives

$$2 \ln\left(\frac{1+2\hat{\theta}_C}{2+2\hat{\theta}_C}\right) + \ln\left(\frac{\hat{\theta}_C}{1+\hat{\theta}_C}\right) = \ln\left(\frac{1+\hat{\theta}_C}{1+2\hat{\theta}_C}\right) / \hat{\theta}_C^{\,2} ,$$

which has solution $\hat{\theta}_C = .282$.

<u>U-statistic estimate, $\hat{\theta}_U$</u>.

$$\hat{W} = -\frac{-1}{n} \Sigma_{\delta=0} \hat{F} = \frac{1}{2}\left[\frac{1}{2}+0\right] = \frac{1}{4}$$

so that $\hat{\theta}_U = \hat{W}/(1-\hat{W}) = 1/3$.

<u>Maximal invariant estimate, $\hat{\theta}_{MI}$</u>. For this data the equation $\Sigma_{i=1}^{N} \dfrac{n_i\theta}{m_i+n_i\theta} = n$ is

$$\frac{2\theta}{2+2\theta} + \frac{2\theta}{1+2\theta} + \frac{\theta}{1+\theta} + \frac{\theta}{\theta} = 2,$$

or equivalently, $4\hat{\theta}^2 + \hat{\theta} - 1 = 0$ which has approximate solution $\hat{\theta}_{MI} = .39$.

<u>Increasing failure rate, $\hat{\theta}_{IFR}$</u>. We need to find $\underline{\lambda}$ such that

$$\lambda_2 = 1/r_1 - (a_2+b_2\theta) = 1/r_1 - (.5+\theta)$$

$$\lambda_3 = 1/r_2 - (a_3+b_3\theta) + \lambda_2 = 1/r_2 - (1+\theta) + \lambda_2$$

$$\lambda_4 = 1/r_3 - (a_4+b_4\theta) + \lambda_3 = 1/r_3 - (.25) + \lambda_3$$

and

$$\Sigma_{i=1}^{N-1} b_{i+1}\theta r_i = n.$$

Step 1. Set $\lambda_i = 0$. Find r_i to solve the last equation $\frac{\theta}{.5+\theta} + \frac{\theta}{1+\theta} + \frac{.25\theta}{.25\theta} = 2$ to find $\theta = \sqrt{.5} = .707$.

Step 2. Check the feasibility of r_i: $r_1 = .823$, $r_2 = .584$, $r_3 = 4$. Since $r_1 > r_2$, r_1 and r_2 are averaged to obtain $r_1 = r_2 = 1/(.75+\theta) = .685$ and $\lambda_2 = .96 - \theta = .25$, $\lambda_3 = \lambda_4 = 0$.

Step 3. Solve $\frac{2\theta}{.75+\theta} = 1$ to find $\theta = .75$.

Step 4. Is $\theta_{new} = \theta_{old}$? No.

Step 5. Update $\underline{r}$, $\underline{\lambda}$: $r_1 = r_2 = 2/3$ and $r_3 = 4$, $\lambda_2 = .21$ and $\lambda_3 = \lambda_4 = 0$. Check to see if $\lambda_i \geq 0$.

Step 6. Check feasibility of r_i: $r_1 = r_2 = 2/3 < r_3 = 1$.

Step 7. Solve $\Sigma_{i=1}^{N-1} \theta b_{i+1} r_i = n$. Stop since $\theta_{new} = \theta_{old}$.

<u>Decreasing failure rate</u>. We need to find $\underline{\lambda}$ such that

$$\lambda_1 = a_1 + b_1\theta - 1/r_1 = (2+2\theta) - 1/r_1$$

$$\lambda_2 = a_2 + b_2\theta - 1/r_2 + \lambda_1 = (.5+\theta) - 1/r_2 + \lambda_1$$

$$\lambda_3 = a_3 + b_3\theta - 1/r_3 + \lambda_2 = (1+\theta) - 1/r_3 + \lambda_2$$

$$\lambda_4 = a_4 + b_4\theta - 1/r_4 + \lambda_3 = .25\theta - 1/r_4 + \lambda_3$$

and

$$\Sigma_{i=1}^{N} b_i \theta r_i = n.$$

Step 1. Set $\underline{\lambda} = 0$, find $\underline{r}$ and then solve $\Sigma_{i=1}^{N} b_i \theta r_i = 2$. This is the same equation as in $\hat{\theta}_{MI}$, so that $\theta = .40$.

Step 2. Check the feasibility condition and average where necessary.

$$r_1 = 1/(2+2(.4)) = .357$$

$$r_2 = 1/(.5+.4) = 1.11$$

$$r_3 = 1/(1+.4) = .71$$

$$r_4 = 1/(.25)(.4) = 10;$$

(a) $r_1 < r_2$, set $\bar{r}_1 = \bar{r}_2 = .54$

(b) $\bar{r}_2 < r_3$, set $\bar{r}_1 = \bar{r}_2 = \bar{r}_3 = .588$

(c) $\bar{r}_3 < r_4$, set $\bar{r}_1 = \bar{r}_2 = \bar{r}_3 = \bar{r}_4 = 1/[.875+1.06(.4)] = .63.$

Step 3. Update value of θ. $\Sigma_{i=1}^{N} b_i \theta r_i = \dfrac{4(1.06)\theta}{.875 + 1.06\theta} = 2$ to give $\theta = .825$. Continue since $\theta_{new} = \theta_{old}$.

Step 4. Update values of $\underline{r}$ and $\underline{\lambda}$ using the ordering from Step 2. $\lambda_1 = 1.900$, $\lambda_2 = 1.475$, $\lambda_3 = 1.550$, $\lambda_4 = .03$. (Because $\lambda_i \geq 0$, $i = 1, \ldots, 4$ we know that the orderings in Step 2 has no unnecessary averagings for θ_{new}.)

Step 5. Check the feasibility conditions.

Step 6. Update θ. Algorithm terminates because θ remains the same.

3. EVALUATION OF THE MONTE CARLO STUDY

3.1. Description of Program Used for Empirical Study

A Monte Carlo study of the estimators was carried out on Oregon State University's CDC3300 computer. The system of programs consisted of a main program with several subroutines which had access to common arrays. The main program generated two random samples from Weibull distributions, and calculated summarizing statistics for the various estimators. An ordering subroutine constructed an array containing the combined ordered sample, with corresponding arrays for the indicator variables δ_i, $m_i = \Sigma_{j=i}^{N} \delta_j$ and $n_i = \Sigma_{j=i}^{N}(1-\delta_j)$. The estimator $\hat{\theta}_U$ was also calculated in the ordering subroutine. The next subroutine used $\hat{\theta}_U$ as the starting value to iteratively find the maximal invariant MLE, $\hat{\theta}_{MI}$. Using $\hat{\theta}_U$ as a starting value, the third subroutine found $\hat{\theta}_C$, which is the MLE of θ with respect to the combined sample. In the first stage of the fourth subroutine the conditional maximum likelihood estimator for $r_F(t)$ assuming IFR was calculated; the estimator was conditional on $\hat{\theta}_{MI}$ so that given $\hat{\theta}_{MI}$, the basic estimates r_i, $i = 1, \ldots, N-1$, were constrained to be increasing to obtain $\hat{r}_{CIFR}(\cdot)$. In the second stage the MLE of θ, ($\hat{\theta}_{IFR}$) and the unconditional MLE $\hat{r}_{IFR}(\cdot)$ were obtained. In the third stage the cummulative failure rate functions (corresponding to $\hat{r}_{CIFR}$ and

$\hat{r}_{IFR}$) were obtained, and subsequently the estimates of the reliability functions $\bar{F}_{CIFR}$, $\bar{F}_{IFR}$, $\bar{F}_M$, and $\bar{F}_C$ at $t_i = .2, .4, \ldots, 2.0$, where $\bar{F}_M$ is the empirical reliability function based only on the single sample $x_1, \ldots, x_m$.

Our objective for the empirical study was to compare the estimators under changes in θ, in sample size, and in the intensity of "increasingness" of the failure rate functions. All estimators are scale invariant so that the Weibull distributions with scale parameter equal one, $F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$, were used for the study. Four basic combinations of (c, θ) were considered: (1, 1), (3, 1), (1, 3), (3, 3). These combinations allow pairwise comparisons when only θ or only c is changed. When $c = 1$ the distributions have monotone failure rate, and when $c = 3$ the distributions have strictly increasing failure rate. When $\theta = 1$ and c is constant both distributions are identical; when $\theta < 1$ and c constant $G(t) < F(t)$, when $\theta > 1$ and c constant $G(t) > F(t)$ to make the y observations from G stochastically less than the x observations from F. We used equal samples sizes $m = n = 10, 20,$ and 40.

One mild restriction was imposed on the samples to guarantee solutions to all procedures. All procedures required that the two samples overlap by at least one observation, that is $(\Sigma_{j=1}^{m} \delta_j \neq m$ and

$\Sigma_{j=1}^{n}(1-\delta_j) \neq n)$ with the exception of the IFR case which required $\Sigma_{j=1}^{m}\delta_j \neq m$ and $\Sigma_{j=1}^{n+1}(1-\delta_j) \neq n$. Any sample not satisfying the stronger restriction was rejected.

Comparison of the failure rate function estimators and reliability function estimators were made at time points .2, .4, .6, . . . , 2.0. For each sample, the failure rate function was estimated only on $[0, t_N)$ and is given value of ∞ on $[t_N, \infty)$. If t_N was less than 2.0, the estimate on $[t_{N-1}, t_N)$ was continued to 2.0. A count of the number of "infinities" on each subinterval was recorded.

3.2. Comparison of MLE of θ

Several trends can be observed from the data contained in Tables I-IV, of which many are as expected and will be mentioned only for completeness. The following observations are appropriate.

1. All estimators have a positive bias. This is as expected because "high" estimates have range (θ, ∞) while "low" estimates have range $(0, \theta)$, and hence the arithmetic mean will reflect the bias of the "high" estimates.
2. In general, the bias decreases with increasing sample size but there are two exceptions. For the case $c = 1$, $\theta = 1$ the bias of both $\hat{\theta}_{MI}$ and $\hat{\theta}_C$ increase slightly when the sample size is increased from 20 to 40. (The MSE decreases, however.)

Table I. Empirical moments of four maximum likelihood estimators for θ when true distributions are Weibull.

These results are based on 200 simulations using $F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 1$, $c = 1$.

$m = n = 10$

	$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
Mean	1.147584	1.164617	1.177045	1.432509
Variance	.442442	.524659	.567755	1.270405
3rd central	.600793	.809265	.922167	3.754207
4th central	1.755142	2.556371	3.067744	20.923734

$m = n = 20$

	$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
Mean	1.012018	1.049156	1.014489	1.079914
Variance	.119046	.161899	.133426	.194833
3rd central	.035870	.060349	.045451	.091208
4th central	.068438	.107658	.089054	.210491

$m = n = 40$

	$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
Mean	1.021612	1.018193	1.023609	1.060344
Variance	.064036	.071219	.068635	.094642
3rd central	.026487	.026030	.029532	.051035
4th central	.040286	.036835	.046134	.091277

Table II. Empirical moments of four maximum likelihood estimators for θ when true distributions are Weibull.

These results are based on 200 simulations using $F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 1$, $c = 3$.

$m = n = 10$

	$\hat{\theta}_{MI}$	$\hat{\theta}_{U}$	$\hat{\theta}_{C}$	$\hat{\theta}_{IFR}$
Mean	1.139709	1.167642	1.166007	1.356748
Variance	.355858	.435565	.451970	.787063
3rd central	.344443	.481708	.521000	1.603561
4th central	.838867	1.288920	1.412062	6.791009

$m = n = 20$

	$\hat{\theta}_{MI}$	$\hat{\theta}_{U}$	$\hat{\theta}_{C}$	$\hat{\theta}_{IFR}$
Mean	1.047312	1.085869	1.051300	1.118398
Variance	.138327	.219811	.153205	.199563
3rd central	.044045	.147604	.051978	.094278
4th central	.064398	.285439	.077939	.166487

$m = n = 40$

	$\hat{\theta}_{MI}$	$\hat{\theta}_{U}$	$\hat{\theta}_{C}$	$\hat{\theta}_{IFR}$
Mean	1.017211	1.021277	1.018373	1.051211
Variance	.054346	.069241	.058246	.068007
3rd central	.006457	.011127	.007409	.009897
4th central	.009001	.017137	.010319	.014523

Table III. Empirical moments of four maximum likelihood estimators for θ when true distributions are Weibull.

These results are based on 200 simulations using $F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 3$, $c = 1$.

$m = n = 10$

	$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
Mean	3.917476	4.294263	4.474457	7.412966
Variance	7.013842	21.372088	11.296877	50.979839
3rd central	37.817435	544.395622	83.634934	1012.919882
4th central	387.699589	21523.166416	1088.316865	30684.642187

$m = n = 20$

	$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
Mean	3.302968	3.428827	3.483258	4.523752
Variance	1.926413	3.335900	2.354526	4.631293
3rd central	5.410304	18.858883	7.747019	20.031946
4th central	38.310249	217.578604	61.650333	214.644838

$m = n = 40$

	$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
Mean	3.130677	3.183356	3.207650	3.826419
Variance	.544874	.758789	.591893	1.070823
3rd central	.120377	.554551	.137649	.494119
4th central	.732876	2.197108	.866973	3.228011

Table IV. Empirical moments of four maximum likelihood estimators for θ when true distributions are Weibull.

These results are based on 200 simulations using $F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 3$, $c = 3$.

$m = n = 10$

	$\hat{\theta}_{MI}$	$\hat{\theta}_{U}$	$\hat{\theta}_{C}$	$\hat{\theta}_{IFR}$
Mean	3.916268	4.387027	4.455282	6.908776
Variance	6.621895	21.645630	10.545860	59.545095
3rd central	35.919833	552.789688	78.424926	1830.691414
4th central	374.598958	21724.933405	1039.436698	86197.881708

$m = n = 20$

	$\hat{\theta}_{MI}$	$\hat{\theta}_{U}$	$\hat{\theta}_{C}$	$\hat{\theta}_{IFR}$
Mean	3.470803	3.585685	3.672115	4.475879
Variance	2.843463	3.626121	3.597574	9.405419
3rd central	13.246012	18.920859	20.595340	153.271707
4th central	130.970442	203.916666	239.933249	4178.040535

$m = n = 40$

	$\hat{\theta}_{MI}$	$\hat{\theta}_{U}$	$\hat{\theta}_{C}$	$\hat{\theta}_{IFR}$
Mean	3.142702	3.178296	3.221978	3.540186
Variance	.773018[illegible]	1.476249	.842575	1.171490
3rd central	.845420[illegible]	5.997962	.981437	1.826236
4th central	3.288367[illegible]	50.913108	3.975654	8.800516

Table V. Empirical correlation matrices of four maximum likelihood estimators for θ when true distributions are Weibull.

These results are based on 200 simulations using $F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 1$, $c = 1$.

$m = n = 10$

$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
1.000000	.991648	.999433	.993940
	1.000000	.991142	.986105
		1.000000	.995385
			1.000000

$m = n = 20$

$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
1.000000	.989748	.991649	.997323
	1.000000	.990575	.990896
		1.000000	.998105
			1.000000

$m = n = 40$

$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
1.000000	.995332	.999793	.998351
	1.000000	.995440	.995901
		1.000000	.998377
			1.000000

Table VI. Empirical correlation matrices of four maximum likelihood estimators for θ when true distributions are Weibull.

These results are based on 200 simulations using $F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 1$, $c = 3$.

n = m = 10

$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
1.000000	.996117	.999357	.990245
	1.000000	.996357	.987449
		1.000000	.991392
			1.000000

n = m = 20

$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
1.000000	.984642	.999600	.996382
	1.000000	.983224	.993053
		1.000000	.995631
			1.000000

n = m = 40

$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
1.000000	.996631	.999765	.999154
	1.000000	.996640	.997389
		1.000000	.999321
			1.000000

Table VII. Empirical correlation matrices of four maximum likelihood estimators for θ when true distributions are Weibull.

These results are based on 200 simulations using $F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 3$, $c = 1$.

m = n = 10

$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
1.000000	.908052	.998913	.984313
	1.000000	.911030	.927427
		1.000000	.990373
			1.000000

m = n = 20

$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
1.000000	.988758	.999687	.998149
	1.000000	.991097	.987997
		1.000000	.997930
			1.000000

m = n = 40

$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
1.000000	.986920	.999969	.994991
	1.000000	.987062	.991460
		1.000000	.995099
			1.000000

Table VIII. Empirical correlation matrices of four maximum likelihood estimators for θ when true distributions are Weibull.

These results are based on 200 simulations using $F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 3$, $c = 3$.

$m = n = 10$

$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
1.000000	.909773	.998808	.957368
	1.000000	.914019	.974756
		1.000000	.963833
			1.000000

$m = n = 20$

$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
1.000000	.993166	.999443	.969116
	1.000000	.992712	.960467
		1.000000	.975656
			1.000000

$m = n = 40$

$\hat{\theta}_{MI}$	$\hat{\theta}_U$	$\hat{\theta}_C$	$\hat{\theta}_{IFR}$
1.000000	.951939	.999950	.998014
	1.000000	.952923	.960272
		1.000000	.998286
			1.000000

3. The variance and the mean square error of all estimators decrease with increasing sample size. The variance decreases by about $1/\sqrt{2}$ for each doubling of the sample size, with a greater decrease for the cases with $\theta = 3$.
4. The MSE of all estimators increases with θ.
5. When $\theta = 1$ there is little change in MSE of the estimators between the case where $c = 1$ and $c = 3$. However, when $\theta = 3$ there is generally an increase in MSE when $c = 3$.
6. On the basis of MSE, the best estimator of θ is $\hat{\theta}_{MI}$ because it is uniformly best over all cases. The next best estimator is $\hat{\theta}_C$, which is almost uniformly better than the remaining two estimators; the exception is in the case $c = 1$, $\theta = 1$, $n = m = 10$ where $\hat{\theta}_U$ is slightly better.

3.3. Comparison of Two Estimators of the Failure Rate Function

Two estimators of the F distribution failure rate function are compared at time points .2 units apart on the fixed interval (0,2.0). There are two difficulties encountered in comparing failure rate function estimators on a fixed interval, and both are concerned with the estimation on the latter part of the interval. As noted in the introduction to Chapter 3, the function is only estimated on $(0, t_N)$ for each sample and we use the estimate on $[t_{N-1}, t_N)$ to also estimate on $[t_N, 2.0)$. For high values of θ and/or c the number of such extended estimates become large. The second difficulty concerns the extremely large estimate that can occur near the end of estimating interval $[0, t_N)$. The basic estimate of the failure rate on, $[t_i, t_{i+1})$, for instance, is the reciprocal of the total exposure time between $[t_i, t_{i+1})$. If t_i and t_{i+1} are very close together, then the basic estimate will be very large. Provided t_i is in the first part of the estimating interval, a large estimate will be dampened when averaged with succeeding basic estimators. However, if a large basic estimate occurs near t_N, then it will be less likely to be averaged. In fact, if the basic estimate on $[t_{n-1}, t_N)$ is extremely large, it will not be dampened. As the parameter c is increased, the probability of obtaining a large basic estimate near t_N is also increased. In practice, one would be skeptical of MLE of an increasing failure rate function near t_N.

In the Monte Carlo study, the number of extended estimates and the number of "extreme" estimates on the interval $[0, 2.0)$ are both increased by:

(i) small sample sizes,

(ii) large values of c,

(iii) large values of θ. In each of these cases we can expect extreme estimates on the latter part of $[0, 2.0)$. In comparing two estimators of a failure rate function, we give more weight to the estimation of the first portion of $[0, 2.0)$.

On the basis of the data in Tables IX-XII, the following conclusions seem to be appropriate.

1. The MSE of both estimators decreases with increasing sample size.
2. The MSE of both estimators decreases initially, and then increases on $[0, 2.0)$ when c is increased from 1 to 3.
3. On the basis of MSE at the chosen time points, one estimator is not uniformly better than the other. The conditional MLE of r_N is slightly better for $c = 1$, $\theta = 1$ and the MLE of r_N is slightly better for $c = 3$, $\theta = 3$. The intermediate cases are indeterminate.
4. The bias of $\hat{r}_{CIFR}$ is generally less than the bias of $\hat{r}_{IFR}$ on the initial portion of $[0, 2.0)$.

3.4. Comparison of Four Estimators of the Reliability Function $\bar{F}$

One important application resulting from this study is having access to combined sample estimates of the F-reliability function when $(F, G) \in \mathcal{L}_e$. The purpose of this section of the empirical work is to consider three combined sample estimators of $\bar{F}$, and compare them to the single sample empirical reliability function, denoted by $\hat{\bar{F}}_M$. The first estimator $\hat{\bar{F}}_C$, which we call the "combined sample MLE of $\bar{F}$," is calculated by

Table IX. Two IFR maximum likelihood estimators of the F failure rate when true distributions are Weibull.

These results are based on 200 simulations using

$F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 1$, $c = 1$.

The minimum MSE for each t is underlined.

m = n = 10

t	Conditional MLE		MLE		No. of
	Bias	MSE	Bias	MSE	Infinities
.2	-0.177209	.183363	-0.221106	.220632	0
.4	-0.063880	.182243	-0.109970	.220250	0
.6	.011465	.204439	-0.038826	.242640	0
.8	.068295	.240410	.018060	.279653	0
1.0	.157539	.276094	.111285	.314139	0
1.2	.247278	.357776	.201372	.379583	0
1.4	.360625	.715104	.319702	.747836	0
1.6	.677754	7.395917	.642786	7.471170	1
1.8	.914063	9.664267	.884458	9.740837	2
2.0	1.254949	15.386455	1.236303	15.795541	5

m = n = 20

t	Conditional MLE		MLE		No. of
	Bias	MSE	Bias	MSE	Infinities
.2	-0.102930	.109266	-0.117488	.125381	0
.4	-0.005277	.092315	-0.020229	.108249	0
.6	.046235	.103198	.030743	.119126	0
.8	.091978	.143013	.077618	.163290	0
1.0	.143651	.238688	.130762	.275350	0
1.2	.184370	.264472	.171897	.299642	0
1.4	.237424	.402362	.224737	.431008	0
1.6	.340754	1.143442	.333584	1.345357	0
1.8	.446684	1.348540	.442795	1.559691	0
2.0	.551392	1.748920	.554368	2.035708	2

m = n = 40

t	Conditional MLE		MLE		No. of
	Bias	MSE	Bias	MSE	Infinities
.2	-0.100861	.043925	-0.112769	.049370	0
.4	-0.033072	.031692	-0.045500	.036054	0
.6	.002180	.036400	-0.009964	.041279	0
.8	.031380	.046556	.019607	.052349	0
1.0	.041104	.048623	.030269	.055355	0
1.2	.060149	.052870	.049394	.059362	0
1.4	.083009	.057707	.072034	.062747	0
1.6	.124200	.086178	.113573	.092733	0
1.8	.149761	.099250	.139728	.105733	0
2.0	.189836	.131121	.179928	.136032	0

Table X. Two IFR maximum likelihood estimators of the F failure rate when true distributions are Weibull.

These results are based on 200 simulations using $F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 1$, $c = 3$. The minimum MSE for each t is underlined.

m = n = 10

t	Conditional MLE Bias	Conditional MLE MSE	MLE Bias	MLE MSE	No. of Infinities
.2	-0.65853	.023631	-0.068375	.022811	0
.4	-0.100425	.125825	-0.124801	.121130	0
.6	-0.018754	.448518	-0.076209	.467853	0
.8	.027736	1.062344	-0.080042	1.086036	0
1.0	.547145	5.267537	.400952	5.937668	0
1.2	5.071730	901.707627	4.836420	894.425326	5
1.4	11.364133	1264.879704	10.776984	1215.139534	56
1.6	21.220150	3560.699075	20.274914	3282.564486	158
1.8	20.163981	3563.656241	19.188605	3282.535260	195
2.0	17.883981	3476.906885	16.908605	3200.233620	200

m = n = 20

t	Conditional MLE Bias	Conditional MLE MSE	MLE Bias	MLE MSE	No. of Infinities
.2	-0.052403	.018129	-0.054036	.017801	0
.4	-0.026936	.098378	-0.038235	.096938	0
.6	-0.003348	.191405	-0.028267	.195796	0
.8	.093783	.647476	.046461	.649311	0
1.0	.286951	1.470937	.213431	1.479089	0
1.2	.811075	5.475432	.715024	5.428643	0
1.4	10.564663	3650.108902	10.627350	3912.321928	16
1.6	27.422457	7760.333511	27.293115	7873.969649	99
1.8	35.374923	15283.120859	35.216670	15388.270795	178
2.0	33.094923	15127.009606	32.936670	15232.881181	198

m = n = 40

t	Conditional MLE Bias	Conditional MLE MSE	MLE Bias	MLE MSE	No. of Infinities
.2	-0.048907	.012111	-0.049913	.011990	0
.4	-0.057734	.047817	-0.063527	.047852	0
.6	-0.043211	.105853	-0.057782	.107287	0
.8	-0.024377	.288506	-0.052370	.285083	0
1.0	.044941	.561182	.003191	.559230	0
1.2	.481255	2.747607	.421909	2.682831	0
1.4	2.658527	119.691958	2.576705	116.733859	0
1.6	21.307618	9002.516407	21.184001	9001.582587	53
1.8	30.749555	10614.132569	30.537163	10571.681484	154
2.0	28.778694	10482.581628	28.563312	10440.918018	194

Table XI. Two IFR maximum likelihood estimators of the F failure rate when true distributions are Weibull.

These results are based on 200 simulations using $F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 3$, $c = 1$.

The minimum MSE for each t is underlined.

m = n = 10

t	Conditional MLE		MLE		No. of Infinities
	Bias	MSE	Bias	MSE	
.2	-0.112054	.191522	-0.329539	.259888	0
.4	.032686	.195451	-0.191950	.209319	0
.6	.133641	.276660	-0.068336	.251630	0
.8	.247398	.400891	.076653	.346371	0
1.0	.384073	1.081049	.232909	.989920	0
1.2	.746924	20.460775	.610063	20.343462	3
1.4	.982177	23.370477	.856218	23.259032	5
1.6	2.904931	374.249279	2.793729	374.143143	13
1.8	3.207199	377.012106	3.102148	376.909568	25
2.0	3.693290	387.305008	3.595196	387.200394	42

m = n = 20

t	Conditional MLE		MLE		No. of Infinities
	Bias	MSE	Bias	MSE	
.2	-0.040773	.083362	-0.193370	.111839	0
.4	.036409	.090693	-0.109776	.093724	0
.6	.095842	.113588	-0.035262	.100606	0
.8	.148876	.153033	.032800	.127382	0
1.0	.208937	.203797	.102998	.166378	0
1.2	.371284	2.327480	.277600	2.289528	0
1.4	.395997	2.351634	.310693	2.315802	1
1.6	.466210	2.619430	.384073	2.544228	2
1.8	1.475024	58.262123	1.403201	58.194091	5
2.0	1.625729	59.835761	1.556787	59.769597	12

m = n = 40

t	Conditional MLE		MLE		No. of Infinities
	Bias	MSE	Bias	MSE	
.2	-0.051982	.039203	-0.153905	.060382	0
.4	.006680	.033686	-0.089156	.042308	0
.6	.052202	.045667	-0.036800	.045656	0
.8	.088641	.058824	.010138	.053577	0
1.0	.117321	.065386	.045135	.057047	0
1.2	.165340	.113841	.101483	.102536	0
1.4	.198120	.131467	.140424	.119187	0
1.6	.239595	.204860	.187050	.194551	0
1.8	.286716	.262742	.239905	.253452	0
2.0	.511505	5.038156	.469477	5.029031	1

Table XII. Two IFR maximum likelihood estimators of the F failure rate when true distributions are Weibull.

These results are based on 200 simulations using $F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 3$, $c = 3$.

The minimum MSE for each t is underlined.

m = n = 10

t	Conditional MLE		MLE		No. of
	Bias	MSE	Bias	MSE	Infinities
.2	-0.054814	.016904	-0.070742	.013919	0
.4	-0.035902	.138316	-0.131811	.119256	0
.6	.062948	.477582	-0.168923	.419600	0
.8	.288815	1.730615	-0.032209	1.651099	0
1.0	1.100842	27.787878	.870431	27.893885	1
1.2	12.689519	6486.379895	12.524690	6483.278612	24
1.4	26.153085	11587.866827	26.023763	11584.492825	108
1.6	47.513944	51897.277165	47.392521	51894.291625	175
1.8	45.619662	51707.330718	45.503558	51704.761552	194
2.0	43.339662	51504.503449	43.223558	51502.463720	199

m = n = 20

t	Conditional MLE		MLE		No. of
	Bias	MSE	Bias	MSE	Infinities
.2	-0.042445	.012419	-0.052892	.011054	0
.4	.010408	.080326	-0.056034	.069624	0
.6	-0.001622	.239004	-0.133781	.229202	0
.8	.110124	.616392	-0.083411	.560406	0
1.0	.554456	2.553476	.400895	2.427489	0
1.2	5.042380	1002.656459	4.968023	1002.559103	3
1.4	20.946820	5848.576732	20.908349	5848.547674	49
1.6	28.382147	10129.651805	28.357319	10129.662422	138
1.8	28.300899	10333.707002	28.277380	10333.803173	190
2.0	27.230103	10481.363667	27.206584	10481.567082	199

m = n = 40

t	Conditional MLE		MLE		No. of
	Bias	MSE	Bias	MSE	Infinities
.2	-0.019031	.009411	-0.026517	.008664	0
.4	.001292	.040625	-0.035384	.037314	0
.6	.012291	.104575	-0.066282	.098605	0
.8	.130793	.270902	.017844	.241759	0
1.0	.160171	.874311	.071889	.849777	0
1.2	.393453	4.474137	.351045	4.470226	0
1.4	27.595407	84236.950431	27.581422	84236.926879	10
1.6	92.739775	500130.605621	92.732845	500130.593742	104
1.8	116.752326	613773.416610	116.746486	613773.421249	181
2.0	114.472326	613246.224456	114.466486	613246.255722	200

$$\hat{\bar{F}}_C(s) = \Pi_{i=1}^{k-1} \hat{p}_i \quad \text{for} \quad s \in (t_{k-1}, t_k],$$

where

$$\hat{p}_i = \{1-[\delta_i+(1-\delta_i)\hat{\theta}]/N_i(\hat{\theta})\}^{\delta_i+(1-\delta_i)/\hat{\theta}}.$$

Two estimators based on $\hat{r}_{CIFR}$ and $\hat{r}_{IFR}$ are considered, where

$$\hat{\bar{F}}_{CIFR}(s) = \exp[-\int_0^s \hat{r}_{CIFR}(u)du]$$

and

$$\hat{\bar{F}}_{IFR}(s) = \exp[-\int_0^s \hat{r}_{IFR}(u)du].$$

Although the estimators of the reliability functions are compared at the same points as the failure rate function, it must be recognized that $\hat{\bar{F}}_{CIFR}$ and $\hat{\bar{F}}_{IFR}$ are based on estimates of the cumulative failure rate function, and therefore, reflect the cumulative properties of $\hat{r}_{CIFR}$ and $\hat{r}_{IFR}$ respectively, rather than the point properties of the previous section.

The following conclusions are appropriate to the data presented in Tables XIII-XX.

1. The MSE for all estimators decrease approximately at the rate of n^{-1}.
2. Both $\hat{\bar{F}}_C$ and $\hat{\bar{F}}_{IFR}$ have greater MSE for $\theta = 3$ than for $\theta = 1$; while $\hat{\bar{F}}_{CIFR}$ has a smaller MSE for $\theta = 3$, than for $\theta = 1$ when $c = 1$, but is greater when $c = 3$.

Table XIII. The bias of four maximum likelihood estimators of the F reliability function when true distributions are Weibull.

These results are based on 200 simulations using

$F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 1$, $c = 1$.

The maximum bias for each t is underlined.

m = n = 10

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	-0.002730	-0.000415	.070102	.075769	0
.4	.002179	.005245	.078957	.091306	0
.6	.006688	.011894	.075059	.092904	0
.8	.007671	.013585	.067177	.088989	0
1.0	.015620	.015697	.055198	.079321	0
1.2	.006805	.003950	.041294	.066064	0
1.4	.002903	-0.001382	.027537	.051617	0
1.6	.003603	.001123	.015686	.038146	1
1.8	.002201	-0.000091	.006415	.026716	2
2.0	.001664	-0.000506	-0.001125	.016646	5

m = n = 20

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	-0.000230	-0.006850	.049216	.051393	0
.4	-0.010070	-0.013751	.050705	.055098	0
.6	-0.001311	-0.007820	.043212	.049387	0
.8	-0.004078	-0.009375	.033294	.040661	0
1.0	-0.000379	-0.004627	.024114	.032199	0
1.2	-0.004194	-0.003563	.015964	.024338	0
1.4	-0.002596	-0.003981	.008966	.017279	0
1.6	-0.003396	-0.004986	.002737	.010704	0
1.8	-0.002798	-0.004130	-0.002708	.004651	0
2.0	-0.006085	-0.005055	-0.007433	-0.000849	2

m = n = 40

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	-0.000730	.002178	.040249	.042012	0
.4	.003304	.004340	.043445	.046781	0
.6	.001063	.003684	.038884	.043268	0
.8	.000921	.002457	.031820	.036842	0
1.0	.001370	.001485	.025255	.030585	0
1.2	.001805	.001297	.019500	.024892	0
1.4	.006403	.004861	.014367	.019653	0
1.6	.002978	.002844	.009422	.014465	0
1.8	.002326	.002414	.004987	.009686	0
2.0	-0.000835	.000165	.001443	.005737	0

Table XIV. The bias of four maximum likelihood estimators of the $\bar{F}$ reliability function when true distributions are Weibull.

These results are based on 200 simulations using

$F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 1$, $c = 3$.

The maximum bias for each t is underlined.

$n = m = 10$

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	-0.001531	-0.002236	.005240	.005347	0
.4	-0.008005	-0.005413	.019408	.021860	0
.6	-0.012735	-0.007041	.033061	.041731	0
.8	-0.001295	-0.003395	.035799	.053296	0
1.0	.018620	.015034	.028215	.051624	0
1.2	-0.004639	-0.002024	.000429	.018949	5
1.4	-0.000812	.000118	-0.010042	-0.001463	56
1.6	-0.004639	-0.003555	-0.005259	-0.002808	158
1.8	-0.001432	-0.001026	-0.000390	.000296	195
2.0	-0.000335	-0.000335	.000459	.000710	200

$n = m = 20$

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	-0.002281	-0.001058	.004927	.005023	0
.4	-0.001005	-0.003429	.016691	.017791	0
.6	-0.009985	-0.010933	.016346	.020314	0
.8	-0.013545	-0.014007	.011274	.018976	0
1.0	-0.012629	-0.011570	.003836	.013762	0
1.2	-0.006389	-0.007272	-0.005630	.002741	0
1.4	-0.007062	-0.004398	-0.010307	-0.006127	16
1.6	-0.002139	-0.002188	-0.005937	-0.004788	99
1.8	-0.000182	-0.000377	-0.000776	-0.000541	178
2.0	-0.000335	-0.000197	.000261	.000323	198

$n = m = 40$

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	.000343	.000501	.004760	.004802	0
.4	.001870	.002240	.014360	.014989	0
.6	-0.000110	.002072	.019665	.021892	0
.8	.004954	.005024	.021755	.026060	0
1.0	.001745	.003004	.014036	.019524	0
1.2	.001985	-0.000176	.003200	.007821	0
1.4	-0.000062	-0.000746	-0.002991	-0.000555	0
1.6	.000985	.000828	-0.002989	-0.002284	53
1.8	.000692	.000372	-0.000867	-0.000731	154
2.0	-0.000335	-0.000093	.000016	.000043	194

Table XV. The bias of four maximum likelihood estimators of the F reliability function when true distributions are Weibull.

These results are based on 200 simulations using $F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 3$, $c = 1$.
The maximum bias for each t is underlined.

m = n = 10

t	$\hat{F}_M$	$\hat{F}_C$	$\hat{F}_{CIFR}$	$\hat{F}_{IFR}$	No. of Infinities
.2	.001269	.016122	.054997	.085654	0
.4	.006179	.026748	.056327	.115134	0
.6	.010688	.027015	.044557	.119163	0
.8	.010671	.018440	.028903	.108537	0
1.0	.014120	.012950	.012274	.089027	0
1.2	.002805	-0.000153	-0.000022	.070132	3
1.4	.007903	.001223	-0.009263	.052998	5
1.6	.010603	.004352	-0.016845	.036751	13
1.8	.011201	.005038	-0.021698	.024040	25
2.0	.010664	.005510	-0.025428	.012999	42

m = n = 20

t	$\hat{F}_M$	$\hat{F}_C$	$\hat{F}_{CIFR}$	$\hat{F}_{IFR}$	No. of Infinities
.2	-0.000230	.006894	.035703	.059100	0
.4	-0.006320	.003743	.032217	.073225	0
.6	-0.001311	.007501	.022339	.072172	0
.8	-0.000328	.002399	.012464	.064802	0
1.0	-0.004379	-0.002504	.002092	.052734	0
1.2	-0.000944	-0.001794	-0.007244	.039152	0
1.4	.000153	-0.001778	-0.013755	.027486	1
1.6	.002853	.000209	-0.017426	.018513	2
1.8	.000701	-0.001935	-0.021286	.009183	5
2.0	.000164	-0.002216	-0.022875	.002732	12

m = n = 40

t	$\hat{F}_M$	$\hat{F}_C$	$\hat{F}_{CIFR}$	$\hat{F}_{IFR}$	No. of Infinities
.2	.005269	.004495	.027656	.044054	0
.4	.007179	.006412	.026525	.054279	0
.6	.005813	.007262	.019692	.053340	0
.8	.001296	.005369	.011490	.046986	0
1.0	-0.001004	.002077	.003628	.038317	0
1.2	.001055	.001878	-0.003512	.028807	0
1.4	-0.001846	-0.002540	-0.009337	.019674	0
1.6	-0.001146	-0.001346	-0.013905	.011556	0
1.8	-0.002548	-0.004074	-0.016876	.005096	0
2.0	-0.004460	-0.005326	-0.018970	-0.000332	1

Table XVI. The bias of four maximum likelihood estimators of the $\bar{F}$ reliability function when true distributions are Weibull.

These results are based on 200 simulations using

$F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 3$, $c = 3$.

The maximum bias for each t is underlined.

m = n = 10

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	-0.004531	.000051	.005254	.005945	0
.4	.005495	.002510	.017644	.027389	0
.6	-0.011735	.002359	.017218	.052894	0
.8	-0.004795	.006715	.008831	.071852	0
1.0	.017120	.014363	-0.001353	.059572	1
1.2	.006860	.001802	-0.019239	.013254	24
1.4	-0.004812	-0.007396	-0.018682	-0.007939	108
1.6	-0.002139	-0.003163	-0.005216	-0.002176	175
1.8	.000067	-0.000361	.000835	.001896	194
2.0	.000164	.000161	.001059	.001478	199

m = n = 20

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	-0.004781	-0.000553	.003976	.004510	0
.4	-0.007505	-0.003315	.008434	.015553	0
.6	-0.003985	-0.000978	.008442	.030644	0
.8	.001954	.004871	.009073	.046530	0
1.0	-0.005879	-0.002309	-0.003795	.032855	0
1.2	-0.000889	-0.003599	-0.015753	.004195	3
1.4	.003937	.002084	-0.012731	-0.005989	49
1.6	.001860	.001250	-0.003377	-0.001604	138
1.8	-0.000182	-0.000345	.000184	.000618	190
2.0	-0.000085	-0.000091	.000524	.000649	199

m = n = 40

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	-0.000156	-0.000462	.003155	.003510	0
.4	-0.002380	-0.003673	.003816	.008111	0
.6	-0.008360	-0.001705	.005745	.018550	0
.8	-0.011170	-0.003195	-0.000134	.021277	0
1.0	-0.004254	-0.002372	-0.004569	.016263	0
1.2	.001735	-0.000999	-0.006116	.005914	0
1.4	-0.000437	-0.001897	-0.008425	-0.004149	10
1.6	-0.000639	-0.001062	-0.005893	-0.005050	104
1.8	-0.000432	-0.000467	-0.001016	-0.000864	181
2.0	-0.000335	-0.000335	.000127	.000165	200

Table XVII. The MSE of four maximum likelihood estimators of the F reliability function when true distributions are Weibull.

These results are based on 200 simulations using

$F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 1$, $c = 1$.

The minimum MSE for each t is underlined.

m = n = 10

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	.013551	.008435	.008541	.009567	0
.4	.022098	.015308	.015754	.019188	0
.6	.024214	.018731	.020182	.025959	0
.8	.025509	.022250	.022152	.029539	0
1.0	.023821	.020865	.020947	.029100	0
1.2	.017482	.016802	.017823	.025618	0
1.4	.015008	.013781	.014827	.021608	0
1.6	.013432	.011735	.012454	.018246	1
1.8	.011798	.009948	.010148	.014740	2
2.0	.010333	.008735	.008336	.011950	5

m = n = 20

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	.008157	.005538	.004987	.005416	0
.4	.013358	.010086	.008584	.009822	0
.6	.014220	.010026	.009954	.011799	0
.8	.014281	.010769	.010249	.012360	0
1.0	.013543	.010471	.010007	.012121	0
1.2	.012108	.008897	.009098	.011002	0
1.4	.010720	.008320	.007659	.009465	0
1.6	.008309	.006675	.006518	.007811	0
1.8	.007076	.005934	.005328	.006336	0
2.0	.005118	.004679	.004274	.005034	2

m = n = 40

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	.003532	.002468	.002678	.002887	0
.4	.005424	.003533	.004135	.004655	0
.6	.005941	.004635	.004669	.005431	0
.8	.005344	.004444	.004755	.005659	0
1.0	.005406	.004730	.004658	.005603	0
1.2	.005519	.004712	.004334	.005248	0
1.4	.005175	.004079	.003790	.004606	0
1.6	.004238	.003585	.003208	.003892	0
1.8	.003222	.003016	.002639	.003185	0
2.0	.002691	.002337	.002148	.002570	0

Table XVIII. The MSE of four maximum likelihood estimators of the $\bar{F}$ reliability function when true distributions are Weibull.

These results are based on 200 simulations using

$F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 1$, $c = 3$.

The minimum MSE for each t is underlined.

m = n = 10

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	.000862	.000601	.000088	.000087	0
.4	.006164	.003373	.001979	.001994	0
.6	.015613	.009739	.007946	.008525	0
.8	.022497	.019753	.017522	.020045	0
1.0	.022314	.019166	.019375	.023785	0
1.2	.014892	.013054	.011746	.014715	5
1.4	.005718	.004725	.003577	.004688	56
1.6	.001177	.000975	.000481	.000676	158
1.8	.000149	.000162	.000039	.000062	195
2.0	.000000	.000000	.000006	.000011	200

m = n = 20

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	.000437	.000229	.000065	.000063	0
.4	.003232	.001950	.001439	.001448	0
.6	.009794	.006186	.004799	.005006	0
.8	.012367	.008960	.007925	.008696	0
1.0	.010694	.009721	.009414	.010639	0
1.2	.007276	.005486	.005071	.005796	0
1.4	.002609	.002051	.001589	.001809	16
1.6	.000644	.000389	.000228	.000258	99
1.8	.000129	.000072	.000018	.000021	178
2.0	.000000	.000001	.000002	.000002	198

m = n = 40

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	.000163	.000091	.000047	.000047	0
.4	.001416	.000723	.000726	.000740	0
.6	.004015	.002551	.002551	.002665	0
.8	.005175	.003798	.003792	.004146	0
1.0	.005527	.004256	.004096	.004532	0
1.2	.003385	.002902	.002804	.003083	0
1.4	.001596	.001208	.001140	.001240	0
1.6	.000468	.000337	.000219	.000238	53
1.8	.000084	.000053	.000015	.000017	154
2.0	.000000	.000002	.000000	.000000	194

Table XIX. The MSE of four maximum likelihood estimators of the F reliability function when true distributions are Weibull.

These results are based on 200 simulations using

$F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 3$, $c = 1$.

The minimum MSE for each t is underlined.

m = n = 10

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	.012601	.008637	.006823	.010630	0
.4	.018935	.015734	.013557	.023597	0
.6	.022923	.020624	.018005	.031766	0
.8	.023113	.023636	.019707	.034401	0
1.0	.024975	.024852	.019761	.032964	0
1.2	.021091	.020962	.018156	.029334	3
1.4	.017742	.017274	.016046	.025397	5
1.6	.014806	.014153	.013593	.021267	13
1.8	.014223	.013317	.011245	.017619	25
2.0	.012597	.011568	.008920	.013959	42

m = n = 20

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	.006607	.003503	.003033	.005087	0
.4	.010268	.007120	.005844	.010076	0
.6	.012095	.010150	.007953	.012962	0
.8	.013099	.011499	.009202	.014139	0
1.0	.012536	.011938	.009417	.013661	0
1.2	.010813	.010446	.008893	.012179	0
1.4	.009976	.009543	.007962	.010486	1
1.6	.007623	.007406	.006836	.008791	2
1.8	.006319	.006159	.005981	.007498	5
2.0	.005639	.005526	.005006	.006219	12

m = n = 40

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	.002958	.001468	.001543	.002734	0
.4	.004845	.003049	.002770	.005192	0
.6	.005303	.004318	.003496	.006346	0
.8	.005429	.004658	.003803	.006458	0
1.0	.004913	.004713	.003709	.005882	0
1.2	.004614	.004487	.003484	.005113	0
1.4	.004194	.003970	.003111	.004240	0
1.6	.003600	.003448	.002743	.003512	0
1.8	.003168	.003076	.002424	.002972	0
2.0	.002832	.002791	.002107	.002489	1

Table XX. The MSE of four maximum likelihood estimators of the F reliability function when true distributions are Weibull.

These results are based on 200 simulations using $F(t) = 1 - e^{-t^c}$ and $G(t) = 1 - e^{-\theta t^c}$ where $\theta = 3$, $c = 3$.

The minimum MSE for each t is underlined.

m = n = 10

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	.001114	.000182	.000058	<u>.000054</u>	0
.4	.005487	.002291	<u>.001707</u>	.001804	0
.6	.018301	.009672	<u>.007774</u>	.009626	0
.8	.025242	.022626	<u>.018994</u>	.025675	0
1.0	.023468	.023642	<u>.021877</u>	.030795	1
1.2	.015556	.015212	<u>.012385</u>	.016846	24
1.4	.005432	.005046	<u>.003146</u>	.004280	108
1.6	.001644	.001415	<u>.000405</u>	.000623	175
1.8	.000291	.000219	<u>.000058</u>	.000098	194
2.0	.000049	.000049	<u>.000011</u>	.000020	199

m = n = 20

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	.000647	.000127	.000049	<u>.000046</u>	0
.4	.003751	.001215	<u>.000861</u>	.000904	0
.6	.006900	.004850	<u>.004296</u>	.004975	0
.8	.012939	.010760	<u>.009295</u>	.012037	0
1.0	.012890	.012820	<u>.010322</u>	.013079	0
1.2	.007572	.007301	<u>.006735</u>	.007981	3
1.4	.003419	.003253	<u>.002400</u>	.002799	49
1.6	.000936	.000878	<u>.000364</u>	.000443	138
1.8	.000154	.000135	<u>.000032</u>	.000041	190
2.0	.000012	.000011	<u>.000003</u>	.000004	199

m = n = 40

t	$\hat{\bar{F}}_M$	$\hat{\bar{F}}_C$	$\hat{\bar{F}}_{CIFR}$	$\hat{\bar{F}}_{IFR}$	No. of Infinities
.2	.000212	.000060	.000036	<u>.000035</u>	0
.4	.001370	.000622	<u>.000505</u>	.000517	0
.6	.004191	.002108	<u>.001862</u>	.002136	0
.8	.005711	.004299	<u>.003925</u>	.004525	0
1.0	.004698	.004652	<u>.004394</u>	.005019	0
1.2	.003149	.002986	<u>.002798</u>	.003155	0
1.4	.001342	.001285	<u>.001169</u>	.001299	10
1.6	.000388	.000369	<u>.000201</u>	.000219	104
1.8	.000062	.000061	<u>.000012</u>	.000013	181
2.0	<u>.000000</u>	<u>.000000</u>	<u>.000000</u>	.000001	200

3. For $c = 3$ all three estimators generally have smaller MSE.
4. Using MSE as the basis of comparison, $\hat{F}_{CIFR}$ is the best estimator of $\bar{F}$, since it is uniformly better than all other estimators except for $\hat{F}_C$, which is slightly better for some values in the case $c = 1$, $\theta = 1$, $m = 10$. $\hat{F}_C$ is generally the next best estimator, but is inferior to $\hat{F}_{IFR}$ for some values of t in the case $c = 3$, $\theta = 3$. $\hat{F}_{IFR}$ and $\hat{F}_M$ are competitive with $\hat{F}_{IFR}$ generally better when $c = 3$.
5. Thus both $\hat{F}_C$ and $\hat{F}_{CIFR}$ are better estimators than $\hat{F}_M$, with $\hat{F}_{CIFR}$ being the superior estimator provided the failure rate is nondecreasing (IFR).

A possible explanation for the poor performance of $\hat{F}_{IFR}$ is that the bias of $\hat{F}_{IFR}$ is almost uniformly greater than the bias of $\hat{F}_{CIFR}$. Although the bias of $\hat{r}_{IFR}$ is only slightly greater than the bias of $\hat{r}_{CIFR}$, the cumulative effect makes a dramatic difference between $\hat{F}_{IFR}$ and $\hat{F}_{CIFR}$.

It is instructive to compare the maximal invariant ML equation, that is

$$\Sigma_{i=1}^{N} n_i \hat{\theta}_{MI} / (m_i + \hat{\theta}_{MI} n_i) = n, \tag{2.2}$$

with an equation equivalent to the final equation of the K-T conditions,

$$\Sigma_{i=1}^{N} \tilde{\theta} b_{i+1} r_i = n$$

which, before constraining the vector $\underline{r}$, may be written as

$$\Sigma_{i=2}^{N} n_i \tilde{\theta}_{IFR} / (m_i + \tilde{\theta}_{IFR} n_i) = n. \tag{3.1}$$

Equation (3.1) is the same as (2.2) except that the first term of (2.2) is not in (3.1). This term, incidentally, is deleted by the requirement that $r_0 = 0$, given in Section 2.5. Thus, the unconstrained estimate $\tilde{\theta}_{IFR}$ must be greater than $\hat{\theta}_{MI}$. From the empirical work we observe that both the bias and variance of constrained estimate $\hat{\theta}_{IFR}$ are less desirable than the bias and variance of $\hat{\theta}_{MI}$, leading us to the conclusion that the ML method is not suitable for estimating θ in the IFR case.

If the difficulty is generated by the omission of the weight given by the term $\tilde{\theta} n_1 / (m_1 + \tilde{\theta} n_1)$, then one might consider either the conditional procedure that we use in the IFR case, or perhaps consider a modified MLE scheme in which the term $\tilde{\theta} n_1 / (m_1 + \tilde{\theta} n_1)$ is added to the left side of Equation (3.1). It would be interesting to see if the latter scheme would provide better estimators than the conditional procedure.

Some insight may be gained from the DFR case. We did not include the DFR case in the empirical study because the number of applications for the DFR case appears to be fewer than for the IFR case. However, it would also be interesting to see if the ML estimators are superior to the conditional ML estimators using $\hat{\theta}_{MI}$, since the appropriate equations have the same number of terms.

4. TESTS FOR PROPORTIONAL FAILURE RATE FUNCTIONS

4.1. Introduction

The use of any estimation procedure given in Chapter 2 is contingent upon satisfying the assumption of proportional failure rate functions, or equivalently that $(F, G) \in \mathcal{L}_C$. When sufficient data are available, graphical procedures are an appropriate means to assess the propriety of the proportionality assumption. If a hypothesis test is deemed necessary, then perhaps the standard approach is to consider a more general model than simple proportionality, from which, a Neyman C-α test, a local test, or a likelihood ratio test could be constructed. In any such test, it is desirable to remove as many nuisance parameters as possible, especially values of the distribution functions at the order statistics.

In our research for this dissertation we considered a more general model where distribution functions F and G are related by $\ln \bar{G} = (\theta \ln \bar{F})^{\gamma}$. We tried to develop a locally best rank test of the hypothesis $H_0: \gamma = 1$, but we were unable to find an explicit form for the expected value of a particular function of the order statistics. Although the integration could be done numerically for a grid of values of θ, m and n, the cost was considered to be too great.

Professor Cox considers a generalized model which generates a straight forward test procedure using the likelihood ratio method.

For the uncensored case, the test procedure is based on the rank order statistics, and therefore has the desirable property which we seek.

A likelihood ratio test using a coarser grid than the order statistics is described in Section 4.3. A test procedure established by Thomas (1969) is described in Section 4.5.

4.2. Graphical Methods

The maximum likelihood method of estimation, as used in Chapter 2, can be regarded as a very general procedure since it does not assume a functional form for the failure rate function. However, this procedure may have severe difficulties. All the methods of Chapter 2 require an overlap of at least one observation (the IFR procedure require an overlap of two) to obtain a nontrivial solution. Since the MLE's of θ depend principally upon the observations in the interval of overlap, the procedures are at their best when the overlap between the samples is large, which is the case when θ is close to one. When the overlap is small a second difficulty is encountered if no functional form is assumed; namely, how can one given credulence to the proportionality assumption? Certainly the data will be of little help when the basis of comparison is restricted to a short interval of overlap. Perhaps on this question, one should be willing to use graphical methods which will utilize certain functional forms for the cumulative failure rate. For example, if the estimators of both cumulative failure rate functions are approximately linear with the same intercept, it would be appropriate to assume proportionality in this case even if the interval of overlap is small.

The following are some graphical procedures which may be useful for certain situations to determine the adequacy of the proportionality model.

1. Comparison of a two-sample estimator of a reliability function with a one-sample estimator. Suppose that the (two sample) MLE of $\bar{F}$ and $\bar{G}$ have been obtained with respect to a particular class, and suppose also, that the corresponding one-sample MLE's of $\bar{F}$ and $\bar{G}$ are available. Then a straight forward graphical procedure to check the appropriateness of the proportionality hypothesis is to plot both estimators of $\bar{F}$ and $\bar{G}$ versus t. A discrepancy in the model of proportional failure rate functions will be detected by any large deviation between the two estimators of the reliability function. Cox (1971) uses this procedure with his two sample estimator of $\bar{F}$, $\bar{G}$ on some medical data. If (F, G) are assumed to belong to a class such as $\mathcal{L}_{IFR}$, then single sample IFR-MLE of $\bar{F}$ and $\bar{G}$ would be used if only the proportionality is in question, and perhaps, the product limit estimates if both proportionality and IFR

assumptions are in question.

2. A method using the cumulative failure rate function. A variation of 1. is to plot estimates of the cumulative failure rate function $R(t) = -\ln(1-F(t))$ versus t. Under the model of proportional failure rate functions, the estimators $R(t|\hat{\bar{G}})/\theta$, $R(t|\hat{\bar{F}})$, and $R(t|\hat{\bar{F}}_C)$ would all be estimating the same function of t. Plotting $(\bar{G}^{1/\theta})^{-1}$, $\hat{\bar{F}}^{-1}$ and $\hat{\bar{F}}_C^{-1}$ on semi-log graph paper would be a suitable method.

3. A method using $\hat{\bar{F}}$ and $\hat{\bar{G}}$. Using the terminology of Wilk and Gnanadesikan (1968), a percent (P-P) plot of the empirical reliability functions may be used to assess the adequacy of the model. If for various values of t the sample percentiles of $\bar{F}$ and $\bar{G}$ are plotted on log-log paper, then a linear graph would be indicative of proportional failure rate functions. The slope of the line is an estimate of θ. This method is only suitable if there is considerable overlap of the samples.

4. Residual analysis. If we consider the random walk of the sample distribution described in Section 2.4, we might plot the residual between $\hat{G}(\hat{F}^{-1}(X_{(i)})$ and $G(\hat{F}^{-1}(X_{(i)})|\hat{\theta})$ for $i = 1, \ldots, m$, where $G(u|\hat{\theta}) = 1-(1-u)^{\hat{\theta}}$, $0 \leq u \leq 1$. If $\hat{G}(\hat{F}^{-1}(X_{(i)})) - G(\hat{F}^{-1}(X_{(i)})|\hat{\theta})$ is plotted for each i, i=1, ..., m,

then either a large residual or perhaps a trend in the residuals would lead to a rejection of the proportionality model. This method would be most suitable for the situation in which m and n are large and θ is close to one.

4.3. Likelihood Ratio Test for Proportionality of Failure Functions

Suppose that we have random samples $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ from distribution functions F and G respectively. If the time axis is divided into k intervals $(\tau_1, \tau_2], (\tau_2, \tau_3], \ldots, (\tau_k, \infty)$ where $k \geq 3$ and $\tau_1 = \inf\{\tau : F(\tau) > 0 \text{ or } G(\tau) > 0\}$. Let $m_i(n_i)$ denote the number of $F(G)$ items in the risk set at τ_i, $i = 1, \ldots, k$. Then if $(F, G) \in \mathcal{L}$ we define $p_0 = q_0 = 1$,

$$p_i = P(X > \tau_{i+1} | X > \tau_i), \quad i = 1, \ldots, k-1,$$

$$q_i = P(Y > \tau_{i+1} | Y > \tau_i), \quad i = 1, \ldots, k-1,$$

so that

$$P(\tau_i < X \leq \tau_{i+1}) = \Pi_{j=0}^{i-1} p_j (1-p_i)$$

and,

$$P(\tau_i < Y \leq \tau_{i+1}) = \Pi_{j=0}^{i-1} q_j (1-q_i)$$

for $i = 1, \ldots, k-1$. We can write the likelihood function in terms of these conditional probabilities by

$$L(\underline{m},\underline{n}|\underline{p},\underline{q}) = \Pi_{i=1}^{k-1}[\{\Pi_{j=0}^{i-1}p_j\}^{m_{i+1}}(1-p_i)^{m_i-m_{i+1}}\{\Pi_{j=0}^{i-1}q_j\}^{n_{i+1}}(1-q_i)^{n_i-n_{i+1}}]$$

$$= \Pi_{i=1}^{k-1}\{p_i^{m_{i+1}}(1-p_i)^{(m_i-m_{i+1})}q_i^{n_{i+1}}(1-q_i)^{(n_i-n_{i+1})}\}$$

where

$$p_k = q_k = 0.$$

Now,

$$\ln L(\underline{m},\underline{n}|\underline{p},\underline{q}) = \Sigma_{i=1}^{k-1}\{m_{i+1}\ln p_i + (m_i-m_{i+1})\ln(1-p_i) + n_{i+1}\ln q_i + (n_i-n_{i+1})\ln(1-q_i)\}$$

is a nonnegative sum of concave functions so that $\ln L(\underline{m},\underline{n}|\underline{p},\underline{q})$ is a concave function. To find the vector $(\hat{\underline{p}},\hat{\underline{q}})$ which maximizes $\ln L(\underline{m},\underline{n}|\underline{p},\underline{q})$ we must consider those points interior to the parameter set for which the gradient is zero, and also those points on the boundary.

Setting the gradient equal to zero we obtain

$$\hat{p}_i = m_{i+1}/m_i \quad \text{and} \quad \hat{q}_i = n_{i+1}/n_i, \quad i = 1,\ldots,k-1.$$

Those solutions obtained on the boundary of the parameter set are also described by these formulas, where

$$\hat{p}_i(\hat{q}_i) = 0 \quad \text{if} \quad m_{i+1}(n_{i+1}) = 0$$

and

$$\hat{p}_i(\hat{q}_i) = 1 \quad \text{if} \quad m_{i+1} = m_i(n_{i+1}=n_i).$$

Thus

$$\max L(\underline{m},\underline{n}|\underline{p},\underline{q}) = L(\underline{m},\underline{n}|\hat{\underline{p}},\hat{\underline{q}})$$

$$= \Sigma_{\substack{i=1 \\ \hat{p}_i, \hat{q}_i \neq 0,1}}^{k-1} \{m_{i+1} \ln \hat{p}_i + (m_i - m_{i+1}) \ln(1-\hat{p}_i)$$

$$+ n_{i+1} \ln \hat{q}_i + (n_i - n_{i+1}) \ln(1-\hat{q}_i)\}. \quad (4.1)$$

If $(F, G) \in \mathcal{L}_C$ then we may write $q_i = \bar{G}(\tau_{i+1})/\bar{G}(\tau_i) = p_i^{\theta}$, $i = 1, \ldots, k-1$, to obtain the constrained logarithmic likelihood function

$$\ln L(\underline{m},\underline{n}|\underline{p},\theta) = \Sigma_{i=1}^{k-1} (m_{i+1} + \theta n_{i+1}) \ln p_i + (m_i - m_{i+1}) \ln(1-p_i)$$

$$+ (n_i - n_{i+1}) \ln(1-p_i^{\theta}).$$

Provided that for all i, m_i and n_i are both not zero, also $(m_i - m_{i+1})$ and $(n_i - n_{i+1})$ are both not zero, then $0 < p_i < 1$. Furthermore, if there are at least two intervals in which there are observations from both distributions, then a necessary condition for $(\hat{\underline{p}}, \hat{\theta})$ to be optimal is for $\nabla \ln L = \underline{0}$.

$$\left.\frac{\partial \ln L}{\partial p_i}\right|_{\hat{\underline{p}},\hat{\theta}} = (m_{i+1} + \hat{\theta} n_{i+1})/\hat{p}_i - (m_i - m_{i+1})/(1-\hat{p}_i)$$

$$- \hat{\theta}(n_i - n_{i+1})\hat{p}_i^{\hat{\theta}-1}/1-\hat{p}_i^{\hat{\theta}} = 0, \quad i = 1, \ldots, k-1$$

$$\left.\frac{\partial \ln L}{\partial \theta}\right|_{\hat{\underline{p}},\hat{\theta}} = \Sigma_{i=1}^{k-1} \{n_{i+1} - (n_i - n_{i+1})\hat{p}_i^{\hat{\theta}}/(1-\hat{p}_i^{\hat{\theta}})\} \ln \hat{p}_i = 0. \quad (4.2)$$

Since $\hat{\underline{p}} \neq \underline{0}$ then $(\hat{\underline{p}}, \hat{\theta})$ is a solution to these equations iff $(\hat{\underline{p}}, \hat{\theta})$ is a solution to

$$g_i(\underline{p}, \theta) = m_{i+1} + \theta n_{i+1} - (m_i - m_{i+1})p_i/(1-p_i)$$

$$- \theta(n_i - n_{i+1})p_i^{\theta}/(1-p_i^{\theta}) = 0, \quad i = 1, \ldots, k-1,$$

$$g_k(\underline{p}, \theta) = \Sigma_{i=1}^{k-1}\{n_{i+1} - (n_i - n_{i+1})p_i^{\theta}/(1-p_i^{\theta})\} \ln p_i = 0. \qquad (4.3)$$

To solve Equations (4.3) one could for given ${}^{n}\theta$ sequentially solve each of the first $k-1$ equations for a unique $\hat{p}_i$ because

$$\left.\frac{\partial g_i}{\partial p_i}\right|_{\hat{p}_i, \hat{\theta}} = -(m_i - m_{i+1})/(1-\hat{p}_i)^2 - \hat{\theta}^2(n_i - n_{i+1})\hat{p}_i^{\hat{\theta}-1}/(1-\hat{p}_i^{\hat{\theta}})^2 < 0.$$

Then considering the p's fixed, one could solve $g_k(\hat{\underline{p}}, \theta) = 0$ for a unique ${}^{n+1}\hat{\theta}$ since

$$\left.\frac{\partial g_k}{\partial \theta}\right|_{\hat{\underline{p}}, \hat{\theta}} = -\Sigma_{i=1}^{k-1}\{(n_i - n_{i+1})(\ln \hat{p}_i)^2 \hat{p}_i^{\hat{\theta}}/(1-\hat{p}_i^{\hat{\theta}})^2\} < 0.$$

Although a solution will be obtained, the algorithm is very slow to converge.

A version of Newton's method is proposed although the sufficiency conditions for convergence have not been established for Equation (4.2). Since we have available a good initial estimate of θ, namely $\hat{\theta} = \hat{\theta}_{MI}$, we may solve the first $k-1$ equations of (4.2) for $\hat{p}_i$, $i = 1, \ldots, k-1$, to obtain the vector ${}^{1}(\hat{\underline{p}}, \hat{\theta})$. We may then use ${}^{1}(\hat{\underline{p}}, \hat{\theta})$ to start Newton's method in which

$$ {}^{n+1}(\hat{\underline{p}}, \hat{\theta})^t = (\hat{\underline{p}}, \hat{\theta})^t - J({}^{n}(\hat{\underline{p}}, \hat{\theta}))^{-1} \underline{f}({}^{n}(\hat{\underline{p}}, \hat{\theta})) $$

where J is the Hessian matrix, and

$$ \underline{f}(\underline{p}, \theta) = \nabla \ln L . $$

J is of the form

$$ \left[\begin{array}{ccc|c} a_1 & & & b_1 \\ & \ddots & & \vdots \\ & & a_{k-1} & b_{k-1} \\ \hline b_1 & \cdots & b_{k-1} & c \end{array} \right] $$

where

$$ a_i = \frac{\partial^2 \ln L}{\partial p_i} $$

$$ b_i = \frac{\partial^2 \ln L}{\partial p_i \partial \theta} $$

$$ c = \frac{\partial^2 \ln L}{\partial \theta^2} $$

and J^{-1} equals

$$ \frac{1}{S} \left[\begin{array}{c|c} \left(\frac{b_i b_j}{a_i a_j} + \frac{S \delta_{ij}}{a_i} \right) & \left(- \frac{b_i}{a_i} \right) \\ \hline \left(- \frac{b_j}{a_j} \right) & 1 \end{array} \right] $$

where

$$S = c - \Sigma_{j=1}^{k-1} \frac{b_j^2}{a_j}$$

so that,

$$^{n+1}\hat{p}_i = {}^n\hat{p}_i - \frac{1}{S}\frac{b_i}{a_i}\{\Sigma_{j=1}^{k-1} \frac{b_j}{a_j} f_j(^n(\underline{\hat{p}}, \hat{\theta})) - f_k(^n(\underline{\hat{p}}, \hat{\theta}))\} - \frac{1}{a_i} f_i(^n(\underline{\hat{p}}, \hat{\theta}))$$

$$^{n+1}\hat{\theta} = {}^n\hat{\theta} + \frac{1}{S}\{\Sigma_{j=1}^{k-1} \frac{b_j}{a_j} f_j(^n(\underline{\hat{p}}, \hat{\theta})) - f_k(^n(\underline{\hat{p}}, \hat{\theta}))\}.$$

We have the constraints $0 < p_i < 1$ and $\theta > 0$ on the maximization problem. It is possible to change variables by the logit-transformation for p_i, that is $s_i = \ln\{p_i/(1-p_i)\}$ or $p_i = e^{s_i}/(1+e^{s_i})$, and by using $w = \ln\theta$ to obtain an unconstrained maximization problem. Thus

$$^{n+1}s_i = {}^n s_i - \frac{1}{S}\frac{b_i}{a_i}\{\Sigma_{i=1}^{k-1} \frac{b_j}{a_j} f_j(^n(\underline{s}, w)) - f_k(^n(\underline{s}, w))\} - \frac{1}{a_i} f_i(^n(\underline{s}, w))$$

$$^{n+1}w = {}^n w + \frac{1}{S}\{\Sigma_{j=1}^{k-1} \frac{b_j}{a_j} f_j(^n(\underline{s}, w)) - f_k(^n(\underline{s}, w))\}$$

where, if we write $e^{s_i}/(1+e^{s_i})$ as p_i and e^w as θ

$$f_i = [m_{i+1} - (m_i - m_{i+1})e^{s_i} + \theta\{n_{i+1} - (n_i - n_{i+1})p_i^{\theta}\}]/(1-p_i^{\theta}), \quad i = 1, \ldots, k-1,$$

$$f_k = \Sigma_{i=1}^{k-1} \theta \ln p_i\{n_{i+1} - (n_i - n_{i+1})p_i^{\theta}/(1-p_i^{\theta})\},$$

$$a_i = \frac{\partial^2 \ln L}{\partial s_i^2} = -\{m_i e^{s_i} + \theta[n_{i+1} e^{s_i} - (n_i - n_{i+1})\{p_i^\theta/(1-p_i^\theta)\}$$

$$\times \{e^{s_i} - \theta/(1-p_i^\theta)\}]\}/(1+e^{s_i})^2,$$

$$b_i = \frac{\partial^2 \ln L}{\partial w \partial s_i} = \theta[n_{i+1} - (n_i - n_{i+1}) p_i^\theta/(1-p_i^\theta)\{1 + \theta \ln p_i/(1-p_i^\theta)\}]/(1+e^{s_i}),$$

$$c = \frac{\partial^2 \ln L}{\partial w^2} = \Sigma_{i=1}^{k-1} \theta \ln p_i [n_{i+1} - (n_i - n_{i+1}) p_i^\theta/(1-p_i^\theta)\{1 + \theta \ln p_i/(1-p_i^\theta)\}].$$

When the solution vector $(\hat{\underline{s}}, \hat{w})$ is substituted into Equations (4.2) we obtain

$$\ln L(\underline{m}, \underline{n} | \hat{\underline{p}}, \hat{\theta}) = \max_{(\underline{p}, \theta)} \ln L(\underline{m}, \underline{n} | \underline{p}, \theta),$$

so that we can define the likelihood ratio statistic

$$LR = -2(\ln L(\underline{m}, \underline{n} | \hat{\underline{p}}, \hat{\underline{q}}) - \ln L(\underline{m}, \underline{n} | \hat{\underline{p}}, \hat{\theta}))$$

for testing the assumption of simple proportionality. It is reasonable to expect that when $(F, G) \in \mathcal{L}_C$ LR will be distributed asymptotically as a chi-square distribution with $k-2$ degrees of freedom.

4.4. Cox's Likelihood Ratio Test

Cox (1971) gives, in a preliminary draft of a paper, a general model for studying continuous distribution functions whose failure rate functions are related by

$$r(t|Z) = r_0(t)e^{Z\underline{\beta}},$$

where $r_0(t)$ is a failure rate function under standard conditions, $\underline{\beta}$ is a $p \times 1$ vector of unknown parameters, Z is an $n \times p$ matrix where the jth row $(\underline{z}_j)$ contains the realization of the variables associated with the jth individual. For two sample problem, $\underline{z}_j$ consists of only one variable equalling 0 or 1 so that,

$$r(t|0) = r_0(t)$$

$$r(t|1) = r_0(t)e^{\beta}.$$

For this example, the indicator variable z_j in Cox's notation equals $1-\delta_j$ in our notation of Chapter 2.

If the time dependent component $\psi(t)$, being a known function of t, is contained in z, that is,

$$\underline{z}_j = (0,0) \quad \text{if jth individual is from the standard distribution,}$$
$$= (1, \psi(t)) \quad \text{if jth individual from related distribution,}$$

then

$$r(t|(0,0)) = r_0(t)$$

$$r(t|(1,\psi(t))) = r_0(t)e^{\beta_0+\beta_1\psi(t)} \tag{4.3}$$

Cox suggests using this model for testing the consistency of the simple model of proportionality. He suggests for the case $\psi(t) = t$ to reparameterize (4.3) to $r_0(t)e^{\{\gamma+\beta_1(t-t^*)\}}$, (4.4) where t^* is an arbitrary constant somewhere near the overall mean. This parameterization is to avoid the more extreme non-orthogonalities of fitting.

We use Cox's argument to develop a test for proportionality of the failure rate functions. Let $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ be random samples from distribution functions F and G having respective failure rate functions $r_0(t)$ and $r_0(t)e^{\underline{z}\,\underline{\beta}}$, and let $t_1, \ldots, t_{m+n}$ be the vector of ordered failure times with associated matrix Z having row vectors

$$\begin{aligned} \underline{z}_j &= (0, 0) \quad \text{if F observation} \\ &= (1, t) \quad \text{if G observation.} \end{aligned}$$

Let $R(t_j)$ denote the risk set, the items on test at time t_j. Since the failure times are distinct, then conditionally on the risk set $R(t_j)$ the probability that individual j is observed is

$$\frac{\exp(\underline{z}_j \underline{\beta})}{\Sigma_{\ell \in R(t_j)} \exp(\underline{z}_\ell \underline{\beta})} .$$

Thus the required conditional log likelihood is

$$L(\beta) = \Sigma_{j=1}^{n} \underline{z}_j \underline{\beta} - \Sigma_{j=1}^{n} \ln \{\Sigma_{\ell \in R(t_j)} \exp(\underline{z}_\ell \underline{\beta})\} . \qquad (4.5)$$

If we use the parameterization in (4.4) then

$$L(\gamma, \beta_1) = \Sigma_{j=1}^{n} (1-\delta_j)\{\gamma+\beta_1(t_j - t^*)\} - \ln[m_j + n_j \exp\{\gamma+\beta_1(t_j - t^*)\}],$$

where $\delta_j = 1(0)$ if F(G) observation, $N = m+n$, $m_j = \Sigma_{i=j}^{N} \delta_i$ and $n_j = \Sigma_{i=j}^{N}(1-\delta_i)$. Let $d_j = m_j + n_j \exp\{\gamma+\beta_1(t_j - t^*)\}$. By differentiation we obtain

$$\frac{\partial L}{\partial \gamma} = \Sigma_{j=1}^{N} (1-\delta_j) - n_j \exp\{\gamma+\beta_1(t_j - t^*)\} / d_j$$

$$\frac{\partial L}{\partial \beta_1} = \Sigma_{j=1}^{N} (1-\delta_j)(t_j - t^*) - n_j (t_j - t^*) \exp\{\gamma+\beta_1(t_j - t^*)\} / d_j$$

$$\frac{\partial^2 L}{\partial \gamma^2} = - \Sigma_{j=1}^{N} n_j m_j \exp\{\gamma+\beta_1(t_j - t^*)\} / d_j^2$$

$$\frac{\partial^2 L}{\partial \beta_1^2} = - \Sigma_{j=1}^{N} n_j m_j (t_j - t^*)^2 \exp\{\gamma+\beta_1(t_j - t^*)\} / d_j^2$$

$$\frac{\partial^2 L}{\partial \beta_1 \partial \gamma} = - \Sigma_{j=1}^{N} n_j m_j (t_j - t^*) \exp\{\gamma+\beta_1(t_j - t^*)\} / d_j^2 .$$

Lemma 4. For the log likelihood function $L(\gamma, \beta_1)$, there is a unique solution vector $(\hat{\gamma}, \hat{\beta}_1)$ provided $\Sigma_{i=1}^{m} \delta_i \neq m$ and $\Sigma_{i=1}^{n} (1-\delta_i) \neq n$.

Proof: The Hessian matrix $H(\gamma, \beta_1)$ associated with $L(\gamma, \beta_1)$ is negative semi definite since for $\forall\ \underline{q}$

$$\underline{q}^t H(\gamma, \beta_1)\underline{q} = -\Sigma_{j=1}^N \{q_1^2 + 2q_1q_2(t_j - t^*) + q_2^2(t_j - t^*)^2\} \times n_j m_j \exp\{\gamma + \beta_1(t_j - t^*)\} / d_j^2 \leq 0.$$

Thus $L(\gamma, \beta_1)$ is a concave function, moreover by the same argument as in the proof of Theorem 2.3 $L(\gamma, \beta_1)$ has limiting value $-\infty$ as γ or β_1 tend to $+\infty$ or $-\infty$. A necessary and sufficient condition for $(\hat{\gamma}, \hat{\beta})$ to maximize $L(\gamma, \beta)$ is for $\nabla L(\hat{\gamma}, \hat{\beta}) = \underline{0}$. (Newton's method is suitable for finding the solution vector.)

To test the adequacy of the assumption of proportional failure rate functions one may consider the hypothesis test $H_0: \beta_1 = 0$. Rejection of H_0 would indicate that the relationship between the failure rate functions is a time dependent. The proposed test statistic is

$$C_t = -2(\ln L(\hat{\gamma}, \hat{\beta}) - \ln L(\hat{\hat{\gamma}}, 0)).$$

Cox assumes that asymptotically C_t will have a chi-square distribution with 1 degree of freedom. This assumption is checked in a Monte Carlo study. A statistic related to C_t is also studied, namely, $C_{t'}$ where by substituting $\psi(t) = \ln t$ in (4.3) we obtain

$$r(t | (0, 0)) = r_0(t)$$

and

$$r(t | (1, \psi(t))) = r_0(t) t^{\beta_1} e^{\beta_0}.$$

This relationship between the failure rate functions is of the Weibull form where the two distributions differ in both shape and scale parameters.

4.5. Paired Observation Test

As an application of property (vi) of Chapter 1, Thomas (1969) proposes a conditional test of exact size for $H_0: G = 1 - (1-F)^{\theta}$, for some θ, $0 < \theta < \infty$ for the case of equal sample size. Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be random samples from F and G respectively. After pairing arbitrarily, we obtain Bernolli random variables $\delta_1, \ldots, \delta_n$ where $\delta_i = 1(0)$ if the minimum of the pair is an $x(y)$. Then under H_0 $D = \Sigma_{i=1}^{n} \delta_i$ has a binomial distribution where D is a sufficient statistic. Suppose the minimum of the pairs are divided into two groups of size n_1 and $n-n_1$ and Z is the sum of the δ's in the first group. Then the conditional distribution of Z for given D is hypergeometric with

$$P(Z=z|H_0,D) = \frac{\binom{D}{z}\binom{n-D}{n_1-z}}{\binom{n}{D}}$$

$$z = 0, \ldots, \min(D, n_1), \quad n_1 < n, \quad D < n.$$

Randomization is necessary to construct a test of exact size for given (D, n_1, n). For example, let $n = 20$, $n_1 = 10$ and let $\alpha = .10$.

From the tables of Owen we need to solve for γ, $.05 = P_{z-1} + \gamma p_z$ where $P_{z-1} = \Sigma_{i=0}^{z-1} P(z=i)$, $z \geq 1$, and $p_z = P(Z=z)$. The following table gives for each D and z the randomization probability γ corresponding to p_z in the lower critical region. By symmetry, γ is the randomization probability for corresponding value of z' in the upper critical region. For any trial for which $m = n = 20$, $2 \leq D \leq 18$, the rejection rule for a size $\alpha = .10$ test is:

reject H_0, if $Z < z$ or if $Z > z'$;

reject H_0 with probability γ, if $Z = z$ or if $Z = z'$;

retain H_0, otherwise.

Table XXI. Randomization probabilities associated with z and z′ for the paired observation test, n = 20.

D	z	z′	γ
2	0	2	.250000
3	0	3	.475000
4	1	3	.184375
5	1	4	.357143
6	1	5	.760710
7	2	5	.322222
8	2	6	.653333
9	3	6	.310000
10	3	7	.626747
11	4	7	.310000
12	4	8	.653333
13	5	8	.322222
14	5	9	.760710
15	6	9	.357143
16	7	9	.184375
17	7	10	.475000
18	8	10	.250000

5. COMPARISON OF TESTS BASED ON A MONTE CARLO STUDY

5.1. Description of Program Used for Empirical Study

The system of programs used to compare the test statistics of the tests for proportional failure rate functions was similar to the system described in Section 3.1. The main program, the ordering subroutine, and the maximal invariant subroutine were retained, while two subroutines for calculating the test statistics replaced the remaining subroutines. Both of the likelihood ratio statistics derived from Cox's paper were calculated in the same subroutine, the first stage of which, used $\hat{\theta}_{MI}$ to calculate $L(\hat{\hat{\gamma}}, 0)$, and then used $\hat{\theta}_{MI}$ to start the iterative procedure to find $L(\hat{\gamma}, \hat{\beta}_1)$ for both the statistics C_t and $C_{t'}$.

Some difficulties were encountered in the subroutines for the LR statistic. Only three intervals were used for the procedure, so that to try to satisfy the necessary condition that at least two intervals contain observations from both distributions, the following scheme was used. The first (t_1) and last (t_N) observation as well as the combined sample median $(t_{m'})$ were found. Division points were placed at $(t_1+2t_{m'})/3$ and $(2t_{m'}+t_N)/3$ and a check of the counts was made; if two of the three cells did not have observations from both distributions, the first mark was moved to $t_{m'}$. If the necessary condition was still not satisfied, the sample was rejected. Also, if the iterative scheme for this subroutine did not converge the sample was rejected. A count of the total number of such rejections was recorded. A count of the number of simulations in which there was no overlap of the samples was also recorded.

Our first objective in the empirical study was to ascertain if the null distributions of the statistics were approximately chi-square for moderate sample size. Our second objective was to estimate the power of tests for different values of the parameter and for different sample size.

Simulated samples were drawn from Weibull distributions $F(t) = 1 - e^{-t^{c_1}}$ and $G(t) = 1 - e^{-\theta t^{c_2}}$. Sample sizes of 10 and 20 constituted the majority of the study, however, some samples of size 40 were used for the case $c_1 = 2$, $c_2 = 1$. Most of the combinations of c_1 and c_2 were repeated for both $\theta = 1$ and $\theta = 3$.

5.2. Empirical Size of the Test Statistics

The following combinations of $c_1 = c_2$ and θ were considered: $(\theta = 1,\ c_1 = c_2 = 1)$, $(\theta = 3,\ c_1 = c_2 = 1)$ and $(\theta = 3,\ c_1 = c_2 = 3)$. Each combination was used for both sample sizes of 10 and 20. Table XXII contains a summary of the sample distributions based on 200 simulations. Rejects of type 0 and L correspond to the rejections for no overlap and likelihood ratio criterion, respectively.

Table XXII. Empirical size of three test statistics for H_0: $G = 1-(1-F)^{\theta}$, $\theta \epsilon (0,\infty)$, based on 200 simulations from Weibull distributions $F(t) = 1-e^{-t^c}$ and $G(t) = 1-e^{-\theta t^c}$.

			$\theta = 1$				$\theta = 3$			
Asym Size	c	m=n	C_t	$C_{t'}$	LR	Rejects 0 L	C_t	$C_{t'}$	LR	Rejects 0 L
.01	1	10	.035	.035	.040	1, 1	.01	.005	.005	4, 0
		20	.010	.005	.005	0, 0	.015	.010	.015	0, 0
	3	10					.010	.010	.035	13, 4
		20					.005	.010	.010	0, 1
.05	1	10	.080	.095	.100	1, 1	.065	.070	.075	4, 0
		20	.035	.025	.035	0, 0	.035	.040	.035	0, 0
	3	10					.065	.070	.100	13, 4
		20					.055	.050	.050	0, 1
.10	1	10	.145	.165	.165	1, 1	.130	.125	.125	4, 0
		20	.110	.065	.095	0, 0	.065	.090	.065	0, 0
	3	10					.120	.130	.155	13, 4
		20					.110	.100	.125	0, 1
.15	1	10	.205	.210	.215	1, 1	.160	.185	.220	4, 0
		20	.170	.125	.155	0, 0	.130	.120	.110	0, 0
	3	10					.165	.160	.210	13, 4
		20					.155	.150	.220	0, 1
.20	1	10	.270	.260	.270	1, 1	.220	.270	.285	4, 0
		20	.245	.165	.205	0, 0	.180	.170	.150	0, 0
	3	10					.215	.235	.250	13, 4
		20					.190	.185	.305	0, 1
.25	1	10	.310	.320	.305	1, 1	.260	.270	.215	4, 0
		20	.245	.210	.275	0, 0	.210	.215	.185	0, 0
	3	10					.285	.295	.325	13, 4
		20					.245	.220	.350	0, 1

From the empirical size of the statistics given in Table XXII, it is fair to conclude that the null distribution of the test statistics is close to the asymptotic distribution and relatively insensitive to moderate changes in $c_1 = c_2$ and θ. The best approximation is given by $C_{t'}$ and the worst by LR.

5.3. Empirical Power of the Test Statistics

Our objective in the empirical study was to compare the power of the test statistics when the sample size is increased, when the ratio of c_1/c_2 is increased, and when θ is changed from 1 to 3. A summary of the empirical power is given in Table XXII and Table XXIV. The following conclusions are appropriate.

1. Even for samples as small as 10, the statistics C_t and $C_{t'}$ have appreciable power when $c_1/c_2 = 2$.
2. $C_{t'}$ is the best statistic in terms of both empirical power and empirical size. This result is not surprising since $C_{t'}$ is constructed to have a failure rate function of Weibull form. If other distributions are used for simulation, then $C_{t'}$ may not be superior to C_t.
3. There is a considerable drop in power when θ is changed from 1 to 3.

It is worthwhile to look at two other tests, the paired comparison test and Thoman's test (1969). Thoman gives a scale invariant test

Table XXIII. Empirical power of three test statistics for $H_0: G = 1-(1-F)^\theta$, based on 200 simulations from Weibull distributions, $F(t) = 1 - e^{-t^{c_1}}$ and $G(t) = 1 - e^{-\theta t^{c_2}}$.

				θ = 1				θ = 3			
							Rejects				Rejects
Asym Size	c_1	c_1	m=n	C_t	$C_{t'}$	LR	0 L	C_t	$C_{t'}$	LR	0 L
			10	.180	.185	.105	3, 0	.040	.060	.100	36, 18
.01	2	1	20	.350	.435	.300	0, 0	.125	.220	.115	7, 4
			40	.775	.805	.575	0, 0	.455	.620	.330	1, 1
			10	.405	.440	.280	3, 0	.100	.190	.220	36, 18
.05	2	1	20	.585	.690	.495	0, 0	.300	.435	.325	7, 4
			40	.885	.905	.750	0, 0	.660	.820	.570	1, 1
			10	.535	.585	.430	3, 0	.180	.300	.275	36, 18
.10	2	1	20	.715	.780	.610	0, 0	.415	.555	.440	7, 4
			40	.935	.960	.840	0, 0	.785	.875	.675	1, 1
			10	.620	.680	.515	3, 0	.260	.330	.305	36, 18
.15	2	1	20	.785	.820	.680	0, 0	.485	.625	.490	7, 4
			40	.950	.965	.876	0, 0	.845	.900	.715	1, 1
			10	.675	.730	.590	3, 0	.295	.420	.335	36, 18
.20	2	1	20	.815	.870	.755	0, 0	.545	.700	.545	7, 4
			40	.970	.985	.895	0, 0	.855	.920	.760	1, 1
			10	.705	.750	.615	3, 0	.345	.475	.355	36, 18
.25	2	1	20	.840	.885	.770	0, 0	.605	.710	.590	7, 4
			40	.975	.985	.920	0, 0	.880	.945	.785	1, 1

Table XXIV. Empirical power of three test statistics $H_0: G = 1-(1-F)^{\theta}$, based on 200 simulations from Weibull distributions, $F(t) = 1-e^{-t^{c_1}}$ and $G(t) = 1-e^{-\theta t^{c_2}}$.

				θ = 1				θ = 3			
							Rejects				Rejects
Asym Size	c_1	c_2	m=n	C_t	$C_{t'}$	LR	0 L	C_t	$C_{t'}$	LR	0 L
.01	3	1	20	.72	.88	.635	8, 0	.420	.520	.370	15, 8
.05	3	1	20	.975	.975	.850	8, 0	.640	.750	.590	15, 8
.10	3	1	20	.990	.995	.900	8, 0	.760	.805	.650	15, 8
.15	3	1	20	1.00	.995	.935	8, 0	.790	.880	.710	15, 8
.20	3	1	20	1.00	1.00	.965	8, 0	.830	.910	.735	15, 8
.25	3	1	20	1.00	1.00	.975	8, 0	.860	.910	.770	15, 8

Table XXV. Empirical power for Thoman's test. Empirical power for the paired comparison test based on 200 samples from Weibull distributions. Size of each test is .10.

c_1	c_2	m=n	Thoman Power	Paired comparison Power (θ=1)	Paired comparison Power (θ=3)
2	1	10	.55	.	
2	1	20	.85	.435	.230
3	1	20	.98	.700	.235

for the equality of the shape parameters in two Weibull distributions with unknown scale parameters. We compare the estimated power of both tests in Table XXV. It is clear that the paired comparison test has smaller power than the other tests considered. An interesting point is that $C_{t'}$ is competitive with Thoman's test when $\theta = 1$, although $C_{t'}$ is quite inferior when $\theta = 3$.

6. SUMMARY

This study is directed towards comparing various inference procedures associated with the class $\mathcal{L}_C = \{(F, G): G = 1 - (1-F)^{\theta}, \theta > 0,$ where F and G are continuous$\}$, which is equivalent to the class of pairs of continuous distributions having proportional failure rate (hazard) functions. The dissertation divides naturally into three parts, estimation of θ, estimation of the reliability function $F(x) = 1 - F(x)$ using samples from both $F(x)$ and $G(x)$, and a comparison of statistics for testing the adequacy of the simple proportionality model. Consideration is restricted to the method of maximum likelihood estimation, which becomes a constrained optimization problem when the failure rates are assumed to be either increasing (IFR) or decreasing (DFR).

To clarify the meaning of nonparametric maximum likelihood estimation, we modify Kiefer and Wolfowitz's generalized definition of MLE for the two sample situation. This generalized definition specifies the form of the MLE's w.r.t. particular classes of pairs of distributions, and consequently allows a meaningful two-sample likelihood function to be expressed for the "combined-sample" case, the IFR case, and the DFR case. Algorithms are given to find the values of parameters which maximize these two sample likelihood functions, and consequently MLE's of θ and of $\bar{F}(t)$ (or $\bar{G}(t)$) are obtained.

Two other MLE's of θ are available, neither of which are obtained from a two-sample likelihood function; $\hat{\theta}_{MI}$ is a function of the rank order statistics, and $\hat{\theta}_U$ is a function of the single-sample empirical distributions corresponding to F and G.

Weibull distributions having failure rate functions with different degrees of "increasingness" were chosen for the Monte Carlo study. The most significant result from the Monte Carlo study, is that although the distributions had IFR, $\hat{\theta}_{MI}$ is superior to $\hat{\theta}_{IFR}$ by having smaller bias and smaller variance. In fact, $\hat{\theta}_{MI}$ is the best of the four estimators and $\hat{\theta}_{IFR}$ is the worst. An explanation for the poor performance of $\hat{\theta}_{IFR}$ is in Section 3.4. Although $\hat{\theta}_{MI}$ and $\hat{\theta}_C$ are derived from the same assumptions, it is to be expected that $\hat{\theta}_{MI}$ is the superior, since $m+n$ nuisance parameters $(F(x_i)$ and $G(y_j))$ need to be estimated for $\hat{\theta}_C$.

Another important result coming from the Monte Carlo study, is that the combined sample estimators of $\bar{F}(\cdot)$ generally have smaller MSE than the single-sample empirical estimator of the F-sample, with $\hat{\bar{F}}_{IFR}(\cdot)$ being the occasional exception. $\hat{\bar{F}}_{CIFR}(\cdot)$, which is the IFR-MLE of $\bar{F}(\cdot)$ for given $\hat{\theta}_{MI}$ (and consequently called a conditional IFR-MLE), is the best estimator of $\bar{F}(\cdot)$ on the basis of MSE. Because of the IFR nature of the simulated distributions, the superiority of $\hat{\bar{F}}_{CIFR}$ over $\hat{\bar{F}}_C$ is as expected.

The superiority of $\hat{\bar{F}}_{CIFR}$ over $\hat{\bar{F}}_{IFR}$ in the Monte Carlo study suggests that it may be advisable, in practical applications, to estimate the proportionality constant first, and then use any failure rate assumptions to estimate either one of the reliability functions. There does not seem to be a theoretical justification for using the conditional IFR procedure in preference to the true IFR procedure. If in general, the given conditional method of estimating $\overline{F}(\cdot)$ gives better results for both the IFR and DFR cases, then perhaps a theoretical justification can be given by means of an invariance argument.

Testing the adequacy of the simple proportionality model was the core of the third part of this study. Cox constructs a class of test statistics based on a method of conditional likelihood ratio. Two such test statistics which were used in a Monte Carlo study, had empirical size close to the asymptotic size and appreciable power for even small $(m = n = 10)$ sample sizes. Their power, however, was affected by large deviations of θ from 1.

BIBLIOGRAPHY

Allen, W.R. (1963). A note on conditional probabilities when hazards are proportional. Opns. Res. 11, 658-659.

Barlow, R.E. and Proschan, F. (1965). Mathematical Theory of Reliability, John Wiley & Sons, Inc., New York.

Barlow, R.E. (1968). Likelihood ratio tests for restricted families of probability distributions. Ann. Math. Statist., 39, 547-560.

Barlow, R.E. and van Zwet, W.R. (1969). Comparison of several nonparametric estimators of the failure rate function. Operations Research Center Report ORC 69-10, University of California, Berkely.

Bray, T.A., Crawford, G.B., and Proschan, F. (1967). Maximum likelihood estimation of a U-shaped failure rate function. Boeing Scientific Research Lab., Doc. D1-82-0660, Seattle, Washington.

Brunk, H.D. (1958). On the estimation of parameters restricted by inequalities. Ann. Math. Statist. 29, 437-454.

Church, J.D. and Harris, B. (1970). The estimation of reliability from stress-strength relationships. Techn. 12, 49-54.

Cox, D.R. (1971). Regression models and life tables. A draft of a paper to be submitted for publication.

Dubey, S. (1965a). Characterization theorems of the Weibull and Weibull-Gamma distributions. Ann. of Math Statist., 36, abstract no. 7, 1321.

Dubey, S. (1965b). Characterization theorems for extreme value and logistic distributions. Annals of Math. Statist., 36, abstract no. 4, 1324.

Efron, B. (1967). The two sample problem with censored data. Proceedings of the Fifth Berkeley Symposium 4, 831-853.

Ferguson, Thomas S. (1967). Mathematical Statistics: A decision theoretic approach, Academic Press, New York and London.

Ghosh, B. K. (1970). Sequential tests of statistical hypotheses, Addison-Wesley Publishing Co., Massachusetts.

Govindarajulu, Z. (1968). Distribution-free confidence bounds for Pr(X < Y). Ann. Inst. Stat. Math. (Tokyo), Vol. 20, 229-238.

Grenander, Ulf. (1956). On the theory of mortality measurement I and II, Skand. Akt. 39, 90-96 and 125-153.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc., 53, 457-480.

Kiefer, J. and J. Wolfowitz. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Ann. Math. Statist., 27, 887-906.

Lehmann, E. L. (1953). The power of rank tests. Ann. Math. Statist., 24, 23-43.

Lehmann, E. L. (1959). Testing statistical hypotheses, John Wiley & Sons, Inc., New York.

Macon, Nathaniel. (1963). Numerical Analysis, John Wiley & Sons, Inc., New York.

Marshall, A. W. and Proschan, F. (1965). Maximum likelihood estimation for distributions with monotone failure rate. Ann. Math. Statist., 36, 69-77.

Nádas, Arthur. (1970). On proportional hazard functions. Techn. 12, 413-416.

Owen, D. B. (1962). Handbook of statistical tables, Addison-Wesley Co., Inc., Massachusetts.

Shorack, R. A. (1967). On the power of precedence life tests. Techn. 9, 154-158.

Thoman, D. R. and Bain, L. J. (1969). Two sample tests in the Weibull distribution. Techn. 11, 805-815.

Thomas, David R. (1969). A useful distributional property for paired observations from Lehmann alternatives. An unpublished note.

van der Laan, P. (1970). Simple distribution-free confidence intervals for a difference in location. Philips Res. Repts Suppl. 1970 No. 5.

Wilk, M.B. and Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. Biometrika, 55, 1-17.

Zangwill, Willard I. (1969). Nonlinear programming: A unified approach, Prentice-Hall, Inc., Englewood Cliffs, N.J.