AN ABSTRACT OF THE THESIS OF

Mohamed Gamal Hassan Habib for the degree of Doctor of Philosophy

in Statistics presented on May 1, 1981

Title: A CHI-SQUARE GOODNESS-OF-FIT TEST FOR CENSORED DATA

Abstract approved [signature]
David R. Thomas

Generalizations of the classical chi-square goodness-of-fit statistic

$$Q = \sum_{i=1}^{r+1} \frac{(O_i - E_i)^2}{E_i}$$

are considered for arbitrarily right-censored data. Assume the hypothesis of interest is specified by the family of distributions

$$\mathcal{F} = \{F_\theta ; \theta \in \Theta\} .$$

The maximum likelihood estimator for θ is used and the asymptotic covariance matrix is modified accordingly to produce a statistic with a limiting chi-square distribution with r degrees of freedom (a generalization of Moore's (1977) result). With the uncorrected covariance, Chernoff and Lehmann's (1954) result is also generalized. These generalizations are also shown to extend to random cell boundaries.

A Chi-square Goodness-of-fit Test for Censored Data

by

Mohamed Gamal Hassan Habib

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed May 1981
Commencement June 1981

APPROVED:

Associate Professor of Statistics
in charge of major

Chairman of Department of Statistics

Dean of Graduate School

Date thesis is presented May 1, 1981

Typed by Muriel Davis for Mohamed Gamal Hassan Habib

ACKNOWLEDGMENT

I am deeply grateful to Dr. David Thomas for his kind supervision, help and encouragement during the preparation of this thesis. I am deeply grateful also to Dr. David Birkes for his continuous encouragement and help. His careful review and suggestions for improving the thesis are especially appreciated. I am also grateful to Dr. Donald Pierce for his helpful suggestions.

My sincere thanks are also due to Professor Lyle Calvin, Chairman, Department of Statistics for his encouragement and guidance during my graduate studies. My special thanks to my friend Adel Bargal for encouragement and helpful discussion. Last, but not least, I wish to express my appreciation to my parents whose love, care, and teachings shape my life.

TABLE OF CONTENTS

LIST OF FIGURES

THE CHI-SQUARE GOODNESS-OF-FIT TEST FOR CENSORED DATA

I. INTRODUCTION

A common problem arising in life testing and clinical trials is that of testing the goodness-of-fit of a distribution model. Such models may either completely specify the distribution (simple hypothesis) or specify a parametric family of distributions (composite hypothesis). Frequently, the experimenter wishes to analyze the data before all the experimental units have failed, resulting in censored observations. The simplest kind of censoring is that of single censoring which occurs when all observations are censored at the same time. The single censoring is likely to occur in industrial applications when the experiment starts with N experimental units at time zero and the censoring occurs either after a prespecified time (Type I censoring) or a prespecified ordered failure (Type II censoring). Due to limited testing facilities or availability of experimental units (as in clinical studies) it may be impossible to put all units on test at the same time. Consequently, censored observations may not all be subject to the same amount of test time (arbitrary censoring).

In this thesis, the problem of testing goodness-of-fit for arbitrarily censored data is considered. Assuming a random censorship model we use the maximum likelihood estimators for the nuisance parameters in the case of a composite hypothesis and generalize a result of Moore (1977). Also, Chernoff and Lehmann's (1954) result is generalized for the random censorship model. A study of the more practical case of allowing the cell boundaries to depend on the estimates is considered with a conclusion of negligible asymptotic effect on the above results.

For the random censorship model, the random sample of failure times $X_1, \ldots, X_N$ is censored by a corresponding random sample of censoring times $U_1, \ldots, U_N$. That is, for $j=1, \ldots, N$ we observe only $Y_j = \min(X_j, U_j)$ and $\delta_j = I_{[X_j \leq U_j]}$. If we let $F_\theta(x)$ and $H(u)$ denote the cumulative distribution functions of X and U, respectively, and $\hat{\overline{F}}_N = (\hat{\overline{F}}_N(a_1), \ldots, \hat{\overline{F}}_N(a_r))'$ be the product-limit estimator for $\overline{F}_\theta = 1 - F_\theta$ at the partition points $0 < a_1 < \ldots < a_r < \infty$ of the real line, then the vector $\hat{T}_N = \sqrt{N}(\hat{\overline{F}}_N - \overline{F}_{\hat{\theta}_N})$ is shown to converge in distribution to $\hat{T} \sim N(0, \hat{\Sigma})$, where $\hat{\theta}_N$ is the maximum likelihood estimator for the parameter vector $\theta = (\theta_1, \ldots, \theta_k)'$. Then $\hat{\hat{Q}}_N(\hat{\theta}_N) = \hat{T}'_N \hat{\hat{\Sigma}}^{-1}_N \hat{T}_N$ is proposed as a test of fit and its limiting distribution is shown to be a chi-square distribution with r degrees of freedom. The quadratic form $Q_N(\hat{\theta}_N) = \hat{T}'_N \hat{V}^{-1} \hat{T}_N$ with $\hat{V}$ obtained from V (the asymptotic covariance of $T_N = \sqrt{N}(\hat{\overline{F}}_N - \overline{F}_\theta))$ upon substituting $\hat{\theta}_N$ for θ without making correction for the estimation is shown to have a limiting distribution which is bounded by the two chi-square distributions with degrees of freedom $r-k$ and r. Allowing $a_1, \ldots, a_r$ to depend on the data is shown to have asymptotically negligible effect on the distribution of both $\hat{Q}_N(\hat{\theta}_N)$ and $Q_N(\hat{\theta}_N)$.

In Chapter II, a brief review of the classical chi-square test for the complete sample case and the product-limit estimator for the random censorship model are given. Also a discussion of the chi-square test when using the modified minimum chi-square estimator for θ is presented. Chen (1975) used the modified minimum chi-square estimator $\tilde{\theta}_N$ and showed that $Q_N(\tilde{\theta}_N)$ has a limiting chi-square distribution with $r-k$ degrees of freedom.

In Chapter III, the chi-square goodness-of-fit statistic $\hat{Q}_N(\hat{\theta}_N)$ is developed and illustrated for the exponential distribution using a sample of survival times from aortic value replacement patients. In Chapter IV, the goodness-of-fit statistic $Q_N(\hat{\theta}_N)$ is developed. The asymptotic effect of using the critical value from a chi-square distribution with $r-k$ degrees of freedom is investigated for the cases of exponential and Weibull distributions assuming a uniform censoring distribution. In Chapter V the asymptotic effect of using random cell boundaries is studied and the results of Chapters III and IV are found to remain valid.

II. PREVIOUS RESEARCH

II.1. Distribution Structure

Let $X_1, \ldots, X_N$ denote the failure times for N individuals. Assume the X_i's are independent and identically distributed with survival function $\overline{F}_\theta(x) = P\{X \geq x\}$, which is absolutely continuous in x with corresponding distribution function $F_\theta(x) = 1 - \overline{F}_\theta(x)$ and density $f_\theta(x) = F'_\theta(x)$. The parameter vector θ is assumed to belong to an open set Θ of k-dimensional Euclidean space $\mathbb{R}^k$. Further, assume that the X_i's are censored from the right by random variables U_i's which are independent and identically distributed with distribution function $H(u)$ and continuous density $h(u)$. Also, the U_i's and the X_i's are assumed to be independent of each other. One observes then only

$$\left.\begin{aligned} Y_i &= \min(X_i, U_i), \quad \text{and} \\ \delta_i &= I_{[X_i \leq U_i]} \end{aligned}\right\} \; i=1, \ldots, N \tag{2.1}$$

where δ_i indicates whether Y_i is a censoring time $(\delta_i = 0)$ or failure time $(\delta_i = 1)$. So, $Y_1, \ldots, Y_N$ constitute a set of independent and identically distributed random variables with a survival function $\overline{G}_\theta(y) = P\{Y \geq y\}$ given by

$$\overline{G}_\theta(y) = \overline{F}_\theta(y)\overline{H}(y) \; . \tag{2.2}$$

The sub-distribution functions of an uncensored observation,

$\widetilde{G}_\theta(y) = P\{Y \leq y, \ \delta = 1\}$, and of a censored observation
$G^*_\theta(y) = P\{Y \leq y, \ \delta = 0\}$, may be written as

$$\widetilde{G}_\theta(y) = \int_0^y \overline{H}(z) d\, F_\theta(z) \ , \qquad (2.3)$$

$$G^*_\theta(y) = \int_0^y \overline{F}_\theta(z)\, d\, H(z) \qquad (2.4)$$

respectively. The joint distribution of Y and δ has the density function

$$g_\theta(y, \delta) = [\, f_\theta(y)\, \overline{H}(y)]^{\delta} [\, \overline{F}_\theta(y) \ h(y)]^{1-\delta}$$

with respect to the product of Lebesgue measure on $(0, \infty)$ and counting measure on $\{0, 1\}$. Hence, the likelihood function for a sample of size N is

$$L = \prod_{i=1}^{N} [\, f_\theta(y_i)\, \overline{H}(y_i)]^{\delta_i} [\, \overline{F}_\theta(y_i)\, h(y_i)]^{1-\delta_i} . \qquad (2.5)$$

II. 2. The Classical Chi-Square Goodness-of-Fit Test for Uncensored Data

We consider the problem of testing the hypothesis

$$H_0 : X \sim \overline{F}_\theta(x) \qquad (2.6)$$

where $\overline{F}_\theta(x) = P\{X \geq x\}$ is some member of a specified family of survival distributions $\{\overline{F}_\theta(x) : \theta \in \Theta\}$ and $\theta = (\theta_1, \ldots, \theta_k)' \in \Theta \subset R^k$. For the uncensored data case, failure times for all sample units are observed, and hence the classical chi-square goodness-of-fit test, which is a comparison of the observed empirical survival function, $\overline{F}_N^e(x)$, to the hypothesized $\overline{F}_\theta(x)$, can be applied to test the hypothesis (2.6). Let $0 = a_0 < a_1 < \ldots < a_{r+1} = \infty$ be partition points on the real line. Let $F_N^e(x)$ be the empirical distribution function based on the N observations $X_1, \ldots, X_N$; that is, $F_N^e(x)$ is the proportion of observations which are less than or equal to x. The empirical survival function is $\overline{F}_N^e(x) = 1 - F_N^e(x)$. The chi-square measure of discrepancy or "distance" between the two survival functions $\overline{F}_N(x)$ and $\overline{F}_\theta(x)$ is

$$Q_N(\theta) = T_N' V^{-1} T_N \tag{2.7}$$

where $T_N = \sqrt{N}(\overline{F}_N^e - \overline{F}_\theta)$ with $\overline{F}_N^e = (\overline{F}_N^e(a_1), \ldots, \overline{F}_N^e(a_r))'$, $\overline{F}_\theta = (\overline{F}_\theta(a_1), \ldots, \overline{F}_\theta(a_r))'$, and $V = \mathrm{cov}(T_N)$. Using the facts that $NF_N^e(a_i) = O_1 + \ldots + O_i$, $NF_\theta(a_i) = E_1 + \ldots + E_i$, and the $i, j^{\underline{th}}$ element of the covariance matrix V can be written $V_{ij} = F_\theta(a_i)\overline{F}_\theta(a_j); i \leq j$ where O_i and E_i are the observed and expected numbers in the i^{th} interval $(a_{i-1}, a_i]$ we can write $Q_N(\theta)$ in the form

$$Q_N(\theta) = \sum_{i=1}^{r+1} \frac{(O_i - E_i)^2}{E_i}$$

which is the familiar Pearson chi-square statistic (if θ is known). Since no such simplification exists in the case of censored observations we will consider only the form (2.7) for $Q_N(\theta)$.

Karl Pearson (1900) established that, under the hypothesis (2.6), $Q_N(\theta)$ has a limiting chi-square distribution with r degrees of freedom. Hence if the null hypothesis is simple (θ known), the statistic $Q_N(\theta)$ can be used as a test which rejects the null hypothesis if $Q_N(\theta)$ exceeds the $100(1-\alpha)$ percentage point of a chi-square random variable with r degrees of freedom. In the case when the null hypothesis (2.6) is composite, θ may be replaced in (2.7) by an estimator $\tilde{\theta}_N$, such as the modified minimum chi-square estmator or the asymptotically equivalent maximum likelihood estimator based on the observed cell frequencies. The resulting statistic $Q_N(\tilde{\theta}_N)$, was shown (Cramér, 1966, p. 426) to have a limiting chi-square distribution with $r-k$ degrees of freedom. However, this limiting distribution does not hold for the more convenient maximum likelihood estimator $\hat{\theta}_N$ based on the original (ungrouped) data. In fact Chernoff and Lehmann (1954) proved that the asymptotic null distribution of $Q_N(\hat{\theta}_N)$ is that of

$$\sum_{i=1}^{k} \lambda_i Z_i^2 + \sum_{i=k+1}^{r} Z_i^2$$

where $Z_1, \ldots, Z_N$ are independent standard normal random variables and the λ_i's, which may depend on the parameters, lie between 0 and 1. Roy (1956) and Watson (1958) proved the Chernoff and Lehmann result when the cell boundaries are allowed to vary with the estimators of the parameters while fixing the class probabilities. Watson conjectured that "if the class intervals are allowed to vary [appropriately] with the estimators of the parameters, it is possible that the resulting asymptotic distribution may become independent of the unknown parameters." Moore (1971) extended the Roy and Watson result to the case of multivariate observations. Moore also gave a method for numerical computation of the asymptotic distribution and provided a table of critical values for testing the goodness-of-fit to the univariate normal family. Nikulin (1973) considered scale-location families of distributions and modified the statistics $Q_N(\hat{\theta}_N)$ to produce goodness-of-fit test statistics which have a limiting chi-square distribution with r degrees of freedom. Moore (1977) gave a more general procedure of constructing a test statistic which has a limiting chi-square distribution.

II. 3. The Product-Limit Estimator for Censored Data

The product limit estimator, $\hat{F}_N(t)$, for the distribution function, $F_\theta(t)$, was derived by Kaplan and Meier (1958) as the distribution which maximizes the likelihood of the observations. If we recall the observations described in (2.1) and order them so that $Y_1 < \ldots < Y_N$ (note that the probability of having any ties is zero under the assumption of continuity of F_θ and H), then the product limit estimator, $\hat{F}_N(t)$, of the failure distribution function, $F_\theta(t)$, is defined by

$$1-\hat{F}_N(t) = \begin{cases} 1 & t \leq Y_1 \\ \prod_{i=1}^{j-1} [(N-i)/(N-i+1)]^{\delta_i} & t \in (Y_{j-1}, Y_j] \\ 0 & t > Y_N \end{cases} \qquad (2.8)$$

The product-limit estimator, $\hat{H}_N(t)$, for the censoring distribution, $H(t)$, can be defined in the same way with $1-\delta_i$ instead of δ_i.

In their famous 1958 paper, Kaplan and Meier proved the consistency of $\hat{F}_N(t)$ as an estimator for $F_\theta(t)$. Breslow and Crowley (1974) have shown under some conditions that the random vector $\sqrt{N}(\hat{F}_N - F_\theta)$ has a limiting normal distribution with zero mean and a covariance matrix

$V = \{V_{ij}\}$, where, for $1 \leq i \leq j \leq r$,

$$V_{ij} = V_{ji} = \overline{F}_\theta(a_i)\overline{F}_\theta(a_j) \int_0^{a_i} [\overline{H}(z)\overline{F}_\theta^2(z)]^{-1} f_\theta(z)dz \qquad (2.9)$$

with the vectors $\hat{F}_N = (\hat{F}_N(a_1), \ldots, \hat{F}_N(a_r))$ and $F_\theta = (F_\theta(a_1), \ldots, F_\theta(a_r))$ evaluated at the partition points $0 < a_1 < \ldots < a_r < \infty$.

II.4. A Chi-Square Goodness-of-Fit Test Under Random Censoring

Corresponding to the partition points $0 < a_1 < \ldots < a_r < \infty$, we have from the previous section that the r-dimensional random vector

$$T_N = \sqrt{N}(\hat{\overline{F}}_N - \overline{F}_\theta) , \tag{2.10}$$

which is evaluated at $a_1, \ldots, a_r$, converges in law to a random variable $T \sim N_r(0, V)$ with V given in (2.9). Hence the quadratic form $Q_N(\theta)$ defined by

$$Q_N(\theta) = T_N' V^{-1} T_N \tag{2.11}$$

has a limiting chi-square distribution with r degrees of freedom. (To show this, one needs only to note that quadratic forms are continuous functionals and a continuous functional in T_N converges in law to the same functional in T (Rao, 1973, p. 124)).

Note that, in our theoretical development, we will assume that the censoring distribution, $H(t)$, is known. However, in practice we will use the product-limit estimator, $\hat{H}_N(t)$. Asymptotically, this can be expected to have negligible effect on our tests, simply because $\hat{H}_N(t)$ is a consistent estimator of $H(t)$, as shown by Kaplan and Meier (1958). This will allow us to prove that the quadratic form $Q_N(\theta)$ is asymptotically equivalent, for each fixed θ, to the quadratic form obtained by substituting $\hat{H}_N(t)$ for $H(t)$.

Lemma 2.1 (Chen, 1975, Lemma 2.7). Let $\{A_N(y)\}$ and $\{B_N(y)\}$ be sequences of monotone random functions such that $\{B_N(y)\}$ is uniformly bounded, $A_N(y) \xrightarrow{P} A(y)$ and $B_N(y) \xrightarrow{P} B(y)$ where $A(y)$ and $B(y)$ are bounded and continuous. Then, for arbitrary $0 \leq a < b < \infty$,

$$\int_a^b A_N(y)\,dB_N(y) \xrightarrow{P} \int_a^b A(y)\,dB(y) .$$

Setting $A_N(z) = [\hat{\bar{H}}_N(z)\bar{F}_\theta^2(z)]^{-1}$, $A(z) = [\bar{H}(z)\bar{F}_\theta^2(z)]^{-1}$, and $B_N(z) = B(z) = F_\theta(z)$ in Lemma 2.1, it follows that

$$\hat{D}_i = \int_{a_{i-1}}^{a_i} [\hat{\bar{H}}_N(z)\,\bar{F}_\theta^2(z)]^{-1} dF_\theta(z)$$

$$\xrightarrow{P} \int_{a_{i-1}}^{a_i} [\bar{H}(z)\,\bar{F}_\theta^2(z)]^{-1} dF_\theta(z) = D_i \tag{2.12}$$

Lemma 2.2 (Chen, 1975, Section 3.4). If $Q_N(\theta)$ is defined as in (2.11), then

$$Q_N(\theta) = N \sum_{i=1}^{r} \frac{\left(\frac{\hat{\bar{F}}_N(a_{i-1})}{\bar{F}_\theta(a_{i-1})} - \frac{\hat{\bar{F}}_N(a_i)}{\bar{F}_\theta(a_i)} \right)^2}{D_i} \tag{2.13}$$

Using (2.12) and (2.13) one can show that

$$Q_N(\theta)\Big|_{H=\hat{H}_N} = N \sum_{i=1}^{r} \frac{\left(\frac{\hat{\bar{F}}_N(a_{i-1})}{\bar{F}_\theta(a_{i-1})} - \frac{\hat{\bar{F}}_N(a_i)}{\bar{F}_\theta(a_i)} \right)^2}{\hat{D}_i} \tag{2.14}$$

has a limiting chi-square distribution with r degrees of freedom. Hence if θ in the null hypothesis (2.6) is known (i.e., a simple hypothesis), then we can use the statistic specified by (2.14) as our test statistic.

In practical applications, a completely specified distribution is rarely assumed. More generally the hypothesis (2.6) specifies a family of distributions and the unknown parameter vector, $\theta = (\theta_1, \ldots, \theta_k)'$, needs to be estimated. In his Ph. D. thesis, Chen (1975) used the modified minimum chi-square method to estimate θ, by $\tilde{\theta}_N$ say, and substituted $\tilde{\theta}_N$ for θ in (2.14). In the limiting distribution of this statistic, one degree of freedom is lost for each parameter estimated from the sample, thus generalizing a result of Fisher (1922, 1924). Note that in a situation like this one needs to have the number of partition points greater than the number of parameters to be estimated in the null hypothesis (2.6), i.e., $r > k$.

In the following chapters we are going to use the maximum likelihood estimator, $\hat{\theta}_N$, for θ in constructing test statistics. In Chapter III, a test statistic $\hat{Q}_N(\hat{\theta}_N)$ is proposed and shown to have an asymptotic chi-square distribution. In Chapter IV, the effect of using the maximum likelihood estimator in the quadratic form (2.11) is examined. And Chapter V discusses the effect of allowing the cell boundaries to vary with the estimator while leaving the estimated cell probabilities fixed.

III. THE CHI-SQUARE GOODNESS-OF-FIT STATISTIC $\hat{Q}_N(\hat{\theta}_N)$

In the previous chapter the modified minimum chi-square estimator, or the asymptotically equivalent maximum likelihood estimator based on the observed cell frequencies, were used in constructing a test statistic which has an asymptotic chi-square distribution with $r-k$ degrees of freedom under the null hypothesis (2.6) for the case of k unknown parameters $\theta_1, \ldots, \theta_k$ and r partition points $0 < a_1 < \ldots < a_r < \infty$. This chapter will be devoted to the use of the maximum likelihood estimator, $\hat{\theta}_N$, based on the original, ungrouped, observations in constructing a statistic which has an asymptotic chi-square distribution with r degrees of freedom regardless of the number of parameters estimated.

The first section is concerned with the asymptotic distribution of the vector $\hat{T}_N = \sqrt{N}(\hat{\bar{F}}_N - \bar{F}_{\hat{\theta}_N})$. The limiting normal distribution of $\hat{T}_N$ is given in Theorem 3.1. The second section is concerned with the construction of the statistic $\hat{Q}_N(\hat{\theta}_N)$. The limiting chi-square distribution of $\hat{Q}_N(\hat{\theta}_N)$ is given in Lemma 3.5. And the last section deals with the application of $\hat{Q}_N(\hat{\theta}_N)$ to testing the goodness-of-fit of two sets of data to the exponential distribution.

Although the approach is different, the result in Lemma 3.5 can be considered as a generalization of Moore's (1977) results under the random censorship assumption.

III. 1 Asymptotic Theory

Let $Y_1, \ldots, Y_N$ and $\delta_1, \ldots, \delta_N$ be defined as in (2. 1). Let $\hat{\theta}_N$ be the maximum likelihood estimator for the parameter vector $\theta = (\theta_1, \ldots, \theta_k)'$ in the family of distributions specified by the hypothesis (2. 6). Corresponding to the partition points $a_1 < \ldots < a_r$, let $\hat{T}_N$ be the random vector obtained from (2. 10) by substituting $\hat{\theta}_N$ for θ, i. e.,

$$\hat{T}_N = \sqrt{N}(\overline{F}_N - \overline{F}_{\hat{\theta}_N}) . \tag{3. 1}$$

In Theorem (3. 1) we will show, under certain regularity conditions, that $\hat{T}_N$ has a limiting normal distribution.

Condition R. Assume that the distributions of the random variables (Y, δ) satisfy the following conditions

(i) Θ is an open set in R^k (k-dimensional Euclidian space),

(ii) $G_\theta(y, \delta)$ is continuous in θ, and $\dfrac{\partial g_\theta(y, \delta)}{\partial \theta_i}$ and $\dfrac{\partial^2 g_\theta(y, \delta)}{\partial \theta_i \partial \theta_j}$

exist for almost all (y, δ) and are continuous in θ; $j=1, \ldots, k$,

(iii) $\displaystyle\iint \frac{\partial}{\partial \theta_i} g_\theta(y, \delta)\, dy\, d\nu(\delta) = 0$; $i=1, \ldots, k$

where $\nu(\delta)$ is the counting measure on the set $\{0, 1\}$,

(iv) $J_{ij}(\theta) = E\left[\frac{\partial \ln g_\theta(Y,\delta)}{\partial\theta_i}\ \frac{\partial \ln g_\theta(Y,\delta)}{\partial\theta_j}\right]$

$= E\left[-\frac{\partial^2 \ln g_\theta(Y,\delta)}{\partial\theta_i\partial\theta_j}\right]$ is finite for all $i,j=1,\ldots,k$

and the information matrix $J(\theta) = \{J_{ij}(\theta)\}$ is positive definite and continuous in θ.

(v) The maximum likelihood estimator $\hat{\theta}_N$ exists and is unique.

In the above condition recall that the density $g_\theta(y,\delta)$ is of the form

$$g_\theta(y,\delta) = [f_\theta(y)\overline{H}(y)]^\delta [\overline{F}_\theta(y)h(y)]^{1-\delta}$$

So, the conditions (ii), (iii) and (iv) can be thought of in the following equivalent forms

(ii)' $F_\theta(y)$ is continuous in θ, and $\frac{\partial f_\theta(y)}{\partial\theta_i}$, $\frac{\partial^2 f_\theta(y)}{\partial\theta_i\partial\theta_j}$, $\frac{\partial F_\theta(y)}{\partial\theta_j}$ and $\frac{\partial^2 F_\theta(y)}{\partial\theta_i\partial\theta_j}$ exist for almost all y and are continuous in θ ; $i,j=1,\ldots,k$,

(iii)' $\int_0^\infty h(y)\frac{\partial \overline{F}_\theta(y)}{\partial\theta_i}dy + \int_0^\infty \overline{H}(y)\frac{\partial f_\theta(y)}{\partial\theta_i}dy = 0$; $i=1,\ldots,k$

(iv)' $J_{ij}(\theta) = E\left[\delta\frac{\partial \ln f_\theta(Y)}{\partial\theta_i}\ \frac{\partial \ln f_\theta(y)}{\partial\theta_j} + (1-\delta)\frac{\partial \ln \overline{F}_\theta(Y)}{\partial\theta_i}\ \frac{\partial \ln \overline{F}_\theta(Y)}{\partial\theta_j}\right]$

$= -E\left[\delta\frac{\partial^2 \ln f_\theta(Y)}{\partial\theta_i\partial\theta_j} + (1-\delta)\frac{\partial^2 \ln \overline{F}_\theta(Y)}{\partial\theta_i\partial\theta_j}\right]$ is finite for all

$i, j=1, \ldots, k$ and the information matrix $J=J(\theta)=\{J_{ij}(\theta)\}$ is positive definite and continuous in θ.

Theorem 3.1. Under the condition R the random vector $\hat{T}_N$ in (3.1) has a limiting normal distribution with zero mean and covariance matrix

$$\Sigma = V - BJ^{-1}B' \qquad (3.2)$$

i.e.,

$$\hat{T}_N \xrightarrow{L} \hat{T} \sim N(0, \Sigma)$$

where V is defined in (2.9), J is the information matrix defined in (iv) above, and B is the gradient matrix whose (i, j) component is

$$B_{ij} = \frac{\partial F_\theta(a_i)}{\partial \theta_j} \qquad i=1, \ldots, r; \quad j=1, \ldots, k \qquad (3.3)$$

Proof: Let us order the observations so that $Y_1 < \ldots < Y_N$. Define the empirical cumulative hazard function

$$\Lambda_N^e(t) = \Sigma_{Y_i \le t} \delta_i/(N-i+1) \qquad (3.4)$$

and let

$$\Lambda_\theta(t) = -\ln \overline{F}_\theta(t)$$

denote the corresponding population cumulative hazard function. Evaluate $\Lambda_N^e(t)$ and $\Lambda_\theta(t)$ at the partition points $a_1 < \ldots < a_r$ to form the vectors

$$\Lambda_N^e = [\Lambda_N^e(a_1), \ldots, \Lambda_N^e(a_r)]'$$

and

$$\Lambda_\theta = [\Lambda_\theta(a_1), \ldots, \Lambda_\theta(a_r)]' .$$

First, we prove the joint asymptotic normality of the vectors $\sqrt{N}(\hat{\theta}_N - \theta)$ and $\sqrt{N}(\Lambda_N^e - \Lambda_\theta)$.

Denote by

$$G_N^e(t) = N^{-1} \sum_{j=1}^{N} I_{[Y_j < t]} \tag{3.5}$$

$$\tilde{G}_N^e(t) = N^{-1} \sum_{j=1}^{N} I_{[Y_j < t,\ \delta_j = 1]}$$

$$= N^{-1} \sum_{j=1}^{N} \delta_j I_{[Y_j < t]} \tag{3.6}$$

and

$$G_N^{*e}(t) = N^{-1} \sum_{j=1}^{N} (1-\delta_j) I_{[Y_j < t]} \tag{3.7}$$

respectively, the left continuous versions of the empirical distribution function of the observations, the sub-empirical distribution function of the uncensored observations and the sub-empirical distribution function of the censored observations. Then, equation (7.9) in Breslow and Crowley (1974) reads

$$\sqrt{N}(\Lambda_N^e(t) - \Lambda_\theta(t)) = \sqrt{N}\left[\int_0^t \frac{G_N^e(y) - G_\theta(y)}{(1-G_\theta(y))^2}\, d\tilde{G}_\theta(y)\right.$$

$$- \int_0^t \frac{\tilde{G}_N^e(y) - G_\theta(y)}{(1 - G_\theta(y))^2}\, dG_\theta(y)$$

$$\left. + \frac{G_N^e(t) - G_\theta(t)}{(1 - G_\theta(t))^2}\right] + o_p(1)$$

Using (3.5) and (3.6) we can write

$$\sqrt{N}\,(\Lambda_N^e(t) - \Lambda_\theta(t)) = N^{-\frac{1}{2}} \sum_{j=1}^{N} S_{2,t}(Y_j, \delta_j) + o_p(1) \qquad (3.8)$$

where

$$S_{2,t}(Y_j, \delta_j) = \int_0^t \frac{I_{[Y_j < y]} - G_\theta(y)}{(1 - G_\theta(y))^2}\, d\tilde{G}_\theta(y)$$

$$- \int_0^t \frac{\delta_j I_{[Y_j < y)} - \tilde{G}_\theta(y)}{(1 - G_\theta(y))^2}\, dG_\theta(y)$$

$$+ \frac{I_{[Y_j < t]} - G_\theta(t)}{(1 - G_\theta(t))^2}$$

Also, under conditions R-i and R-ii, the maximum likelihood estimator $\hat{\theta}_N$ of θ is a solution of

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta}\Big]_{\hat{\theta}_N} = 0$$

and from condition R-iv and the weak law of large numbers $-\frac{1}{N}\frac{\partial^2 \mathscr{L}(\theta)}{\partial\theta\,\partial\theta'}$ converges in probability to the information matrix $J(\theta)$, where $\mathscr{L}(\theta)$ is the log-likelihood function obtained from (2.5) as

$$\mathscr{L}(\theta) = \sum_{j=1}^{N} [\, \delta_j \{\ln f_\theta(Y_j) + \ln \overline{H}(Y_j)\} + (1-\delta_j)\{\ln \overline{F}_\theta(Y_j) + \ln h(Y_j)\}] \; . \tag{3.9}$$

Then

$$\frac{\partial \mathscr{L}(\theta)}{\partial\theta} = \sum_{j=1}^{N} S_1(Y_j, \delta_j) \tag{3.10}$$

where

$$S_1(Y_j, \delta_j) = \delta_j \frac{\partial \ln f_\theta(Y_j)}{\partial\theta} + (1-\delta_j)\frac{\partial \ln \overline{F}_\theta(Y_j)}{\partial\theta} \; .$$

Expand $\frac{\partial \mathscr{L}(\theta)}{\partial\theta}$ around $\theta = \hat{\theta}_N$ and use (3.10) to write (see Wilks, 1963, p. 361)

$$\sqrt{N}(\hat{\theta}_N - \theta)\; N^{-\frac{1}{2}} = J^{-1}(\theta) \sum_{j=1}^{N} S_1(Y_j, \delta_j) + o_p(1) \tag{3.11}$$

Summarize (3.8) and (3.11) as

$$\begin{bmatrix} \sqrt{N}(\hat{\theta}_N - \theta) \\ \sqrt{N}(\Lambda_N^e - \Lambda_\theta) \end{bmatrix} = N^{-\frac{1}{2}} \sum_{j\;1}^{N} \begin{bmatrix} J^{-1}(\theta)\, S_1(Y_j, \delta_j) \\ S_2(Y_j, \delta_j) \end{bmatrix} + o_p(1) \tag{3.12}$$

where $S_2(Y_j, \delta_j) = [S_{2,a_1}(Y_j, \delta_j), \ldots, S_{2,a_r}(Y_j, \delta_j)]'$. From (3.12) and the multivariate central limit theorem it follows that $\sqrt{N}(\hat{\theta}_N - \theta)$ and $\sqrt{N}(\Lambda_N^e - \Lambda_\theta)$ jointly have a limiting multivariate normal distribution.

The joint asymptotic normality of the vectors $\sqrt{N}(\hat{\theta}_N-\theta)$ and $T_N = \sqrt{N}(\hat{\bar{F}}_N-\bar{F}_\theta)$ follows from the result (Breslow and Crowley, 1974, Equation 7.12) that

$$T_N(t) = -e^{-\Lambda_\theta(t)} \sqrt{N}\,(\Lambda_N^e(t) - \Lambda_\theta(t)) + o_p(1) \tag{3.13}$$

$$= -e^{-\Lambda_\theta(t)} Z_N(t) + o_p(1) .$$

Therefore,

$$\begin{bmatrix} \sqrt{N}(\hat{\theta}_N-\theta) \\ \\ T_N \end{bmatrix} = h \begin{bmatrix} \sqrt{N}(\hat{\hat{\theta}}_N-\theta) \\ \\ \sqrt{N}(\Lambda_N^e-\Lambda_\theta) \end{bmatrix} + o_p(1) ,$$

where h is a function with continuous first partial derivatives. Also $\sqrt{N}(\hat{\hat{\theta}}_N-\theta)$ has a limiting distribution which is $N(0, J^{-1}(\theta))$. Hence (see Rao, 1973, p. 387) we have

$$\begin{bmatrix} \sqrt{N}(\hat{\theta}_N-\theta) \\ \\ T_N \end{bmatrix} \xrightarrow{L} N\left(O\ ,\ \begin{bmatrix} J^{-1}(\theta) & | & V_{12} \\ \hline V_{21} & | & V \end{bmatrix} \right) \tag{3.14}$$

Now, in $\hat{T}_N = \sqrt{N}(\hat{\bar{F}}_N-\bar{F}_{\hat{\theta}_N})$ we expand $\bar{F}_{\hat{\theta}_N}$ around θ using the same result from Rao to obtain

$$\hat{T}_N-T_N = B\sqrt{N}(\hat{\theta}_N-\theta) + o_p(1) \tag{3.15}$$

Finally, the proof of Theorem 3.1 is completed by an application of the following lemma:

<u>Lemma 3.2</u> (Pierce 1980) Let $Y_1, Y_2, \ldots$ be a sequence of random variables with distribution dependent on the parameter vector

θ. Let $\hat{\theta}_N$ be an asymptotically normal and efficient sequence of estimators. Let $T_N = T_N(Y_1, \ldots, Y_N; \theta)$ be such that

$$\begin{bmatrix} \sqrt{N}(\hat{\theta}_N - \theta) \\ T_N \end{bmatrix} \xrightarrow{L} N \left(0, \left[\begin{array}{c|c} V_{11} & V_{12} \\ \hline V_{21} & V_{22} \end{array} \right] \right).$$

Assume the dependence of T_N on θ is smooth enough so that for each θ there exists a matrix B such that

$$\hat{T}_N = T_N + B\sqrt{N}(\hat{\theta}_N - \theta) + o_p(1)$$

where $\hat{T}_N = T_N(Y_1, \ldots, Y_N; \hat{\theta}_N)$. Then $\hat{T}_N$ has a limiting normal distribution with zero mean and covariance $\Sigma = V_{22} - BV_{11}B'$, i. e.

$$\hat{T}_N \doteq N(0, \Sigma).$$

III. 2 The Test Statistic

Assume that Σ is invertible (After the proof of Lemma 3. 5 we discuss the case when Σ is not invertible.) Let $\hat{V}_N$ be the matrix obtained from V by substituting $\hat{\theta}_N$ and $\hat{H}_N$ for θ and H respectively in (2. 9). Similarly define $\hat{\Sigma}_N$.

Lemma 3. 3. $\hat{V}_N$ converges in probability to V.

Proof: Let $A_N(z) = [\bar{\hat{H}}_N(z) \bar{F}^2_{\hat{\theta}_N}(z)]^{-1}$ and $B_N(z) = F_{\hat{\theta}_N}(z)$ in Lemma 2. 1.

Lemma 3. 4. Assume that $\int_0^\infty \left| \frac{\partial^2 \ln f_\theta(y)}{\partial\theta_i \partial\theta_j} \right| f_\theta(y)\,dy$ is continuous in θ for $i, j = 1, \ldots, k$. Then $\hat{\Sigma}_N$ converges in probability to Σ.

Proof: Recall that

$$\Sigma = V - BJ^{-1}B'$$

where V and J depend on both H and θ, but B, as defined in (3.3), depends only on θ. Since B is assumed continuous in θ (condition R(ii)') and since

$$\hat{V}_N = V + o_p(1)$$

from Lemma 3.3, then, the lemma will be concluded by showing that $J]_{\hat{H}_N, \hat{\theta}_N}$ converges in probability to J. To show the dependence of J on H and θ explicitly let us write it $J(H, \theta)$.

Now, the $i, j^{\underline{th}}$ element of $J(H, \theta)$ is

$$J_{ij}(H, \theta) = -\int_0^\infty \frac{\partial^2 \ln f_\theta(y)}{\partial\theta_i \partial\theta_j} \overline{H}(y) dF_\theta(y) - \int_0^\infty \frac{\partial^2 \ln \overline{F}_\theta(y)}{\partial\theta_i \partial\theta_j} \overline{F}_\theta(y) dH(y) \tag{3.16}$$

From the triangle inequality we have

$$\begin{aligned} |J_{ij}(\hat{H}_N, \hat{\theta}_N) - J_{ij}(H, \theta)| &\leq |J_{ij}(\hat{H}_N, \hat{\theta}_N) - J_{ij}(H, \hat{\theta}_N)| + |J_{ij}(H, \hat{\theta}_N) - J_{ij}(H, \theta)| \\ &\leq |J_{ij}(\hat{H}_N, \hat{\theta}_N) - J_{ij}(H, \hat{\theta}_N)| + o_p(1) \end{aligned} \tag{3.17}$$

where the last inequality follows from the condition R-iv. Write,

$$W_{ij}^{(1)}(y; \theta) = -\frac{\partial^2 \ln f_\theta(y)}{\partial\theta_i \partial\theta_j} ,$$

and

$$W^{(2)}_{ij}(y;\theta) = -\frac{\partial^2 \ln \bar{F}_\theta(y)}{\partial\theta_i \partial\theta_j} \bar{F}_\theta(y) .$$

Then

$$J_{ij}(H,\theta) = J^{(1)}_{ij}(H,\theta) + J^{(2)}_{ij}(H,\theta) \tag{3.18}$$

where

$$J^{(1)}_{ij}(H,\theta) = \int_0^\infty W^{(1)}_{ij}(y;\theta)\bar{H}(y)dF_\theta(y) ,$$

and

$$J^{(2)}_{ij}(H,\theta) = \int_0^\infty W^{(2)}_{ij}(y;\theta)\, d\, H(y)$$

Now,

$$|J^{(1)}_{ij}(\hat{H}_N,\hat{\theta}_N)-J^{(1)}_{ij}(H,\hat{\theta}_N)| = |\int_0^\infty (\bar{\hat{H}}_N(y)-\bar{H}(y))\, W^{(1)}_{ij}(y;\hat{\theta}_N)d\, F_{\hat{\theta}_N}(y)|$$

$$\leq \sup_{y\in R} |\hat{H}_N(y)-H(y)| \left|\int_0^\infty W^{(1)}_{ij}(y;\hat{\theta}_N)\, dF_{\hat{\theta}_N}(y)\right|$$

$$= o_p(1) \tag{3.19}$$

where the last equality follows because $\sqrt{N}(\hat{H}_N(y)-H(y))$ has a limiting normal distribution and $\int_0^\infty W^{(1)}_{ij}(y,\theta)\, d\, F_\theta(y)$ is bounded for all θ close to the true value. Also since $W^{(2)}_{ij}(y;\theta)$, being continuous as a function of y on the closed interval $[0,T]$, is uniformly continuous, then for a given $\epsilon > 0$ there exists a finite positive number M and a set of partition points $0 = t_0 < t_1 < \ldots < t_M = T$ such that for $t \epsilon I_m = (t_{m-1}, t_m]$ we have

$$|W_{ij}^{(2)}(t,\theta) - W_{ij}^{(2)}(t_m,\theta)| < \frac{\epsilon}{3} \tag{3.20}$$

Let θ_0 denote the true value of θ and

$$W_{ij} = \sup_{|\theta-\theta_0| \leq \epsilon} \{|W_{ij}^{(2)}(t_m,\theta)| ;\ m=0,\ldots,M\}$$

So, since $\hat{\theta}_N$ is a consistent estimator for θ and since $W_{ij}^{(2)}(y;\theta)$ is continuous in θ, we can find N_1 such that for $N > N_1$ we have

$$P\{|W_{ij}^{(2)}(t_m,\hat{\theta}_N)| \leq W_{ij} ;\ m=0,\ldots,M\} > 1 - \frac{\epsilon}{2} \tag{3.21}$$

Now,

$$|J_{ij}^{(2)}(\hat{H}_N,\hat{\theta}_N) - J_{ij}^{(2)}(H,\hat{\theta}_N)| \leq \sum_{m=1}^{M} \left|\int_{I_m} W_{ij}^{(2)}(t,\hat{\theta}_N)d\hat{H}_N(t) - \int_{I_m} W_{ij}^{(2)}(t,\hat{\theta}_N)dH(t)\right|$$

$$= \sum_{m=1}^{M} \left|\int_{I_m} W_{ij}^{(2)}(t,\hat{\theta}_N)d[\hat{H}_N(t)-H(t)]\right|$$

$$\leq \sum_{m=1}^{M} \left|\int_{I_m} [W_{ij}^{(2)}(t,\hat{\theta}_N) - W_{ij}^{(2)}(t_m,\hat{\theta}_N)]\, d[\hat{H}_N(t)-H(t)]\right|$$

$$+ \sum_{m=1}^{M} \left|\int_{I_m} W_{ij}^{(2)}(t_m,\hat{\theta}_N)\, d[\hat{H}_N(t)-H(t)]\right|$$

$$\leq \frac{\epsilon}{3}\int_0^T d[\hat{H}_N(t)+H(t)] + \sum_{m=1}^{M} |W_{ij}^{(2)}(t_m,\hat{\theta}_N)| \left|\int_{I_m} d[\hat{H}_N(t)-H(t)]\right|$$

where the last inequality follows from (3.20). Now, using (3.21) in the above inequality we have for $N > N_1$

$$P\{|J_{ij}^{(2)}(\hat{H}_N, \hat{\theta}_N) - J_{ij}^{(2)}(H, \hat{\theta}_N)| \leq \frac{2\epsilon}{3} + W_{ij} \sum_{m=1}^{M} | \hat{H}_N(t_m) - H(t_m)$$

$$+ \hat{H}_N(t_{m-1}) - H(t_{m-1}) |\} > 1 - \frac{\epsilon}{2}$$

i.e., using the triangle inequality we have for $N > N_1$

$$P\{|J_{ij}^{(2)}(\hat{H}_N, \hat{\theta}_N) - J_{ij}^{(2)}(H, \hat{\theta}_N)| \leq \frac{2\epsilon}{3} + 2W_{ij} \sum_{m=1}^{M} |\hat{H}_N(t_m) - H(t_m)|\} > 1 - \frac{\epsilon}{2} .$$

But, since $\hat{H}_N(t_m)$ is consistent estimator for $H(t_m)$, there exists N_2 such that for $N > N_2$ we have

$$P\{\sum_{m=1}^{M} |\hat{H}_N(t_m) - H(t_m)| < \frac{\epsilon}{6W_{ij}}\} > 1 - \frac{\epsilon}{2} .$$

So, for $N > \max\{N_1, N_2\}$ we have

$$P\{|J_{ij}^{(2)}(\hat{H}_N, \hat{\theta}_N) - J_{ij}^{(2)}(H, \hat{\theta}_N)| < \epsilon\} > 1 - \epsilon .$$

Since ϵ is arbitrary, then

$$|J_{ij}^{(2)}(\hat{H}_N, \hat{\theta}_N) - J_{ij}^{(2)}(H, \hat{\theta}_N)| = o_p(1) \tag{3.22}$$

Using (3.19), (3.22) and (3.18) we have

$$|J_{ij}(\hat{H}_N, \hat{\theta}_N) - J_{ij}(H, \hat{\theta}_N)| = o_p(1)$$

Substitute in (3.17) to conclude that

$$|J_{ij}(\hat{H}_N, \hat{\theta}_N) - J_{ij}(H, \theta)| = o_p(1)$$

<u>Lemma 3.5:</u> Let $\hat{T}_N = \sqrt{N}(\hat{\bar{F}}_N - \bar{F}_{\hat{\theta}_N})$ as defined in (3.1) and assume the condition R holds. Set

$$\hat{Q}_N(\hat{\theta}_N) = \hat{T}'_N \hat{\Sigma}_N^{-1} \hat{T}_N . \tag{3.23}$$

Then, under the hypothesis (2, 6), $\hat{Q}_N(\hat{\theta}_N)$ has a limiting chi-square distribution with r degrees of freedom.

Proof: From Lemma 3.4 we can write

$$\hat{\Sigma}_N^{-1} = \Sigma^{-1} + o_p(1).$$

So,

$$\begin{aligned}\hat{Q}_N(\hat{\theta}_N) &= \hat{T}_N'(\Sigma^{-1} + o_p(1))\hat{T}_N \\ &= \hat{T}_N'\Sigma^{-1}\hat{T}_N + \hat{T}_N' o_p(1)\hat{T}_N \\ &= \hat{T}_N'\Sigma^{-1}\hat{T}_N + o_p(1)\end{aligned}$$

where $\hat{T}_N' o_p(1)\hat{T}_N = o_p(1)$ because of Theorem 3.1. Hence $\hat{Q}_N(\hat{\theta}_N)$ and $\hat{T}_N'\Sigma^{-1}\hat{T}_N$ have the same limiting distribution. The limiting distribution of $\hat{T}_N'\Sigma^{-1}\hat{T}_N$ is that of $\hat{T}'\Sigma^{-1}\hat{T}$ ($\hat{T}$ from Theorem 3.1) which is chi-square with r degrees of freedom.

If $\hat{\Sigma}_N$ in Lemma 3.5 is not invertable use a generalized inverse $\hat{\Sigma}_N^-$ in (3.23) to give a statistic with a limiting chi-square distribution with rank $(\hat{\hat{\Sigma}}_N)$ as its degrees of freedom.

The above lemma suggests that under the random censorship assumption, when the hypothesis is composite and the sample size is large, the statistic (3.23) can be used to perform the goodness-of-fit test. To actually calculate the test statistic, the following may somewhat simplify the task. Recall that (3.2) gives

$$\Sigma = V - BJ^{-1}B'$$

From Rao (1973, p. 33) we can write

$$\Sigma^{-1} = V^{-1} + V^{-1}BC^{-1}B'V^{-1}$$

where

$$C = J - B'V^{-1}B.$$

So, we want to evaluate the quadratic form

$$\hat{Q}_N(\hat{\theta}_N) = \hat{T}'_N \hat{V}_N^{-1}\hat{T}_N + \hat{T}'_N\hat{V}_N^{-1}\hat{B}\hat{C}^{-1}\hat{B}'\hat{V}_N^{-1}\hat{T}_N \tag{3.24}$$

where a hat means θ and H are replaced by $\hat{\theta}_N$ and $\hat{H}_N$ respectively. Set $V^*=AVA'$, $\hat{T}^*_N=A\hat{T}_N$ and $B^*=AB$, where A is given by

$$A = \begin{bmatrix} -\frac{1}{\overline{F}_\theta(a_1)} & 0 & \cdots & 0 \\ \frac{1}{\overline{F}_\theta(a_1)} & -\frac{1}{\overline{F}_\theta(a_2)} & & \\ \vdots & & \ddots & \\ 0 & & \frac{1}{\overline{F}_\theta(a_{r-1})} & -\frac{1}{\overline{F}_\theta(a_r)} \end{bmatrix}$$

Then we have the diagonal matrix

$$V^* = \begin{bmatrix} D_1 & 0 & \cdots & 0 \\ 0 & D_2 & & \\ \vdots & & \ddots & \\ 0 & & & D_r \end{bmatrix}$$

where

$$D_i = \int_{a_{i-1}}^{a_i} [\bar{H}(z)\bar{F}_\theta^2(z)]^{-1} f_\theta(z)\,dz \quad ; \; i=1,\ldots,r, \qquad (3.25)$$

and

$$\hat{T}_N^* = \sqrt{N}\left[\frac{\hat{\bar{F}}_N(a_0)}{\bar{F}_{\hat{\theta}_N}(a_1)} - \frac{\hat{\bar{F}}_N(a_1)}{\bar{F}_{\hat{\theta}_N}(a_1)}, \ldots, \frac{\hat{\bar{F}}_N(a_{r-1})}{\bar{F}_{\hat{\theta}_N}(a_{r-1})} \; \frac{\hat{\bar{F}}_N(a_r)}{\bar{F}_{\hat{\theta}_N}(a_r)}\right].$$

And so,

$$\begin{aligned}\hat{T}'_N V^{-1}\hat{T}_N &= \hat{T}'_N A' A'^{-1} V^{-1} A^{-1} A\hat{T}_N \\ &= \hat{T}_N^* V^{*-1}\hat{T}_N^* \\ &= N\sum_{i=1}^{r}\left(\frac{\hat{\bar{F}}_N(a_{i-1})}{\bar{F}_{\hat{\theta}_N}(a_{i-1})} - \frac{\hat{\bar{F}}_N(a_i)}{\bar{F}_{\hat{\theta}_N}(a_i)}\right)^2 \Big/ D_i . \qquad (3.26)\end{aligned}$$

Similar simplifications for the 2nd term of (3.24) follow from

$$\hat{T}'_N V^{-1} B = \hat{T}_N^{*\prime} V^{*-1} B^* ,$$

and

$$B' V^{-1} B = B^{*\prime} V^{*-1} B^* .$$

In evaluation of the test statistic we replace H and θ in (3.25) and consequently in (3.26) by $\hat{H}_N$, the product-limit estimator, and $\hat{\theta}_N$, the maximum likelihood estimator.

In the following section the statistic specified by (3.23) is used as a test of fit for the exponential distribution using two different sets of data.

III. 3. Examples

In this section, the generalized chi-square goodness-of-fit test for the exponential distribution is illustrated.

Example 1. (Grunkemeier, et al. (1980)). The data used in this example is from 104 patients who survived isolated aortic valve replacement at the University of Oregon Health Science Center between 1965 and 1968. During the 12 years of follow up 35 patients died and the remaining were censored after varying periods of time.

For testing exponential fit $H_o: \bar{F}_\theta(x) = \exp(-x/\theta)$, maximization of the likelihood function (equation (2.5)) yields $\hat{\theta}_N = \sum_{i=1}^{N} y_i / \sum_{i=1}^{N} \delta_i$ which gives for the set of 104 observations $\hat{\theta}_{104} = 25.48$. Using partition points 2, 4, 6, 8, 10, the value of the generalized chi-square statistic (3.23) was 3.66 with 5 degrees of freedom. In Table 1 below, columns 2 and 3 include respectively the product-limit estimator and the fitted exponential distribution evaluated at the partition points given in column 1. In Figure 1-a, the product-limit estimator and the fitted exponential distribution are plotted. The product-limit estimator of the underlying censoring distribution is plotted in Figure 1-b. It is interesting to see that the fitted distribution agrees quite well with the product-limit estimator.

Table 1. Product-limit estimator and the exponential fit for the aortic valve replacement data fixed cell boundaries

Years	Product-limit Estimator	Exponential Fit
0	1.0	1.0
2	0.9044	0.9245
4	0.8455	0.8547
6	0.7959	0.7902
8	0.7449	0.7306
10	0.6591	0.6754

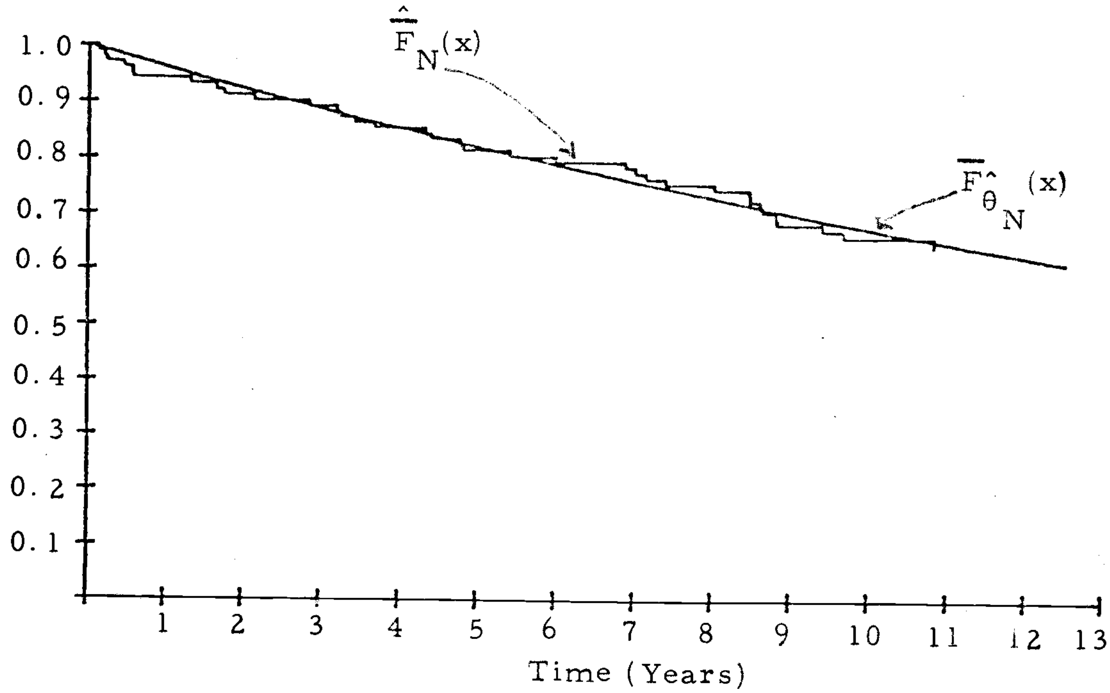

Figure 1-a. Product-limit estimator, $\hat{\bar{F}}_N(x)$, and the exponential fitted distribution, $\bar{F}_{\hat{\theta}_N}(x)$, for the aortic valve replacement data.

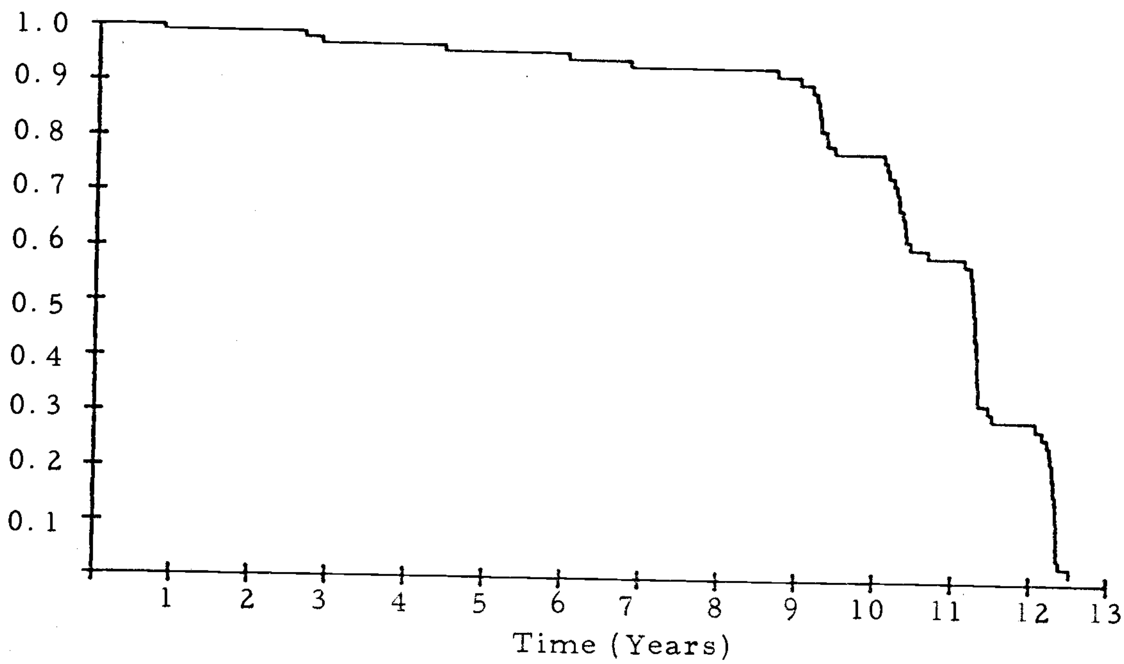

Figure 1-b. Product-limit estimator for the censoring distribution for the aortic valve replacement data.

Example 2. One hundred and four observations were generated from the Weibull survival distribution $\bar{F}(x) = \exp(-(\frac{x}{15})^2)$ with a uniform censoring distribution $\bar{H}(x) = 1-x/26.25$ such that the expected proportion of censored observations is 1/2. The generalized chi-square statistic was used to test the fit of the data to the exponential distribution, i.e., to test $H_o : \bar{F}_\theta(x) = \exp(-x/\theta)$. From the 104 generated observations 51 were failed and 53 were censored at varying times. The maximum likelihood estimator for θ was $\hat{\theta}_{104} = 18.62$ and the corresponding value of the generalized chi-square statistic (3.23) was 41.56 based on 5 degrees of freedom. In Table 2 below, the selected set of partition points used in computing the test statistic is given in column 1 and the corresponding values of the product-limit estimator and the fitted exponential values of the product-limit estimator and the fitted exponential survivals are given in columns 2 and 3 respectively. In Figure 2-a, the product-limit estimator, $\hat{\bar{F}}_N(x)$, and the fitted exponential survival function, $\bar{F}_{\hat{\theta}_N}(x)$, are plotted. It is easy to see that the exponential estimator is under-estimating the early survival function and is over-estimating the late survival function. It is interesting to see how the test statistic reflects the disagreement between the fitted exponential distribution and the product-limit estimator. In Figure 2-b, the product-limit estimator of the underlying censoring distribution is plotted. Note how close the estimator is to being linear.

Table 2. Product-limit estimator and exponential fit for the simulated Weibull data fixed cell boundaries

Years	Product-limit Estimator	Exponential Fit
0	1.0	1.0
2	0.990	0.898
4	0.948	0.807
6	0.813	0.725
8	0.777	0.651
10	0.581	0.585

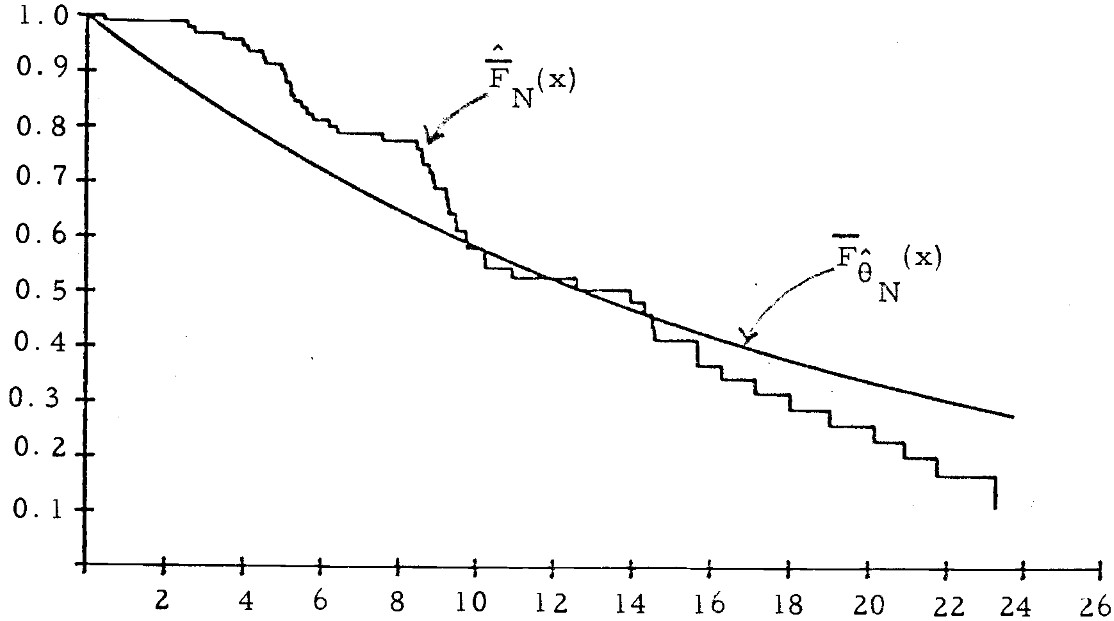

Figure 2-a. Product-limit estimator, $\hat{\bar{F}}_N(x)$, and the exponential fitted distribution, $\bar{F}_{\hat{\theta}_N}(x)$, for the simulated Weibull data.

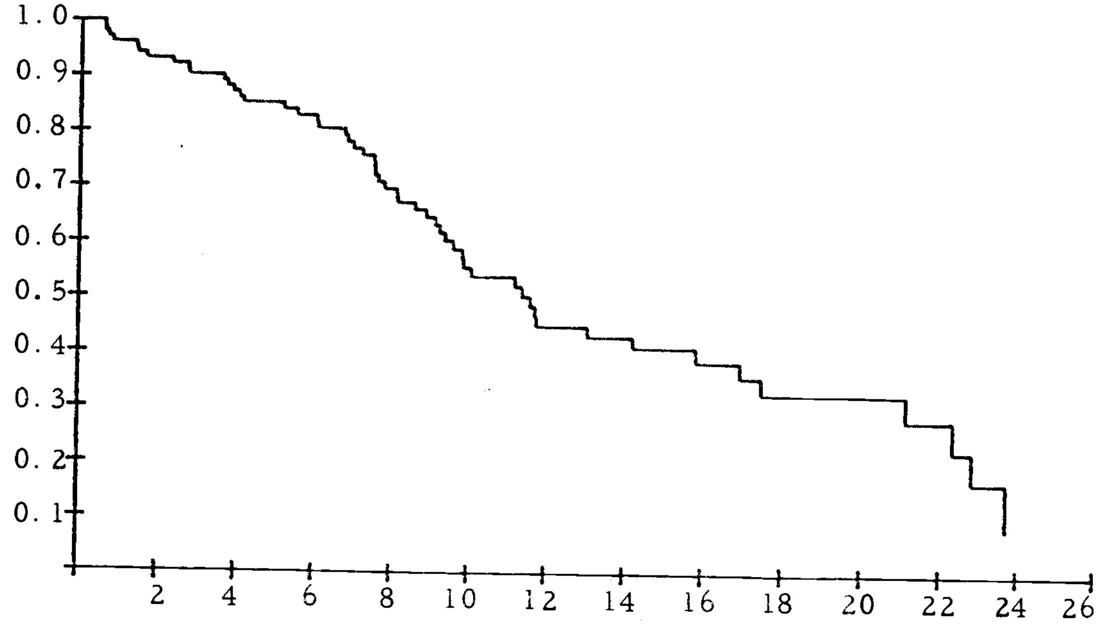

Figure 2-b. Product-limit estimator for the censoring distribution for the simulated Weibull data.

IV. THE GOODNESS-OF-FIT STATISTIC $Q_N(\hat{\theta}_N)$

In the previous chapter we proposed the statistic $\hat{Q}_N(\hat{\theta}_N)$ as a goodness-of-fit test and proved its limiting chi-square distribution.

This chapter will be devoted to examining the limiting distribution of the quadratic form $Q_N(\theta) = \hat{T}'_N V^{-1} \hat{T}_N$ when $\hat{\theta}_N$ is substituted for θ. In Theorem 4.5, $Q_N(\hat{\theta}_N)$ is shown to have a limiting distribution which is bounded by chi-square with $r-k$ degrees of freedom and chi-square with r degrees of freedom. The exact limiting distribution is concluded in Lemma 4.6. In Section 2 a numerical investigation is made of the excess of the asymptotic size of the test when using a critical value from a chi-square distribution with $r-k$ degrees of freedom.

Although the approach is different, the results contained in this chapter (especially Theorem 4.5 and Lemma 4.6) can be considered as generalizations of Chernoff and Lehmann's (1954) results.

IV.1 Limiting Distribution of $Q_N(\hat{\theta}_N)$

Let $Y_1, \ldots, Y_N$ and $\delta_1, \ldots, \delta_N$ be as given by (2.1). Let $\hat{\theta}_N$ be the maximum likelihood estimator for the parameter vector $\theta = (\theta_1, \ldots, \theta_k)'$ in the family of distributions specified by the hypothesis (2.6). Corresponding to the partition points $a_1 < \ldots < a_r$, let

$\hat{T}_N$ be the random vector defined by

$$\hat{T}_N = \sqrt{N}(\hat{\bar{F}}_N - \bar{F}_{\hat{\theta}_N}) \qquad (4.1)$$

where $\hat{F}_N = (\hat{F}_N(a_1), \ldots, \hat{F}_N(a_r))'$ and $F_{\hat{\theta}_N} = (F_{\hat{\theta}_N}(a_1), \ldots, F_{\hat{\theta}_N}(a_r))'$. Let $\hat{V}_N$ be the consistent estimator of V given by Lemma 3.3. Define $Q_N(\hat{\theta}_N)$ by

$$\tilde{Q}_N(\hat{\theta}_N) = \hat{T}'_N \hat{V}_N^{-1} \hat{T}_N. \qquad (4.2)$$

In Theorem 4.5 we will show, under certain regularity conditions, that using the maximum likelihood estimator $\tilde{\theta}_N$ in constructing the quadratic form $Q_N(\hat{\theta}_N)$ will produce a statistic whose asymptotic distribution is not a chi-square but is the distribution of a linear combination of chi-square distributed random variables.

First, we present the following lemmas and corollaries.

Lemma 4.1. Let U be a random vector that has a normal distribution with mean zero and covariance matrix V. Let L be an indempotent matrix such that

$$LV = VL'. \qquad (4.3)$$

Then

$$\tilde{Q} = ||LU||^2_{V^{-1}}$$

has a chi-square distribution with $r(L) = tr(L)$ degrees of freedom where $||a||^2_{V^{-1}} = a'V^{-1}a$ is the norm of a inducted by V^{-1} and $tr(L) = trace(L)$.

Proof: Write

$$\tilde{Q} = U'AU$$

where

$$A = L'V^{-1}L .$$

Then,

$$AVA = (L'V^{-1}L)V(L'V^{-1}L)$$

which, by using (4.3), is

$$\begin{aligned} AVA &= L'V^{-1}VL'L'V^{-1}L \\ &= (L')^3V^{-1}L \\ &= A \end{aligned}$$

The last equality follows from the fact that L is indempotent (i.e. $L^2 = L$). Also

$$\begin{aligned} tr(AV) &= tr(L'V^{-1}LV) \\ &= tr(L'V^{-1}VL') \\ &= tr(L) . \end{aligned}$$

The lemma is concluded by an application of (vii) in page 188 of Rao (1973).

As a result of the above lemma we have the following fact from large-sample theory.

Corollary 4.2. If the random vector U_N converges in distribution to U in Lemma 4.1, then

$$\tilde{Q}_N = ||LU_N||^2_{V^{-1}}$$

has a limiting chi-square distribution with tr(L) as its degrees of freedom.

Lemma 4.3. The nonzero roots of $|BJ^{-1}B'-\mu V| = 0$ and of $|\tilde{J}-\mu J| = 0$ are identical, where $\tilde{J} = B'V^{-1}B$.

Proof: Let $V = AA'$ and $J = LL'$ with A and L non-singular. Then

$$\begin{aligned} 0 &= |BJ^{-1}B'-\mu V| \\ &= |A(A^{-1}BL'^{-1}L^{-1}B'A'^{-1}-\mu I)A'| \\ &= |A(MM'-\mu I)A'| \end{aligned}$$

is equivalent to

$$0 = |MM'-\mu I|$$

where

$$M = A^{-1}BL'^{-1} .$$

Also,

$$\begin{aligned} 0 &= |\tilde{J}-\mu J| \\ &= |L(L^{-1}B'A'^{-1}A^{-1}BL'^{-1}-\mu I)L'| \\ &= |L(M'M-\mu I)L'| \end{aligned}$$

is equivalent to

$$0 = |M'M-\mu I| .$$

To conclude the lemma apply the result in Rao (1973, p. 68).

Lemma 4.4. Let V and W be positive definite matrices such that V-W is non-negative definite. Then $W^{-1}-V^{-1}$ is non-negative definite.

Proof. See Fraser (1957, p. 55).

Theorem 4. 5. Assume that B, defined by (3. 3), has rank k. Then, under the condition R of Chapter III the statistic $Q_N(\hat{\theta}_N)$, defined by (4. 2), has a limiting distribution which is bounded by the two chi-square distributions with degrees of freedom r-k and r respectively.

In other words, the limiting distribution of $Q_N(\hat{\theta}_N)$ is that of

$$\sum_{i=1}^{k} \lambda_i Z_i^2 + \sum_{i=k+1}^{r} Z_i^2 \qquad (4.4)$$

where the Z_i's are independent standard normal random variables and the λ_i's are constants that lie between 0 and 1 and may depend on the parameter $\theta = (\theta_1, \ldots, \theta_k)'$.

Proof: Let $\tilde{\theta}_N$ be the value of θ which minimizes the quadratic form

$$R_N(\theta) = N(\hat{\bar{F}}_N - \bar{F}_\theta)' \hat{V}_N^{-1} (\hat{\bar{F}}_N - \bar{F}_\theta) .$$

That is,

$$\inf_\theta R_N(\theta) = R_N(\tilde{\theta}_N) .$$

Therefore, since $\hat{\theta}_N$ is a valid value of θ, we have

$$R_N(\tilde{\theta}_N) \leq R_N(\hat{\theta}_N) = Q_N(\hat{\theta}_N) . \qquad (4.5)$$

In obtaining $\tilde{\theta}_N$ one may differentiate $R_N(\theta)$ with respect to θ and equate to zero. Thus $\tilde{\theta}_N$ must satisfy the equation

$$\tilde{B}' \hat{V}_N^{-1} \sqrt{N} (\hat{\bar{F}}_N - \bar{F}_{\tilde{\theta}_N}) = 0 \qquad (4.6)$$

where $\tilde{B}$ is obtained from B, defined by (3. 3), by substituting $\tilde{\theta}_N$ for θ. Expansion of $F_{\tilde{\theta}_N}$ around θ (see Rao, 1973, p. 387 ; here we must assume $\tilde{\theta}_N$ is measurable) gives

$$\tilde{T}_N = T_N + B\sqrt{N}(\tilde{\theta}_N - \theta) + o_p(1) \tag{4.7}$$

where $\tilde{T}_N = \sqrt{N}(\hat{\bar{F}}_N - \bar{F}_{\tilde{\theta}_N})$ and $T_N = \sqrt{N}(\hat{\bar{F}}_N - \bar{F}_\theta)$. Pre multiply both sides of (4. 7) by $B'V^{-1}$ to obtain

$$B'V^{-1}\tilde{T}_N = B'V^{-1}T_N + B'V^{-1}B\sqrt{N}(\tilde{\theta}_N - \theta) + o_p(1) \tag{4.8}$$

In his Theorem 3. 1, Chen (1975) expressed $\sqrt{N}(\tilde{\theta}_N - \theta)$ as an approximate linear function of $\sqrt{N}\hat{\bar{F}}_N$ and showed its limiting normal distribution. So, T_N and $\sqrt{N}(\tilde{\theta}_N - \theta)$ have a joint limiting normal distribution. Thus, by (4. 7), $\tilde{T}_N$ has a limiting normal distribution. Therefore, by (4. 6),

$$\begin{aligned} B'V^{-1}\tilde{T}_N &= B'(V^{-1} - \hat{V}_N^{-1})\tilde{T}_N + (B - \tilde{B})'\hat{V}_N^{-1}\tilde{T}_N \\ &= o_p(1) \end{aligned} \tag{4.9}$$

Hence, using (4. 9) in (4. 8), we have

$$\sqrt{N}(\tilde{\theta}_N - \theta) = -(B'V^{-1}B)^{-1}B'V^{-1}T_N + o_p(1).$$

The above equation with (4. 7) gives

$$\tilde{T}_N = T_N - B(B'V^{-1}B)^{-1}B'V^{-1}T_N + o_1(1)$$

Now,

$$R_N(\tilde{\theta}_N) = \tilde{T}'_N \hat{V}_N^{-1} \tilde{T}_N$$

$$= \tilde{T}'_N V^{-1} \tilde{T}_N + o_p(1)$$

$$= ||\tilde{T}_N||^2_{V^{-1}} + o_p(1)$$

$$= ||(I-B(B'V^{-1}B)^{-1}B'V^{-1})T_N||^2_{V^{-1}} + o_p(1)$$

Recall that T_N has a limiting zero mean and a limiting covariance matrix V, so an application of Corollary 4.2 implies that $R_N(\tilde{\theta}_N)$ has a limiting chi-square distribution with the degrees of freedom

$$\text{tr}(I-B(B'V^{-1}B)^{-1}B'V^{-1}) = r-r(B)$$

$$= r-k \ .$$

Now, for the upper bound recall

$$Q_N(\theta) = T'_N V^{-1} T_N,$$

and

$$\hat{Q}_N(\theta) = T'_N \Sigma^{-1} T_N \ ,$$

where

$$\Sigma = V-BJ^{-1}B'.$$

So, $V-\Sigma$ is non-negative definite. An application of Lemma 4.4 yields that $\Sigma^{-1}-V^{-1}$ is non-negative definite. Therefore,

$$\hat{Q}_N(\theta) \geq Q_N(\theta)$$

for all $\theta \epsilon \Theta$. Thus

$$\hat{Q}_N(\hat{\theta}_N) \geq Q_N(\tilde{\theta}_N) \ . \qquad (4.10)$$

The inequalities in (4.5) and (4.10) conclude the result of Theorem 4.5.

Still to be answered is the question of what is the exact asymptotic distribution of $Q_N(\hat{\theta}_N)$. This question will now be answered by determining the values $\lambda_1, \ldots, \lambda_k$ in (4.4).

<u>Lemma 4.6.</u> If $\mu_i = 1-\lambda_i$, $i=1,\ldots,k$, then the μ_i's are the characteristic roots of the determinental equation

$$|\tilde{J} - \mu J| = 0 \tag{4.11}$$

where $\tilde{J} = B'V^{-1}B$ and all μ_i's are between zero and one.

Proof: Let Λ be the diagonal matrix of eigenvalues of V and P be the corresponding orthogonal matrix such that

$$V = P\Lambda P' .$$

Hence

$$Q(\theta) = \hat{T}'V^{-1}\hat{T} = \hat{T}'P\Lambda^{-1}P'\hat{T}$$

i. e.

$$Q(\theta) = \hat{T}'_* \hat{T}_* ,$$

where

$$\hat{T}_* = \Lambda^{-\frac{1}{2}}P'\hat{T}$$

which has a zero mean normal distribution and covariance matrix

$$\Sigma_* = \Lambda^{-\frac{1}{2}} P'\Sigma P\Lambda^{-\frac{1}{2}} .$$

Let P_* be an orthogonal matrix and Λ_* a diagonal matrix such that

$$\Sigma_* = P_* \Lambda_* P'_* .$$

Set $Z = \Lambda_*^{-\frac{1}{2}} P'_* \hat{T}_*$. Then $Z \sim N(0, I)$ and

$$\hat{T}'_* \hat{T}_* = Z' \Lambda_* Z = \sum_{i=1}^{r} \lambda_{*i} Z_i^2 .$$

The λ_{*i} are the eigenvalues of Σ_*, i.e., roots of

$$\begin{aligned} 0 &= |\Sigma_* - \lambda_* I| \\ &= |\Lambda^{-\frac{1}{2}} P' \Sigma P \Lambda^{-\frac{1}{2}} - \lambda_* I| . \end{aligned}$$

This is equivalent to

$$\begin{aligned} 0 &= |\Sigma - \lambda_* P \Lambda P'| \\ &= |\Sigma - \lambda_* V| \\ &= |V - BJ^{-1}B' - \lambda_* V| \\ &= (-1)^k |BJ^{-1}B' - (1-\lambda_*)V| . \end{aligned}$$

Now, Lemma 4.3 implies (4.11). To conclude the lemma note that both Σ_* and $BJ^{-1}B'$ are non-negative definite.

Lemma 4.6 gives an alternative proof for the inequalities (4.5) and (4.10) in Theorem 4.5 since

$$\tilde{Q} \overset{d}{\doteq} \sum_{i=k+1}^{r} Z_i^2 ,$$

$$\hat{Q}_N(\hat{\theta}_N) \overset{d}{\doteq} \sum_{i=1}^{r} Z_i^2 ,$$

and

$$Q_N(\hat{\theta}_N) \overset{d}{\doteq} \sum_{i=1}^{k} \lambda_i Z_i^2 + \sum_{i=k+1}^{r} Z_i^2$$

implies the asymptotic stochastic ordering

$$\tilde{Q} \overset{d}{\lessdot} Q_N(\hat{\theta}_N) \overset{d}{\lessdot} \hat{Q}_N(\hat{\theta}_N) .$$

IV. 2. Numerical Investigation

The result of Theorem 4.5 shows that the widely used procedure of rejecting the null hypothesis if $Q_N(\hat{\theta}_N) > C_\alpha$, where C_α is the upper α quantile of the chi-square distribution with $r-k$ degrees of freedom and $\hat{\theta}_N$ is the maximum likelihood estimator for θ based on the ungrouped sample $Y_1, \ldots, Y_N, \delta_1, \ldots, \delta_n$ as defined in (2.1), will yield an asymptotic size α^* larger than the nominal size α. Theoretically, the difference $\alpha^* - \alpha$ may be quite large in cases where the number of cells is small. If the values $\lambda_1, \ldots, \lambda_k$ in Theorem 4.5 are all close to 1, then we will be losing essentially k degrees of freedom whose effect will be most serious when the number of cells $r+1$ is small. For example, if we have only one parameter θ_1, then, as λ_1 approaches 1, the actual asymptotic size α^* will vary from about .15 to .10 for $r=2, 3$ or 4 when $\alpha = .05$.

In the following, the actual asymptotic size

$$\alpha^* = P\{\sum_{i=1}^{k} \lambda_i Z_i^2 + \sum_{i=k+1}^{r} Z_i^2 > C_\alpha\}$$

is evaluated using the results of Robbins and Pitman (1949) for two cases of one-parameter families of survival distributions. The uniform censoring distribution with density

$$h(x) = \frac{1}{T} I_{[0<x<T]}$$

and the nominal level $\alpha = .05$ are used.

<u>Case i</u> Consider an exponential survival distribution with density

$$f_\theta(x) = (1/\theta)\exp(-x/\theta)$$

In this case the expected proportion of failures is

$$P\{\delta=1\} = 1-(\theta/T)\{1-\exp(-T/\theta)\}, \tag{4.12}$$

The Fisher information quantity is given by

$$J(\theta) = \{1-(\theta/T)[1-\exp(-T/\theta)]\}/\theta^2,$$

and,

$$\tilde{J}(\theta) = B' V^{-1} B$$

$$= \sum_{i=1}^{r} [B(a_{i-1})/\bar{F}_\theta(a_{i-1}) - B(a_i)/\bar{F}_\theta(a_i)]^2/D_i$$

$$= \theta^{-4} \sum_{i=1}^{r} (a_{i-1}-a_i)^2/D_i,$$

where,

$$D_i = \int_{a_{i-1}}^{a_i} [\bar{H}(z)\bar{F}_\theta^2(z)]^{-1} dF_\theta(z)$$

$$= (T/\theta)\exp(T/\theta)\{\ln[(T-a_{i-1})/\theta)]\exp[-(T-a_{i-1})/\theta]$$

$$-\ln[(T-a_i)/\theta]\exp[-(T-a_i)/\theta] + \int_{(T-a_i)/\theta}^{(T-a_{i-1})/\theta} \ln(z)\exp(-z)dz\}.$$

Consider the situation with $\theta = 1$, $T = 3$ and $r = 2$. Compute

$$\lambda_1 = 1-\tilde{J}(1)/J(1)$$

at the partition points in column 1 of Table 3 to produce the values of λ_1 given in column 2. Use Robbins and Pitman's theorem to calculate the corresponding values of α^* given in column 3.

Case ii For the Weibull survival distribution with density

$$f_\theta(x) = (\theta/\sigma_0)(x/\sigma_0)^{\theta-1} \exp\{-(x/\sigma_0)^\theta\}$$

the value of σ_0 is chosen so that when $T=3$ and the shape parameter $\theta=2$ the same probability of failure, $P\{\delta=1\}$, is obtained as that under the exponential survival distribution with unit mean and $T=3$. In this case, using (4.12) we obtain

$$\begin{aligned} P\{\delta=1\} &= (2/3) + (1/3)\exp(-3) \\ &= .683262356 . \end{aligned} \tag{4.13}$$

Under the Weibull survival distribution with shape parameter $\theta=2$ and uniform (0, 3) censoring we have

$$\begin{aligned} P(\delta=1) &= \frac{1}{3}\int_0^3\int_0^v (2/\sigma_0)(x/\sigma_0)\exp(-(x/\sigma_0)^2)\,dx\,dv \\ &= \frac{1}{3}\int_0^3 (1-\exp(-(V/\sigma_0)^2)dv \\ &= \sigma_0/(3\sqrt{2})\int_0^{3\sqrt{2}/\sigma_0} (1-\exp(-y^2/2))dy \\ &= 1-(\sqrt{\pi}\,\sigma_0/3)[\Phi(3\sqrt{2}/\sigma_0)-\tfrac{1}{2}] \end{aligned} \tag{4.14}$$

where $\Phi(a)$ represents the probability that a standard normal random variable is less than a. Substitute (4.13) in (4.14) and solve for σ_0 to obtain $\sigma_0 = 1.07228198$.

Now, consider the general parameter θ ,

$$\bar{F}_\theta(x) = \exp\{-(x/\sigma_0)^\theta\} .$$

So,

$$B = \frac{\partial F_\theta(x)}{\partial \theta}$$

$$= (x/\sigma_0)^\theta \ln(x/\sigma_0) \exp\{-(x/\sigma_0)^\theta\}$$

$$\tilde{J}(\theta) = \sum_{i=1}^{r} [B(a_{i-1})/\bar{F}_\theta(a_{i-1}) - B(a_i)/\bar{F}_\theta/(a_i)]^2/D_i$$

where

$$D_i = \frac{6}{\sigma_0^2} \int_{3-a_i}^{3-a_{i-1}} [(3-y)/y] \exp\{((3-y)/\sigma_0)^2\} dy .$$

To obtain $J(\theta)$, recall the likelihood function from equation (2.5). In our case, the logarithm of the likelihood function will be

$$\mathscr{L} = \sum_{j=1}^{N} \delta_j [\ln\theta - \theta \ln\sigma_0 + (\theta-1)\ln y_j - (y_j/\sigma_0)^\theta + \ln[(3-Y_j)/Y_j]]$$

$$- \sum_{j=1}^{N} (1-\delta_j)[\ln 3 + (y_j/\sigma_0)^\theta] .$$

Differentiate twice with respect to the unknown parameter θ to obtain

$$\frac{\partial^2 \mathscr{L}}{\partial \theta^2} = - \sum_{j=1}^{N} \{\delta_j [\frac{1}{\theta^2} + (Y_j/\sigma_0)^\theta \ln^2(Y_j/\sigma_0)] + (1-\delta_j)(Y_j/\sigma_0)^\theta \ln^2(Y_j/\sigma_0)\} .$$

So,

$$J(\theta) = -\frac{1}{N} E\frac{\partial^2 \mathcal{L}}{\partial\theta^2}$$
$$= E(\delta/\theta^2) + E(Y/\sigma_0)^\theta \ln^2(Y/\sigma_0)$$

where

$$E(Y/\sigma_0)^2 \ln^2(Y/\sigma_0) = 1/T\int_0^T (y/\sigma_0)^2 \ln^2(y/\sigma_0)\exp[-(y/\sigma_0)^2]\,dy$$

$$+ 2/T\int_0^T (y/\sigma_0)^3 \ln^2(y/\sigma_0)(T-y)\exp[-(y/\sigma_0)^2]\,dy$$

$$= \sigma_0 \int_0^{T/\sigma_0} [T^{-1} + 2y - 2y - 2\sigma_0 T^{-1} y^2]\, y^2 e^{-y^2} \ln^2 y\, dy.$$

Put $T=3$ and $\sigma_0 = 1.07228198$ to give $P\{\delta=1\}$ as in (4.14) when the shape parameter $\theta=2$. Then, corresponding to the partition points in column 1 of Table 3 numerical integration leads to the values of $\lambda_1 = 1-\tilde{J}(2)/J(2)$ and α^* that are in columns 4 and 5 respectively.

Recall that the nominal value $\alpha = .05$ was used. So, for the chosen sets of partition points, for only 3 intervals, the actual asymptotic size of the test, α^*, varies from 1.76α to 2.22α in the exponential case and from 1.47α to 2.70α in the Weibull case.

Table 3. The eigenvalues λ_1 and the corresponding asymptotic sizes α^* of the statistic $Q_N(\hat{\theta}_N)$

Partition points a_1, a_2	Exponential, mean $\theta=1$		Weibull, shape $\theta=2$	
	λ_1	$\alpha^{*(1)}$	λ_1	α^*
2.0, 2.5	.746	.111	.774	.115
1.5, 2.5	.647	.098	.808	.119
0.5, 2.5	.640	.097	.594	.091
1.3, 2.7	.625	.095	.920	.135
1.2, 2.2	.597	.092	.838	.123
0.5, 2.0	.589	.091	.431	.074
1.0, 2.0	.575	.089	.771	.114
0.5, 1.5	.567	.088	.445	.075

(1) $\alpha^* = P\{\lambda_1 Z_1^2 + Z_2^2 \geq C\}$ where Z_1, Z_2 are independent standard normal random variables and $C=3.8416$ is the 95 percentile of a chi-square random variable with 1 degree of freedom.

V. RANDOM CELL BOUNDARIES

In the previous chapters we suggested using the statistic $\hat{Q}_N(\hat{\theta}_N)$ defined by (3.16) as a test of fit and showed its limiting chi-square distribution. Also we obtained the limiting distribution of the statistic $Q_N(\hat{\theta}_N)$ defined by (4.2). The limiting distributions of $\hat{Q}_N$ and Q_N were obtained for fixed cell boundaries $0 = a_0 < a_1 < \cdots < a_{r+1} = \infty$.

In this chapter we will show that these results remain true for the random cell boundaries corresponding to specified positive cell probabilities $p_1, \ldots, p_{r+1}$.

Using a different approach, Watson (1957, 1958) proved the corresponding results for those included in Chapter IV in the case of complete samples. Moore (1970, 1971) showed that the same results hold for the case of multivariate observations and gave a table for the percentile points of the statistic for testing the univariate normal distribution.

We are interested in testing the hypothesis (2.6) in an interval $[0, T]$ where $G_\theta(T)$ is less than one. Let us require that $F_\theta^{-1}(x)$ exists and is jointly continuous in x and θ.

V.1. Preliminaries

Let $Y_1, \ldots, Y_N$ and $\delta_1, \ldots, \delta_N$ be as given by (2.1). Let

$\hat{\theta}_N$ be the maximum likelihood estimator for the parameter vector θ in the family of distributions specified by the hypothesis (2.6). Let $p_1, \ldots, p_{r+1}$ be a fixed set of positive probabilities that add up to unity. Set $a_0 = \hat{a}_0 = 0$ and for $j = 1, \ldots, r+1$ define a_j (depends on θ) and $\hat{a}_j$ by

$$\left.\begin{aligned} F_\theta(a_j) &= P_j F_\theta(T) \\ F_{\hat{\theta}_N}(\hat{a}_j) &= P_j F_{\hat{\theta}_N}(T) \end{aligned}\right\} \qquad (5.1)$$

and

$$P_j = \sum_{i=1}^{j} p_i$$

Note that $\hat{a}_j \xrightarrow{p} a_j$ and that a_r and $\hat{a}_r$ are less than T. Corresponding to the set of partition points $\hat{a}_1 < \ldots < \hat{a}_r$, define $\hat{\hat{T}}_N$ by

$$\hat{\hat{T}}_N = \sqrt{N}\,(\hat{\hat{\bar{F}}}_N - \hat{\bar{F}}_{\hat{\theta}_N}) \qquad (5.2)$$

where $\hat{\hat{\bar{F}}}_N = (\hat{\bar{F}}_N(\hat{a}_1), \ldots, \hat{\bar{F}}_N(\hat{a}_r))'$ and $\hat{\bar{F}}_{\hat{\theta}_N} = (\bar{F}_{\hat{\theta}_N}(\hat{a}_1), \ldots, \bar{F}_{\hat{\theta}_N}(\hat{a}_r))'$ are the product-limit and the maximum likelihood estimators for the survival function respectively. Similarly define $\hat{T}_N$ by

$$\hat{T}_N = \sqrt{N}(\hat{\bar{F}}_N - \bar{F}_{\hat{\theta}_N}) \qquad (5.3)$$

corresponding to $a_1 < \ldots < a_r$. Then, the main result of this chapter (contained in Theorem 5.5) is to show that $\hat{\hat{T}}_N$ and $\hat{T}_N$ have the same limiting distribution. First we show, in Theorem

5.4, that

$$\hat{Z}_N(\hat{a}_j) = \hat{Z}_N(a_j) + o_p(1)$$

where

$$\hat{Z}_N(t) = \sqrt{N}(\Lambda_N^e(t) - \Lambda_{\hat{\theta}_N}(t)), \tag{5.4}$$

and $\Lambda_N^e(t)$ and $\Lambda_{\hat{\theta}_N}(t)$ are the empirical and maximum likelihood estimators for the cumulative hazard function respectively. We will need the following lemmas

Lemma 5.1. Let $\{D_N\}$ be a sequence of random variables such that $\{D_N\}$ converges in distribution to the random variable D. Then D_N is bounded in probability.

Proof: See Chen (1975, Lemma 2.1).

Lemma 5.2. If $\{D_N\}$ is defined as in Lemma 5.1 and $\tilde{N}_N \xrightarrow{p} \infty$ as $N \to \infty$, then $\{D_{\tilde{N}_N}\}$ converges in distribution to D.

Proof: Given $\epsilon > 0$, we must find N_0 such that

$$|P\{D_{\tilde{N}_N} \leq t\} - P\{D \leq t\}| < \epsilon \qquad \text{for all } N \geq N_0 .$$

From the assumptions of the lemma, there exists N_1 such that for $N \geq N_1$ we have

$$|P\{D_N \leq t\} - P\{D \leq t\}| < \frac{\epsilon}{2} ,$$

and there exists N_2 such that for $N > N_2$ we have

$$P\{\tilde{N}_N \geq N_1\} > 1 - \frac{\epsilon}{2} .$$

Now,

$$P\{D_{\tilde{N}_N} \leq t\} = \sum_{k=0}^{\infty} P\{D_{\tilde{N}_N} \leq t \mid \tilde{N}_N = k\} P\{\tilde{N}_N = k\}$$

$$= \sum_{k=0}^{\infty} P\{D_k \leq t\} P\{\tilde{N}_N = k\} .$$

So,

$$\left|P\{D_{\tilde{N}_N} \leq t\} - P\{D \leq t\}\right| = \left| \sum_{k=0}^{\infty} [P\{D_k \leq t\} - P\{D \leq t\}] P\{\tilde{N}_N = k\} \right|$$

$$\leq \sum_{k=0}^{\infty} \left|P\{D_k \leq t\} - P\{D \leq t\}\right| P\{\tilde{N}_N = k\}$$

$$= \sum_{k=0}^{N_1 - 1} \left|P\{D_k \leq t\} - P\{D \leq t\}\right| P\{\tilde{N}_N = k\}$$

$$+ \sum_{k=N_1}^{\infty} \left|P\{D_k \leq t\} - P\{D \leq t\}\right| P\{\tilde{N}_N = k\}$$

$$\leq \sum_{k=0}^{N_1 - 1} P\{\tilde{N}_N = k\} + \frac{\epsilon}{2} \sum_{k=N_1}^{\infty} P\{\tilde{N}_N = k\}$$

$$\leq P\{\tilde{N}_N < N_1\} + \frac{\epsilon}{2}$$

$$< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \qquad \text{for} \quad N \geq N_2 .$$

The lemma is concluded by setting $N_0 = N_2$.

<u>Lemma 5.3.</u> Let $\hat{Z}_N = (\hat{Z}_N(a_1), \ldots, \hat{Z}_N(a_r))'$ where $\hat{Z}_N(t)$ is defined by (5.4). Then, under the condition R of Chapter III and assuming $G_\theta(T) < 1$, $\hat{Z}_N \xrightarrow{L} \hat{Z}$ where $\hat{Z}$ has a zero mean normal distribution and $\text{cov}(\hat{Z}_i, \hat{Z}_j) = \sigma_{ij} \exp\{\Lambda_\theta(a_i) + \Lambda_\theta(a_j)\}$ with $\Sigma = \{\sigma_{ij}\}$ as given by (3.2).

Proof: From equation (7.12) in Breslow and Crowley (1974) we have

$$\sqrt{N}(\hat{\bar{F}}_N(t)-\bar{F}_{\hat{\theta}_N}(t))=\hat{Z}_N(t)[-e^{-\Lambda_{\hat{\theta}_N}(t)}+\frac{1}{2}e^{-\Lambda_N^*}(\Lambda_N^e(t)-\Lambda_{\hat{\theta}_N}(t))]+R_{2N}(t) \tag{5.5}$$

where

$$R_{2N}(t) = -e^{-\Lambda_N^{**}(t)}\sqrt{N}(-\ln\bar{F}_N(t)-\Lambda_N^e(t))$$

$$= o_p(1)$$

resulting from the fact that (Lemma 1 in Breslow and Crowley) whenever $t \leq T$ (note $a_r < T$)

$$0 < -\ln\hat{\bar{F}}_N(t)-\Lambda_N^e(t) < \frac{1}{N}\left(\frac{N}{N(t)}-1\right) \leq \frac{1}{N}\left(\frac{N}{N(T)}-1\right)$$

where

$$N(t) = \sum_{i=1}^{N} I_{[Y_i \geq t]}$$

Let A_N be the diagonal matrix whose i^{th} diagonal element is defined by

$$A_N(i) = 1/[-e^{-\Lambda_{\hat{\theta}_N}(a_i)}+\frac{1}{2}e^{-\Lambda_N^*}(\Lambda_N^e(a_i)-\Lambda_{\hat{\theta}_N}(a_i))]$$

for $i = 1, \ldots, r$. So that (5.5) gives

$$\hat{Z}_N = A_N(\hat{T}_N - o_p(1))$$

where $A_N(i) \xrightarrow{p} A(i) = -e^{\Lambda_\theta(a_i)}$ resulting from the consistency of

$\hat{\theta}_N$ and $\Lambda_N^e(a_i)$ as estimates for θ and $\Lambda_\theta(a_i)$ respectively. Therefore, the lemma will be concluded by applications of Slutsky's theorem (Bickel andDoksum, 1977, p. 461) and of Theorem 3.1.

V. 2. Asymptotic Effect of Using Random Cell Boundaries

Watson (1958) studied the effect of using random cell boundaries $\hat{a}_1, \ldots, \hat{a}_r$ in constructing the Pearson chi-square statistic. He showed that the asymptotic difference between this statistic and the quadratic form obtained by using the boundaries $a_1, \ldots, a_r$ is negligible in the case of complete samples. For the case of random censorship we take a different approach by applying the following theorem.

Theorem 5.4. Let $\hat{Z}_N(t)$ be defined as in (5.4). Then $\hat{Z}_N(\hat{a}_j) - \hat{Z}_N(a_j)$ converges to zero in probability where a_j and $\hat{a}_j$ are defined by (5.1) for $j = 1, \ldots, r$.

Proof: For ease of notation let $\hat{t}$ and t denote $\hat{a}_j$ and a_j respectively for any $j = 1, \ldots, r$. So,

$$\hat{Z}_N(\hat{t}) = \sqrt{N}(\Lambda_N^e(\hat{t}) - \Lambda_{\hat{\theta}_N}(\hat{t}))$$

$$= \sqrt{N}\,\{(\Lambda_N^e(t) - \Lambda_{\hat{\theta}_N}(t) + (\Lambda_N^e(\hat{t}) - \Lambda_N^e(t)) + (\Lambda_{\hat{\theta}_N}(t) - \Lambda_{\hat{\theta}_N}(\hat{t}))\}$$

$$= \hat{Z}_N(t) + R_{1N} + R_{2N} \qquad (5.6)$$

where,

$$R_{1N} = \sqrt{N}\,\{\Lambda_{\theta}(\hat{t}) - \Lambda_{\theta}(t) + \Lambda_{\hat{\theta}_N}(t) - \Lambda_{\hat{\theta}_N}(\hat{t})\}\,,$$

$$R_{2N} = Z_N(\hat{t}) - Z_N(t)\,,$$

$$Z_N(t) = \sqrt{N}(\Lambda_N^{e}(t) - \Lambda_{\theta}(t))$$

By Taylor expansions, under the condition R in Chapter III we can write

$$\Lambda_{\theta}(\hat{t}) = \Lambda_{\theta}(t) + \frac{\partial \Lambda_{\theta}(t)}{\partial t}(\hat{t}-t) + o_p(N^{-\frac{1}{2}})\,,$$

$$\Lambda_{\hat{\theta}_N}(t) = \Lambda_{\theta}(t) + \frac{\partial \Lambda_{\theta}(t)}{\partial \theta}(\hat{\theta}_N-\theta) + o_p(N^{-\frac{1}{2}})\,,$$

and

$$\Lambda_{\hat{\theta}_N}(\hat{t}) = \Lambda_{\theta}(t) + \frac{\partial \Lambda_{\theta}(t)}{\partial t}(\hat{t}-t) + \frac{\partial \Lambda_{\theta}(t)}{\partial \theta}(\hat{\theta}_N-\theta) + o_p(N^{-\frac{1}{2}})$$

Therefore,

$$R_{1N} = o_p(1). \tag{5.7}$$

Now, to show that $R_{2N} = o_p(1)$ recall equation (7.9) in Breslow and Crowley (1974) which states that

$$Z_N(t) = A_N(t) + B_N(t) + R_N(t)$$

where $R_N(t) \xrightarrow{p} 0$ in the supremum metric, so

$$\left.\begin{aligned} Z_N(t) &= A_N(t) + B_N(t) + o_p(1)\,, \\ &\text{and} \\ Z_N(\hat{t}) &= A_N(\hat{t}) + B_N(\hat{t}) + o_p(1)\,, \end{aligned}\right\} \tag{5.8}$$

where

$$A_N(t) = \int_0^t X_N(x)(1-G_\theta(x))^{-2} d\tilde{G}_\theta(x) , \tag{5.9}$$

$$B_N(t) = Y_N(t)[1-G_\theta(t)]^{-1} - \int_0^t Y_N(x)(1-G_\theta(x))^{-2} dG_\theta(x), \tag{5.10}$$

$$X_N(x) = \sqrt{N}[G_N^e(x) - G_\theta(x)] ,$$

and

$$Y_N(x) = \sqrt{N}[\tilde{G}_N^e(x) - \tilde{G}_\theta(x)]$$

and $G_N^e(x)$ and $\tilde{G}_N^e(x)$ are defined by (3.5) and (3.6) respectively. From (5.8) and the triangle inequality we have

$$|Z_N(\hat{t}) - Z_N(t)| \leq |A_N(\hat{t}) - A_N(t)| + |B_N(\hat{t}) - B_N(t)| + o_p(1). \tag{5.11}$$

Using (5.9) we have

$$|A_N(\hat{t}) - A_N(t)| = \left| \int_{\min(t,\hat{t})}^{\max(t,\hat{t})} X_N(x)(1-G_\theta(x))^{-2} d\tilde{G}_\theta(x) \right|$$

$$\leq \int_{\min(t,\hat{t})}^{\max(t,\hat{t})} |X_N(x)|(1-G_\theta(x))^{-2} d\tilde{G}_\theta(x)$$

$$\leq \sup_x |X_N(x)| \int_{\min(t,\hat{t})}^{\max(t,\hat{t})} (1-G_\theta(x))^{-2} d\tilde{G}_\theta(x)$$

Therefore,

$$P\{|A_N(\hat{t})-A_N(t)|<\epsilon\}\geq P\{\sup_x |X_N(x)| \int_{\min(t,\hat{t})}^{\max(t,\hat{t})} (1-G_\theta(x))^{-2} d\tilde{G}_\theta(x) < \epsilon\} \tag{5.12}$$

Now, set

$$D_N = \sup_x |X_N(x)|$$

then D_N is a Kolmogorov-Smirnov type statistic. Hence $\{D_N\}$ converges, under $G_\theta(y)$, in law (Rao, 1973, p. 421) to some random variable D say. An application of Lemma 5.1 shows that, given $\eta > 0$, there exist M and N_1 such that for $N > N_1$

$$P\{D_N < M\} > 1-\eta/2. \tag{5.13}$$

Since $\int_a^b (1-G_\theta(x))^{-2} d\tilde{G}_\theta(x)$ is continuous in a and b, for a given $\epsilon > 0$, there exists ϵ_2 such that

$$\int_{t-\epsilon_2}^{t+\epsilon_2} (1-G_\theta(x))^{-2} d\tilde{G}_\theta(x) < \frac{\epsilon}{M} . \tag{5.14}$$

But $\hat{t}$ converging to t in probability implies that given ϵ_2 and η there exists N_2 such that

$$P\{t-\epsilon_2 < \hat{t} < t+\epsilon_2\} > 1-\eta/2 \quad , \quad N > N_2 .$$

So, for $N > N_2$

$$P\{\int_{\min(t,\hat{t})}^{\max(t,\hat{t})} (1-G_\theta(x))^{-2} dG_\theta(x) < \int_{t-\epsilon_2}^{t+\epsilon_2} (1-G_\theta(x))^{-2} d\tilde{G}_\theta(x)\} > 1-\eta/2 \tag{5.15}$$

From (5.14) and (5.15) we have for $N > N_2$

$$P\{\int_{\min(t,\hat{t})}^{\max(t,\hat{t})} (1-G_\theta(x))^{-2} d\tilde{G}_\theta(x) < \frac{\epsilon}{M}\} > 1 - \eta/2 .$$

Combine the above inequality with (5.12) and (5.13) to conclude that: Given $\epsilon > 0$ and $\eta > 0$, then, for $N > \max(N_1, N_2)$ we have

$$P\{|A_N(\hat{t}) - A_N(t)| < \epsilon\} > 1-\eta$$

i.e.,

$$A_N(\hat{t}) - A_N(t) = o_p(1) \tag{5.16}$$

Using (5.10) we have

$$|B_N(\hat{t}) - B_N(t)| \leq \left|\frac{Y_N(\hat{t})}{1-G_\theta(\hat{t})} - \frac{Y_N(t)}{1-G_\theta(t)}\right| + \int_{\min(t,\hat{t})}^{\max(t,\hat{t})} |Y_N(x)|(1-G_\theta(x))^{-2} dG_\theta(x). \tag{5.17}$$

Now, let $p = P\{\delta = 1\}$ and $\tilde{N}_N$ denote the number of failed observations so that

$$\left.\begin{aligned} &\tilde{p} = \frac{\tilde{N}_N}{N} \xrightarrow{P} p \\ &\text{and} \\ &\sqrt{N}(\tilde{p} - p) \xrightarrow{L} N(0, p(1-p)) \end{aligned}\right\} \tag{5.18}$$

Hence, given any $\epsilon > 0$ there exists $N_1 = N_1(\epsilon)$ such that for all $N > N_1$

$$P\{ \frac{1}{\sqrt{\tilde{p}}} \leq \frac{2}{\sqrt{p}} \} > 1 - \frac{\epsilon}{2} \tag{5.19}$$

Also, from Kolmogorov-Smirnov theorem and Lemma 5.2

$$\tilde{D}_N = \sqrt{\tilde{N}_N} \sup_x \left| \frac{1}{\tilde{\tilde{p}}} \tilde{G}_N^e(x) - \frac{1}{p} G_\theta(x) \right| \tag{5.20}$$

converges in distribution to some random variable, $\tilde{D}$ say. Lemma 5.1 implies that there exists M_ϵ such that

$$P\{\tilde{D}_N < M_\epsilon\} > 1 - \frac{\epsilon}{2} \tag{5.21}$$

Now,

$$|Y_N(x)| = \sqrt{N} \, |\tilde{G}_N^e(x) - \tilde{G}_\theta(x)|$$

$$= p\sqrt{N} \left| \frac{1}{p} \tilde{G}_N^e(x) - \frac{1}{p} \tilde{G}_\theta(x) \right|$$

$$\leq Y_N^{(1)}(x) + Y_N^{(2)}(x) \tag{5.22}$$

where

$$Y_N^{(1)}(x) = p\sqrt{N} \left| \frac{1}{p} \tilde{G}_N^e(x) - \frac{1}{\tilde{\tilde{p}}} \tilde{G}_N^e(x) \right| ,$$

and

$$Y_N^{(2)}(x) = p\sqrt{N} \left| \frac{1}{\tilde{\tilde{p}}} \tilde{G}_N^e(x) - \frac{1}{p} \tilde{G}_\theta(x) \right| .$$

And so,

$$Y_N^{(1)}(x) = p\sqrt{N} \left| \frac{1}{p} - \frac{1}{\tilde{\tilde{p}}} \right| \tilde{G}_N^e(x)$$

$$= \frac{1}{\tilde{\tilde{p}}} \sqrt{N} \, |\tilde{p} - p| \, \tilde{G}_N^e(x) .$$

Use (5.18) and the consistency of $\tilde{G}_N^e(x)$ as an estimator for $\tilde{G}_\theta(x)$ to conclude that

$$Y_N^{(1)}(x) \xrightarrow{L} \frac{1}{p} |Z^* \sqrt{\tilde{G}_\theta(x)}, \tag{5.23}$$

where $Z^* \sim N(0, p(1-p))$. Also,

$$Y_N^{(2)}(x) = p \frac{\sqrt{N}}{\sqrt{\tilde{N}_N}} \{\sqrt{\tilde{N}_N} | \frac{1}{\tilde{p}} \tilde{G}_N^e(x) - \frac{1}{p} \tilde{G}_\theta(x) |\} .$$

Use (5.20) to write

$$Y_N^{(2)}(x) \leq \frac{P}{\sqrt{\tilde{p}}} \tilde{D}_N .$$

Finally use (5.19) and (5.21) to conclude that for $N > N_1$

$$P\{Y_N^{(2)}(x) < 2\sqrt{P}\, M_\epsilon \} > 1 - \epsilon . \tag{5.24}$$

Use (5.23) and (5.24) in (5.22) to conclude that $Y_N(x)$ is bounded in probability. Hence the second term in (5.17) can be shown to converge to zero in probability by similar steps to those leading to (5.16). The continuity of $G_\theta(t)$ in t and the fact that $\hat{t}$ converges to t in probability implies that $G_\theta(\hat{t})$ converges to $G_\theta(t)$ in probability. Thus, to conclude that

$$|B_N(\hat{t}) - B_N(t)| = o_p(1) \tag{5.25}$$

it is sufficient to show that

$$|Y_N(\hat{t}) - Y_N(t)| = o_p(1). \tag{5.26}$$

Let $Y(t)$ denote the limit of $Y_N(t)$ as N increases as given by Breslow and Crowley (1974). Then, from the triangle inequality we can write

$$|Y_N(\hat{t}) - Y_N(t)| \leq |Y_N(\hat{t}) - Y(\hat{t})| + |Y(t) - Y_N(t)| + |Y(\hat{t}) - Y(t)|.$$

Since Y_N converges to Y in the uniform measure (Breslow and Crowley, 1974), i.e.,

$$\rho(Y_N, Y) = \sup_x |Y_N(x) - Y(x)| = o_p(1),$$

it follows that for each N

$$|Y_N(t) - Y(t)| \leq \rho(Y_N, Y) = o_p(1)$$

and

$$|Y_N(\hat{t}) - Y(\hat{t})| \leq \rho(Y_N, Y) = o_p(1)$$

To conclude (5. 26) we need to prove that

$$|Y(\hat{t}) - Y(t)| = o_p(1). \tag{5.27}$$

From the Chebyshev's inequality we have for a given $\epsilon > 0$

$$P\{|Y(\hat{t}) - Y(t)| > \epsilon\} \leq \frac{E\{[Y(\hat{t}) - Y(t)]^2\}}{\epsilon^2}$$

i. e.,

$$P\{|Y(\hat{t}) - Y(t)| > \epsilon\} \leq \frac{E\{E[(Y(\hat{t}) - Y(t))^2 | \hat{t}]\}}{\epsilon^2} \tag{5.28}$$

where

$$E[(Y(\hat{t}) - Y(t))^2 | \hat{t}] = (1 - |\tilde{G}_\theta(\hat{t}) - \tilde{G}_\theta(t)|) |\tilde{G}_\theta(\hat{t}) - \tilde{G}_\theta(t)|$$

$$\leq |\tilde{G}_\theta(\hat{t}) - \tilde{G}_\theta(t)| . \tag{5.29}$$

To compute the above conditional expectation we used the fact that $Y(t)$ is a zero mean Gaussian process with covariance function defined for $t \leq t'$ by

$$\mathrm{Cov}(Y(t), Y(t')) = \tilde{G}_\theta(t)(1 - \tilde{G}_\theta(t')) .$$

Use (5.29) in (5.28) to write

$$P\{|Y(\hat{t}) - Y(t)| > \epsilon\} \leq \frac{E|\tilde{G}_\theta(\hat{t}) - \tilde{G}_\theta(t)|}{\epsilon^2} \tag{5.30}$$

Now, since $\tilde{G}_\theta(t)$ is an absolutely continuous function in t then for any given $\epsilon > 0$ there exists a $\eta > 0$ such that

$$|\tilde{G}_\theta(t') - \tilde{G}_\theta(t)| < \frac{\epsilon^3}{2} \tag{5.31}$$

for every t, t' such that $|t' - t| < \eta$. But $\hat{t}$ is a consistent estimator for t which implies that for the above ϵ and η there exists a N_3 such that for $N > N_3$ we have

$$P\{|\hat{t} - t| < \eta\} > 1 - \frac{\epsilon^3}{2} \tag{5.32}$$

Combine (5.31) and (5.32) to conclude for $N > N_3$ that

$$P\{|\tilde{G}_\theta(\hat{t}) - \tilde{G}_\theta(t)| < \frac{\epsilon^3}{2}\} > 1 - \frac{\epsilon^3}{2} \tag{5.33}$$

Now, let $F^*(\hat{t})$ be the distribution function of $\hat{t}$ so,

$$E[|\tilde{G}_\theta(\hat{t})-\tilde{G}_\theta(t)|] = \int_{-\infty}^{\infty} |\tilde{G}_\theta(\hat{t})-\tilde{G}_\theta(t)|dF^*(\hat{t})$$

$$= \int_{|\hat{t}-t|<\eta} |\tilde{G}_\theta(\hat{t})-\tilde{G}_\theta(t)|dF^*(\hat{t}) + \int_{|\hat{t}-t|\geq\eta} |\tilde{G}_\theta(\hat{t})-\tilde{G}_\theta(t)|dF^*(\hat{t})$$

$$\leq \int_{|\hat{t}-t|<\eta} |\tilde{G}_\theta(\hat{t})-\tilde{G}_\theta(t)|dF^*(\hat{t}) + P\{|\hat{t}-t|\geq \eta\}$$

Thus, using (5. 31) and (5. 32) in the above inequality, we have for $N > N_3$

$$E\{|\tilde{G}_\theta(\hat{t}) - \tilde{G}_\theta(t)|\} < \frac{\epsilon^3}{2} P\{|\hat{t}-t|<\delta\} + P\{|\hat{t}-t|>\eta\} .$$

$$< \frac{\epsilon^3}{2} + \frac{\epsilon^3}{2} = \epsilon^3$$

Substitute in (5. 30) to conclude that

$$P\{|Y(\hat{t}) - Y(t)| > \epsilon\} < \epsilon \qquad \text{for} \quad N > N_3$$

hence (5. 27) holds and so (5. 26) is true. Consequently (5. 25) holds. Substitute (5. 16) and (5. 25) in (5. 11) to conclude that $R_{2N} = o_p(1)$. Use this fact and (5. 7) in (5. 6) to conclude the theorem.

<u>Theorem 5. 5.</u> Let $\hat{\hat{T}}_N$ and $\hat{T}_N$ be as given in equations (5. 2) and (5. 3) respectively. Then, under the condition R of Chapter III and the assumption $G_\theta(T) < 1$, $\hat{\hat{T}}_N$ and $\hat{T}_N$ have the same limiting distribution.

Proof: Let $\hat{t}$ and t denote $\hat{a}_j$ and a_j respectively for any $j=1,\ldots,r$. We will show that

$$g_N(\hat{t}, t) = o_p(1) , \tag{5.34}$$

where

$$g_N(\hat{t}, t) = \sqrt{N}(\hat{\bar{F}}_N(\hat{t}) - \bar{F}_{\hat{\theta}_N}(\hat{t})) - \sqrt{N}(\hat{\bar{F}}_N(t) - \bar{F}_{\hat{\theta}_N}(t) .$$

Now, using (5.5) and the triangle inequality, we have

$$|g_N(\hat{t}, t)| \leq |[\tfrac{1}{2} e^{-\hat{\Lambda}^*_N}(\Lambda^e_N(\hat{t}) - \Lambda_{\hat{\theta}_N}(\hat{t})) - e^{\Lambda_{\hat{\theta}_N}(\hat{t})}]\hat{Z}_N(\hat{t})$$

$$-[\tfrac{1}{2} e^{-\Lambda^*_N}(\Lambda^e_N(t) - \Lambda_{\hat{\theta}_N}(t)) - e^{-\Lambda_{\hat{\theta}_N}(t)}]\hat{Z}_N(t)| + |R_{2N}(\hat{t})|$$

$$+ |R_{2N}(t)| ,$$

where

$$R_{2N}(t) = e^{-\Lambda^{**}_N}\sqrt{N}(-\ln\hat{\bar{F}}_N(t) - \Lambda^e_N(t)) ,$$

$$R_{2N}(\hat{t}) = e^{-\hat{\Lambda}^{**}_N}\sqrt{N}(-\ln\hat{\hat{\bar{F}}}_N(\hat{t}) - \Lambda^e_N(t)) ,$$

and Λ^*_N, $\hat{\Lambda}^*_N$, Λ^{**}_N and $\hat{\Lambda}^{**}_N$ are all positive values. As shown in the proof of Lemma 5.3 $R_{2N}(t)$ is $o_p(1)$. Similar treatment will show that $R_{2N}(\hat{t})$ is $o_p(1)$. So, with the aid of Theorem 5.4 we can write

$$|g_N(\hat{t}, t)| \leq |g_{1N}(\hat{t}, t) + g_{2N}(\hat{t}) - g_{2N}(t)||\hat{Z}_N(t)|$$

$$+ |g_{3N}(\hat{t}) \cdot o_p(1)| + o_p(1) , \tag{5.35}$$

where

$$g_{1N}(\hat{t}, t) = e^{-\Lambda_{\hat{\theta}_N}(t)} - e^{-\Lambda_{\hat{\theta}_N}(\hat{t})} ,$$

$$g_{2N}(t) = \frac{1}{2\sqrt{N}} e^{-\Lambda_N^*} \hat{Z}_N(t) ,$$

$$g_{2N}(\hat{t}) = \frac{1}{2\sqrt{N}} e^{-\hat{\Lambda}_N^*} \hat{Z}_N(\hat{t}) ,$$

and

$$g_{3N}(\hat{t}) = -e^{-\Lambda_{\hat{\theta}_N}(\hat{t})} + g_{2N}(\hat{t}) .$$

Now

$$g_{1N}(\hat{t}, t) = o_p(1) \tag{5. 36}$$

since $\Lambda_\theta(t) = -\ln\bar{F}_\theta(t)$ is continuous in both θ and t and since $\hat{\theta}_N$ and $\hat{t}$ are consistent for θ and t respectively. Also

$$g_{2N}(t) = o_p(1) \tag{5. 37}$$

and

$$g_{2N}(\hat{t}) = o_p(1) \tag{5. 38}$$

since $\hat{Z}_N(t)$ and $\hat{Z}_N(\hat{t})$ are bounded in probability as a result from Theorem 5. 4, Lemma 5. 3 and Lemma 5. 1. Finally, since $\Lambda_{\hat{\theta}_N}(\hat{t})$ is non-negative, $g_{3N}(\hat{t})$ is bounded in probability. Therefore, by using (5. 36), (5. 37), (5. 38) and the boundedness of $g_{3N}(\hat{t})$, (5. 35) reduces to (5. 34).

V. 3. Application to Examples

In this section we compute the test statistics $\hat{Q}_N(\hat{\theta}_N)$ and $Q_N(\hat{\theta}_N)$ defined by (3.16) and (4.2) respectively. We consider the case when the cell boundaries are random corresponding to equal estimated cell probabilities. The procedure is summarized in the following steps:

1. estimate the parameter vector θ by $\hat{\theta}_N$,
2. set $T = Y_N$, the largest order statistic in the sample of size N,
3. take $p_1 = \ldots = p_{r+1} = \frac{1}{r+1}$ and compute the cell boundaries $\hat{a}_1, \ldots, \hat{a}_r$ by (5.1), namely

$$F_{\hat{\theta}_N}(\hat{a}_j) = P_j F_{\hat{\theta}_N}(T)$$

where

$$P_j = p_1 + \ldots + p_j$$

4. use $\hat{a}_1, \ldots, \hat{a}_r$ from Step 3 to compute $\hat{Q}_N(\hat{\theta}_N)$ and $Q_N(\hat{\theta}_N)$ as done in the previous chapters.

Now, we apply the above steps for Example 1 and Example 2 from Chapter III in which we are interested in testing

H_o: $X \sim F_\theta(x) = 1 - \exp(-x/\theta)$.

Example 1 continued. In this example $y_N = 12.52$ was the maximum of the 104 observations. Recall that $\hat{\theta}_N = 25.48$ from which $F_{\hat{\theta}_N}(y_N) = .389$ is evaluated. In Table 4 below columns 2 and 3 include respectively the product-limit estimator and the fitted exponential distribution evaluated at the estimated partition points given in column 1. Based on this 5 partition points, the computed values of the test statistics are $\hat{Q}_N(\hat{\theta}_N) = 2.46$ with 5 degrees of freedom and $Q_N(\hat{\theta}_N) = 1.77$.

<u>Example 2 continued</u>. In this example $y_N = 23.72$ was the maximum of the 104 generated observations. The maximum likelihood $\hat{\theta}_N = 18.62$ and $F_{\hat{\theta}_N}(y_N) = .721$ were computed. Using 6 random cells $(r=5)$ we obtain $\hat{Q}_N(\hat{\theta}_N) = 35.64$ with 5 degrees of freedom and $Q_N(\hat{\theta}_N) = 20.40$. The estimated values of the partition points, the product-limit estimator, and the fitted exponential are displayed in Table 5.

In both Example 1 and Example 2 note that the conclusion of the test is the same whether the cell boundaries are fixed or random. In Example 1 the values of $\hat{Q}_N(\hat{\theta}_N)$ were found to be 3.66 and 2.46 for the fixed and random cell boundaries respectively. Neither is significant (based on 5 degrees of freedom). For the second example the two values were found to be 41.56 and 35.64 respectively, both highly significant. Similarly, the values of $Q_N(\hat{\theta}_N)$ in Example 1 were 2.75 and 1.77 respectively, and in Example 2, were 32.36 and 20.40 respectively.

Table 4. Product-limit estimator and exponential fit for the aortic valve replacement data with random cell boundaries.

Estimated Boundaries Years	Product-limit Estimator	Exponential Fit
0.00	1.0	1.0
1.70	0.914	0.935
3.53	0.855	0.871
5.50	0.796	0.806
7.63	0.745	0.741
9.96	0.659	0.677

Table 5. Product-limit estimator and exponential fit for the simulated Weibull data with random cell boundaries

Estimated Boundaries Years	Product-limit Estimator	Exponential Fit
0.00	1.000	1.000
2.38	0.990	0.898
5.11	0.948	0.807
8.31	0.813	0.725
12.18	0.777	0.651
17.07	0.581	0.585

BIBLIOGRAPHY

Bickel, P.J. and Doksum, K.A. (1977). Mathematical Statistics: Basic Ideas and Selected Topics. Holden-day, Inc.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. Annals of Statistics, 2:437-453.

Chernoff, H. (1954). On the distribution of the likelihood ratio. Annals of Mathematical Statistics, 25:573-578.

Chernoff, H. and Lehmann, E. L. (1954). The use of maximum likelihood estimates in χ^2 tests for goodness of fit. Annals of Mathematical Statistics, 25:579-586.

Chen, J. (1975). Goodness of fit tests under random censorship. Ph. D. thesis, Dept. of Statistics, Oregon State University.

Cramér, H. (1966). Mathematical Methods of Statistics. Princeton University Press, Princeton.

Dahiya, R.C. and Gurland, J. (1972). Pearson chi-square test of fit with random intervals, I. Null case. Biometrika, 59: 147-153.

Durbin, J. (1973). Weak convergence of the sampe distribution function when parameters are estimated. Annals of Statistics, 1:279-290.

Durbin, J. (1973b). Weak convergence of the sample distribution function when parameters are estimated. Annals of Statistics, 1:279-290.

Fisher, R.A. (1922). On mathematical foundation of theoretical statistics. Philadelphia: Transactions of Royal Society A, 222:309-365.

Fisher, R.A. (1924). The conditions under which χ^2 measures the discrepancy between observations and hypothesis. Journal of the Royal Statistical Society, 87:442-450.

Fraser, D.A.S. (1957). Nonparametric Methods in Statistics. John Wiley and Sons, Inc., New York.

Grunkemeier, G. L., Macmanus, Q., Thomas, D. R., Luber, J. M., Lambert, L. E. and Starr, A. (1980). The use of time-interrelated covariates to predict survival following aortic valve replacement. The Annals of Theoretic Surgery, 30:240-246.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53:457-480.

Lehmann, E. L. (1959). Testing Statistical Hypotheses. John Wiley and Sons, Inc. New York.

Mann, H. B. andWald, A. (1943). On stochastic limit and order relationships. Annals of Mathematical Statistics, 14:217-226.

Mihalko, D. P. and Moore, D. S. (1980). Chi-square tests of fit for type II censored data. Annals of Statistics 8:625-644.

Moore, D. S. (1971). A chi-square statistic with random cell boundaries. Annals of Mathematical Statistics, 42:147-156.

Moore, D. S. (1977). Generalized inverse, Wald's method, and the construction of chi-squared tests of fit. Journal of the American Statistical Association, 72:131-137.

Moore, D. S. and Spruill, M. C. (1975). Unified large sample theory of general chi-square statistic for tests of fit. Annals of Statistics, 3:599-616.

Nikulin, M. (1973). Chi-square tests for continuous distributions with shift and scale parameters. Theory of Probability and its Applications, 18:559-568.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophy Magazine Series 5, 50:157-172.

Pierce, D. A. (1980). The asymptotic effect of substituting estimators for parameters in certain types of statistics. Submitted for publication.

Rao, C.R. (1973). Linear Statistical Inference and its Applications. John Wiley and Sons, Inc., New York.

Rao, K.C. and Robson, D.S. (1974). A chi-square statistic for goodness of fit tests within the exponential family. Communications in Statistics, 3:1139-1153.

Robbins, H. and Pitman, E.J.G. (1949). Application of the method of mixtures to quadratic forms in normal variates. Annals of Mathematical Statistics, 20:552-560.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society, 54:426-482.

Watson, G.S. (1957). The chi-squared goodness of fit tests for normal distributions. Biometrika, 44:336-348.

Watson, G.S. (1958). The chi-squared goodness of fit tests for continuous distributions. Journal of Royal Statistical Society B, 20:44-61.

Watson, G.S. (1959). Some recent results in chi-square goodness of fit tests. Biometrics, 15:440-468.

Wilks, S.S. (1963). Mathematical Statistics. John Wiley and Sons, Inc., New York.