CORRELATION-TYPE GOODNESS-OF-FIT TESTS
FOR RANDOMLY CENSORED DATA

BY

CHEN-HSIN CHEN

TECHNICAL REPORT NO. 73
DECEMBER 1981

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT 5 RO1 GM21215-07

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
CORRELATION-TYPE GOODNESS-OF-FIT TESTS
FOR RANDOMLY CENSORED DATA

BY

CHEN-HSIN CHEN

TECHNICAL REPORT NO. 73
December 1981

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT 5 RO1 GM21215-07

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
CORRELATION-TYPE GOODNESS-OF-FIT TESTS
FOR RANDOMLY CENSORED DATA

by
Chen-Hsin Chen

Abstract

For testing the goodness of fit of a distribution \( F \) with randomly censored data to the hypothesis \( F(x) = F_0([x-\theta]/\tau) \) with unknown \( \theta \) and \( \tau > 0 \) and a completely specified distribution function \( F_0 \), or the hypothesis \( F(x) = F_0(x/\tau) \) with \( F(0) = 0 \), this paper introduces correlation-type statistics which are the analogues of the Shapiro-Francia statistic. They measure the correlation between a chosen fixed number of quantiles of the Kaplan-Meier estimator and the corresponding hypothesized quantiles. The test statistics lead to quadratic forms of multivariate normal random variables and hence are asymptotically distributed as weighted sums of independent chi-square variables. The correlation-type tests for exponentiality, in the case of light censoring, are asymptotically robust against departures from the Koziol-Green model of random censorship, which assumes that the censoring distribution \( G \) satisfies \( (1-G) = (1-F)^\beta \) for some \( \beta > 0 \). The tests of exponentiality developed with this model depend only on the censoring proportion. Monte Carlo studies of the power against several specified alternatives are reported. Two examples of clinical trial data are provided for testing exponentiality.
Acknowledgements

It gives me special pleasure to express my deep appreciation to my advisor, Professor Rupert Miller, for his invaluable guidance, continual encouragement, considerate patience, and expert writing assistance throughout the entire work. I am sincerely grateful to Professor Bill Brown for his constant inspiration, as well as for his reading the manuscript and comments on it. Thanks are also due to Professor Ingram Olkin for helpful discussions and to Professor Herbert Solomon for his consent to read the thesis. My particular gratitude is extended to Professor Arthur Albert at Boston University who kindled my interest in Biostatistics.

The computing for this work benefited from the unlimited use of the Statistics Department PDP 11/34A computer. I would like to thank my colleagues at the Department computer room, especially Stephen Peters who introduced the EISPACK to me, Terry Therneau who timely installed a new efficient random number generator RANDM in the system, Stephen Ngok Tsun who made technical suggestions, and Peter Bacchetti who supplied the RCPP data used in Example 2. Thanks also go to Karola Decleve for her delicate and speedy typing of this manuscript.

The financial support from the National Institute of General Medical Sciences during my education at Stanford is acknowledged.

Last but not least, I am greatly indebted to my parents for their love and confidence in me and to Bi-Huey for providing me with the greatest encouragement from the beginning of the research.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Correlation-Type Goodness-of-Fit Tests for Composite Hypotheses</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Censored Data and the Kaplan-Meier Estimator</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Goodness-of-Fit Tests for Simple Hypotheses with Randomly Censored</td>
<td>4</td>
</tr>
<tr>
<td>1.4 Goodness-of-Fit Tests for Composite Hypotheses with Censored Data</td>
<td>5</td>
</tr>
<tr>
<td>1.5 Organization of the Paper</td>
<td>6</td>
</tr>
<tr>
<td>II. Preliminaries and Notation</td>
<td>8</td>
</tr>
<tr>
<td>2.1 Some Preliminary Results</td>
<td>8</td>
</tr>
<tr>
<td>2.2 The Quantiles of the Kaplan-Meier Estimator</td>
<td>10</td>
</tr>
<tr>
<td>III. Correlation-Type Test for Two-Parameter Location-Scale Family of</td>
<td>13</td>
</tr>
<tr>
<td>Distributions</td>
<td></td>
</tr>
<tr>
<td>3.1 Test Statistic</td>
<td>13</td>
</tr>
<tr>
<td>3.2 Asymptotic Distributions of the Test Statistic</td>
<td>16</td>
</tr>
<tr>
<td>IV. Correlation-Type Test for Distribution with Origin Known</td>
<td>25</td>
</tr>
<tr>
<td>4.1 Test Statistic</td>
<td>25</td>
</tr>
<tr>
<td>4.2 Asymptotic Distributions of the Test Statistic</td>
<td>29</td>
</tr>
<tr>
<td>V. Correlation-Type Tests for Exponentiality</td>
<td>32</td>
</tr>
<tr>
<td>5.1 Asymptotic Robustness of the Tests in the Case of Light Censoring</td>
<td>33</td>
</tr>
<tr>
<td>5.2 Tests Developed with the Koziol-Green Model of Random Censorship</td>
<td>39</td>
</tr>
<tr>
<td>5.3 Power Studies</td>
<td>45</td>
</tr>
</tbody>
</table>
I. Introduction

Each parametric method in statistical inference requires an assumption with regard to the distribution that underlies the data. This distributional assumption is the topic of concern in testing goodness of fit.

The null hypotheses of goodness-of-fit, like parametric hypotheses, can be categorized into simple and composite ones. A hypothesis is simple if the hypothesized distribution is completely specified.

1.1. Correlation-Type Goodness-of-Fit Tests for Composite Hypotheses with Complete Data.

For the problem of assessing a composite hypothesis with complete data (i.e., uncensored data), Shapiro and Wilk derived statistics for testing normality (1965) and exponentiality (1972) with the motivation coming from probability plots and generalized least-squares analysis. In the power study by Shapiro, Wilk and Chen (1968), the Shapiro-Wilk statistic for normality was shown to be superior to eight alternative omnibus tests for normality. Shapiro and Francia (1972) presented an approximate and simplified version of the Shapiro-Wilk statistic which is exactly the squared product-moment correlation coefficient between the ordered observations and the expectations of the order statistics from a standardized variate of the hypothesized distribution. Their Monte Carlo studies indicated that these two statistics are equally powerful for testing normality with large samples.
Instead of using the expectations of the order statistics from the standard normal distribution, de Wet and Venter (1972) suggested using the quantiles of the standard normal distribution at \( i/(N+1) \), \( 1 \leq i \leq N \), (here \( N \) is the sample size) in the Shapiro-Francia statistic and derived the asymptotic distribution of this modified statistic for testing normality. Sarkadi (1975) proved that the Shapiro-Francia statistic which tests the hypothesis that the distribution belongs to a two-parameter location-scale family of distributions is consistent and claimed that the Shapiro-Wilk statistic is consistent for testing normality.

Sarkadi also concluded that the Shapiro-Wilk statistic for location and scale parameter families is not consistent for testing exponentiality, which involves only a scale parameter. Stephens (1978) modified the Shapiro-Wilk statistic for testing exponentiality for the case where the origin (location parameter) is known and claimed that the modified version performs substantially better in power comparisons.

1.2. Censored Data and the Kaplan-Meier Estimator.

Censored data arise frequently in industrial life testing and medical clinical trials. Unlike complete data whose actual values are recorded, a censored observation is known only to exceed a certain value. Type I censoring and Type II censoring are typical patterns encountered in engineering applications. Due to loss to follow-up, deaths from competing risks,
withdrawal from and termination of the study, biomedical data usually have a more complicated censoring pattern. Random censorship, which is the censoring type of interest in this paper, has been a commonly-used model in biomedical survival analysis.

To describe the random censorship model precisely, let \( X_1, \ldots, X_N \) denote a random sample from the lifetime distribution \( F \) and \( Y_1, \ldots, Y_N \) a random sample, drawn independently of the \( X_i, 1 \leq i \leq N \), from the censoring distribution \( G \). The observed value for the \( i \)th individual and the variable which indicates whether it is censored or uncensored are

\[
T_i = \min(X_i, Y_i) \quad \text{and} \quad \delta_i = I\{X_i \leq Y_i\},
\]

respectively, for \( 1 \leq i \leq N \). (Here \( I\{S\} \) denotes the indicator function of the set \( \{S\} \).

In the nonparametric analysis of censored data, the Kaplan-Meier (1958) estimator for the distribution function has played an immensely important role. Their product-limit estimator for the lifetime distribution \( F \) is defined by

\[
1 - \hat{F}_N(t) = \prod_{i: T_i \leq t} \left( \frac{N - r_i}{N - T_i + 1} \right)^\delta_i \quad \text{for} \quad t \leq T(N),
\]

where \( T(N) = \max(T_1, \ldots, T_N) \) and \( r_i \) is the rank of \((T_i, 1-\delta_i)\) in the lexicographic ordering of the sequence \((T_1, 1-\delta_1), \ldots, (T_N, 1-\delta_N)\).
Efron (1967) introduced the concept of self-consistency for an estimator and showed that the Kaplan-Meier estimator is the unique self-consistent estimator for a distribution if \( \hat{F}_N(t) \) is defined to be equal to 1 for \( t > T(N) \). Under the random censorship model and the assumption of continuous distributions, Breslow and Crowley (1974) proved the weak convergence of \( \sqrt{N}(\hat{F}_N(t) - F(t)) \), as a function of \( t \) on a finite interval, to a Gaussian process.


The presence of a censoring mechanism substantially complicates the goodness-of-fit problem. In recent years, several test statistics have been proposed and investigated in simple hypothesis testing with randomly censored data. A Cramér-von Mises-type statistic based on \( \hat{F}_N \) was introduced by Koziol and Green (1976). They derived the asymptotic distribution of the statistic only for a particular distributional form of random censorship satisfying \( (1-G) = (1-F)^\beta \) with some positive \( \beta \). This special assumption was called the Koziol-Green model of random censorship by Csörgő and Horváth (1981). Hyde (1977) constructed tests for both discrete and continuous underlying lifetime distributions which are essentially the one-sample analogues of Mantel's (1966) two-sample statistic. The one-sample limit of Efron's (1967) two-sample statistic was presented by Hollander and Proschan (1979). In the same paper, they compared
test with the tests of Koziol-Green and Hyde. In addition, Koziol (1980) proposed generalized versions of Kolmogorov-Smirnov, Kuiper and weighted Cramér-von Mises statistics.

Hall and Wellner (1980) developed symmetric asymptotic simultaneous confidence bands for $F$, centered at $\hat{F}_N$, which reduce to the Kolmogorov bands in the absence of censoring. They also made comparisons with the asymmetric bands of Gillespie-Fisher (1979) and Nair (1981). Under the Koziol-Green model of random censorship, Csörgő and Horváth (1981) constructed empirical exact confidence bands for $F$ and proposed a class of functionals corresponding to Efron's (1967) transformation of $\hat{F}_N$ as goodness-of-fit test statistics. This class includes the Cramér-von Mises-type statistic. They presented the limit theory with the censoring parameter $\beta$ of the Koziol-Green model being estimated from the sample.

All the aforementioned statistics have been designed merely for a simple hypothesis of a continuous lifetime distribution except for Hyde's test.

1.4. **Goodness-of-Fit Tests for Composite Hypotheses with Censored Data.**

For discrete or grouped data subject to random censoring, Turnbull and Weiss (1978) presented a likelihood ratio statistic for testing the hypothesis of goodness-of-fit which may be simple or composite.

Gerlach generalized the Shapiro-Francia statistic to the singly or doubly Type II censored sample case for testing the
two-parameter extreme value or Weibull distribution (1979) and normality (1980) and proved the consistency of the test.

Mihalko and Moore (1980) considered chi-square tests for Type II censoring. The unknown parameters of a continuous distribution in the composite hypothesis are estimated by estimators asymptotically equivalent to linear combinations of functions of observed order statistics.

We seek to develop some method for testing the composite hypotheses of goodness-of-fit with randomly censored data. In the preparation of this paper, it is noted that LaRiccia and Wehrly (1981) are working on minimum quantile distance estimators for randomly censored data which also provide a test of the goodness of fit to an assumed parametric family.

1.5. Organization of the Paper.

The quantile Q-Q plot is a popular graphical technique for assessing the goodness-of-fit of a distribution function. To provide a more trustworthy tool than just an eyeball analysis, this paper presents correlation-type omnibus tests of composite hypotheses for randomly censored data.

Section II states a few results from matrix and eigenvalue theory which are needed in the sequel and describes the quantiles of the Kaplan-Meier estimator in order to define the test statistics.

In Section III a statistic for testing a two-parameter location-scale family of distributions is proposed and its asymptotic distributions under null and specified alternative hypotheses are derived.
In Section IV a statistic for testing a distribution with origin known and its asymptotic distribution are considered.

Section V is a discussion of correlation-type tests of exp-

onentiality under the Koziol-Green model of random censorship.
Some calculations support that, under this model and light cen-
soring, our tests are robust with respect to different censoring
distributions which have the same fixed expected censoring propor-
tion. The upper quantiles of the asymptotic and finite-sample
(N = 30 or 100) null distributions for testing exponentiality
are tabulated with various expected censoring proportions. A
power study of the test discussed in Section IV for exponen-
tiality is given.

In Section VI two collections of real data are tested for
exponentiality as illustrations of our technique. One data set
was furnished in Hollander and Proschan (1979). The other set
dealing with clinical data on myocardial infarction was provided
by the Recurrent Coronary Prevention Project at the Mt. Zion
Hospital, San Francisco.

Section VII gives several remarks to conclude the paper.
II. Preliminaries and Notation

2.1. Some Preliminary Results.

The derivations of the asymptotic distributions in the subsequent sections utilize some probabilistic limit theorems and a number of facts from matrix and eigenvalue theory. In this subsection these needed known results are stated without proof.

Proposition 2.1 (Continuous Mapping Theorem of Convergence).
Suppose that $g: \mathbb{R}^m \to \mathbb{R}^n$ is a continuous function and $\{X_j\}$ is a sequence of $m$-dimensional random vectors.

(i) If $X_j \xrightarrow{L} X$, then $g(X_j) \xrightarrow{L} g(X)$.

(ii) If $X_j \xrightarrow{p} X$, then $g(X_j) \xrightarrow{p} g(X)$.

[See Rao (1973), p. 124, results (xii) and (xiii).]

Proposition 2.2. If $X_n \xrightarrow{L} X$, then $X_n = O_p(1)$. (Here $O_p$ is the stochastic-order notation of boundedness in probability.)

[See Bishop et al. (1975), p. 477, Theorem 14.4-2.]

Proposition 2.3 (Slutsky's Theorem). Let $\{X_n, Y_n\}$ be a sequence of pairs of random variables. If $X_n \xrightarrow{L} X$ and $Y_n \xrightarrow{p} c$, where $c$ is a constant, then

(i) $X_n + Y_n \xrightarrow{L} X + c$,

(ii) $X_n Y_n \xrightarrow{L} cX$, and

(iii) $X_n/Y_n \xrightarrow{L} X/c$, if $c \neq 0$.

[See Rao (1973), p. 122, result (x).]
Proposition 2.4. Suppose that $G = (g_{ij})$ is a Green's matrix of order $n$, i.e.,

$$g_{ij} = \alpha_{\min(i,j)} \beta_{\max(i,j)} \quad \text{for} \quad 1 \leq i, j \leq n,$$

where $\{\alpha_k, 1 \leq k \leq n\}$ and $\{\beta_m, 1 \leq m \leq n\}$ are two sequences of real numbers. If $\alpha_k$ and $\beta_m$ are either all positive or all negative and

$$\frac{\alpha_1}{\beta_1} < \frac{\alpha_2}{\beta_2} < \ldots < \frac{\alpha_n}{\beta_n},$$

then $G$ is positive definite.

[This follows from Theorem 3.1, p. 111, of Karlin (1968).]

Proposition 2.5. Let $X$ be a real symmetric matrix of order $m$ and $X$ be an $m$-dimensional column vector of variables. Then the quadratic form $X^TAX$ can be expressed as

$$X^TAX = \lambda_1 Y_1^2 + \lambda_2 Y_2^2 + \ldots + \lambda_m Y_m^2,$$

where the $\lambda_i, 1 \leq i \leq m$, are the eigenvalues of $X$ including the multiplicities, $Y = (Y_1, \ldots, Y_m)^T = P^TX$, $X$ is a diagonal matrix having $\lambda_i$ as its $i$th diagonal element, and $P$ is an orthogonal matrix whose $i$th column vector is a standardized eigenvector corresponding to $\lambda_i$.

[See Rao (1973), pp. 39-40.]
Proposition 2.6. Let $B$ and $C$ be matrices of order $p \times n$ and $n \times p$, respectively. Then

$$\det(\sim_p + BC) = \det(\sim_n + CB),$$

where $I_p$ denotes the identity matrix of order $p$.

[See Mardia et al. (1979), p. 458, (A.2.3n).]

Proposition 2.7. If $A$ is a Hermitian matrix of order $n$ and $H$ is a positive definite Hermitian matrix of order $n$, then

$$\lambda_j(A) \leq \lambda_j(H) \leq \lambda_j(H)$$

for $1 \leq j \leq n$,

where $\lambda_i(C)$ denotes the $i$th largest eigenvalue of the matrix $C$.

[This follows from (A.1.a.), p. 510, of Marshall and Olkin (1979).]

2.2. The Quantiles of the Kaplan-Meier Estimator.

To set up the test statistics, we need to bring in the quantiles of the Kaplan-Meier estimator. The $p$th quantile of a distribution function $H$ is defined by

$$H^{-1}(p) = \inf\{x : H(x) \geq p\},$$

where the infimum of the empty set is taken to be infinity. The quantile function $H^{-1}(\cdot)$ is indeed the left-continuous inverse of the distribution function $H$. To make the quantile function of the Kaplan-Meier estimator well-defined, $\hat{F}_N(t)$ is henceforth defined to be equal to 1 for $t > T(N)$. 

10
Before exhibiting the weak convergence of \( \hat{F}_N^{-1}(\cdot) \), we assume that the following conditions on the lifetime distribution \( F \) and the censoring distribution \( G \) hold throughout this paper:

(A-1) \( F \) is bounded below, i.e., there exists a real number \( x_0 \) such that \( F(x_0) = 0 \).

(A-2) \( F \) and \( G \) are continuous distributions.

(A-3) There exists an \( A \) such that \( 0 < A < 1 \) and \( G \circ F^{-1}(A) < 1 \).

(A-4) There exists a unique \( x \) such that \( F(x) = p \) for each \( p \in [0,A] \).

(A-5) The continuous density \( f \) of the distribution \( F \) exists and \( f(F^{-1}(p)) \neq 0 \) for all \( p \in [0,A] \).

Sander (1975) proved the following theorem on the weak convergence of \( \hat{F}_N^{-1}(\cdot) \):

**Proposition 2.8.** Suppose that (A-1)-(A-5) hold. Then, as \( N \to \infty \),

\[
\sqrt{N}[\hat{F}_N^{-1}(p) - F^{-1}(p)] \overset{w}{\to} \frac{1}{f(F^{-1}(p))} [-X(p)] , \text{ for } p \in [0,A],
\]

where \( X(\cdot) \) is a mean zero Gaussian process on \( [0,A] \) with covariance

\[
\text{Cov}[X(s), X(t)] = (1-s)(1-t) \int_0^S \frac{du}{(1-u)^2 [1-G \circ F^{-1}(u)]}, \text{ s} \leq t.
\]

[See Sander (1975), p. 5, Corollary 1.]
Actually, Sander proved this theorem with the assumption that \( F(0) = 0 \). The weaker assumption (A-1), however, does not affect the assertion.

The number of uncensored lifetime variables \( X_i \) observed in a randomly censored sample of size \( N \) is random. We therefore choose a fixed integer \( K \), which is greater than 2, and take an appropriate set of \( K \) proportions

\[
0 < p_1 < p_2 < \ldots < p_K < A
\]

to be the basis of the tests. For given \( \{p_i, 1 \leq i \leq K\} \), we obtain \( K \) quantiles of the Kaplan-Meier estimator which are denoted by

\[
\hat{Q}_{Ni} = \hat{F}_N^{-1}(p_i), \ 1 \leq i \leq K.
\]
III. Correlation-Type Test for Two-Parameter Location-Scale Family of Distributions

A goodness-of-fit criterion will be examined for testing the composite null hypothesis

\[ H_0: F(x) = F_0 \left( \frac{x - \theta}{\tau} \right) \text{ for some unknown } \theta \text{ and } \tau > 0 , \]

where \( F_0 \) is a completely specified distribution function. \( \theta \) and \( \tau \) are location and scale parameters of \( F \), respectively.

In the Q-Q plot, the empirical values \( \hat{\xi}_{N1} \) are plotted versus the hypothesized values \( q_i = F_0^{-1}(p_i) \), for \( 1 \leq i \leq K \).

3.1. Test Statistic.

The Shapiro-Francia statistic suggests the following statistic for a randomly censored sample:

\[
R_N = \frac{\sum_{i=1}^{K} (\hat{\xi}_{N1} - \bar{\xi}_N)(q_i - \bar{q})}{\sqrt{\sum_{i=1}^{K} (\hat{\xi}_{N1} - \bar{\xi}_N)^2} \sqrt{\sum_{i=1}^{K} (q_i - \bar{q})^2}} \tag{3.1}
\]

\[
\bar{\xi}_N = \frac{\sum_{i=1}^{K} \hat{\xi}_{N1}/K }{S_N S_0},
\]

where

\[
\hat{\xi}_N = \frac{\sum_{i=1}^{K} \hat{\xi}_{N1} }{K},
\]

\[
\bar{q} = \frac{\sum_{i=1}^{K} q_i }{K}.
\]
\[ S_N^2 = \sum_{i=1}^{K} (\hat{Q}_{Ni} - \bar{Q}_N)^2 , \]

and

\[ S_0^2 = \sum_{i=1}^{K} (q_i - \bar{q})^2 . \]

It is easy to see that \( R_N \) is a nonnegative random variable.

Under \( H_0 \), for large \( N \), the regression of \( \hat{Q}_{Ni} \) on \( q_i \) should be approximately linear, i.e.,

\[ \hat{Q}_{Ni} \approx \theta + \tau q_i , \quad 1 \leq i \leq K , \]

and thus \( R_N \) should be close to 1. This suggests rejecting \( H_0 \) if \( R_N \) is significantly small.

We now consider a statistic equivalent to \( R_N \) and obtain a quadratic form representation for it.

**Lemma 3.1.**

\[ N(1-R_N^2) = \hat{Q}_N^T \Sigma_N^{-1} \hat{Q}_N / S_N^2 , \quad (3.2) \]

where \( \hat{Q}_N^T = \sqrt{N}(\hat{Q}_{N1} - q_1, \ldots, \hat{Q}_{NK} - q_K) \) is a K-dimensional row vector, and

\[ D = I_K - uu^T - aa^T \]

with \( u^T = \left( \frac{1}{\sqrt{K}}, \frac{1}{\sqrt{K}}, \ldots, \frac{1}{\sqrt{K}} \right) \), \( a^T = (a_1, a_2, \ldots, a_K) \) with
\[ a_i = (q_i - \bar{q})/S_0 \quad \text{for} \quad 1 \leq i \leq K, \quad \text{and} \quad I_K \quad \text{is the identity matrix of order} \quad K. \]

**Proof:** By definition (3.1)

\[
R_N^2 = \left[ \sum_{i=1}^{K} \tilde{Q}_{Ni}(q_i - \bar{q}) \right]^2 /(S_N^2 S_0^2)
= \left[ \sum_{i=1}^{K} (\tilde{Q}_{Ni} - q_i)(q_i - \bar{q}) + S_0^2 \right]^2 /(S_N^2 S_0^2),
\]

so we have

\[
NS_N^2 S_0^2 (1 - R_N^2) = NS_0^2 \left[ S_N^2 - 2 \sum_{i=1}^{K} (\tilde{Q}_{Ni} - q_i)(q_i - \bar{q}) - S_0^2 \right]
\]

\[
- N \sum_{i=1}^{K} \sum_{j=1}^{K} (\tilde{Q}_{Ni} - q_i)(\tilde{Q}_{Nj} - q_j)(q_i - \bar{q})(q_j - \bar{q}).
\]

(3.3)

With the identity

\[
S_N^2 = \sum_{i=1}^{K} [ (\tilde{Q}_{Ni} - q_i) + (q_i - \bar{q}) - (\tilde{Q}_N - \bar{q}) ]^2,
\]

(3.3) can be rewritten as

\[
NS_N^2 S_0^2 (1 - R_N^2) = NS_0^2 \left[ \sum_{i=1}^{K} (\tilde{Q}_{Ni} - q_i)^2 \left( 1 - \frac{1}{K} a_i^2 \right) \right]
+ \sum_{1 \leq i \neq j \leq K} (\tilde{Q}_{Ni} - q_i)(\tilde{Q}_{Nj} - q_j) \left( \frac{1}{K} - a_i a_j \right),
\]

where \( a_i = (q_i - \bar{q})/S_0 \), \( 1 \leq i \leq K \). In terms of the matrix notation, the result follows. \( \square \)
It is feasible to derive the asymptotic distributions with this quadratic form representation. Therefore, \( N(1-R_N^2) \) is selected as the test statistic. The test rejects \( H_0 \) if \( N(1-R_N^2) \) is significantly large.

3.2. Asymptotic Distributions of the Test Statistic.

The statistic \( R_N \) is location and scale invariant under \( H_0 \) provided that the censoring distribution \( G \) makes the same location and scale changes as the lifetime distribution \( F \) does. Hence, to study the distribution of \( N(1-R_N^2) \) under \( H_0 \), without loss of generality, we may assume that \( F = F_0 \) and let

\[
G_0(y) = G(\theta + \tau y)
\]  

(3.4)

be the censoring distribution corresponding to this case since, under \( H_0 \), we originally have \( F_0(x) = F(\theta + \tau x) \).

It follows directly from the definition of weak convergence and Proposition 2.8 that

**Lemma 3.2.** Suppose that (A-1)-(A-5) hold. If \( F = F_0 \), then for \( 0 < p_1 < p_2 < \ldots < p_K < A \), as \( N \to \infty \),

\[
Q_N \equiv \sqrt{N} (\hat{q}_{1} q_{1}, \ldots, \hat{q}_{K} q_{K})^T \xrightarrow{L} N_K(0, \Sigma)
\]

where \( N_K(0, \Sigma) \) is a \( K \)-variate normal distribution with mean \( 0 \), the zero vector, and covariance matrix \( \Sigma = \left( \sigma_{ij} \right)_{i,j=1}^{K} \) with elements

16
\[ \sigma_{ij} = \frac{1-p_i}{f_0(F_0^{-1}(p_i))} \cdot \frac{1-p_j}{f_0(F_0^{-1}(p_j))} \int_0^{p_i} \frac{du}{(1-u)^2[1-G_0 \circ F_0^{-1}(u)]}, \]

\[ = \sigma_{ji}, \quad \text{for} \quad 1 \leq i \leq j \leq K. \]  

(3.5)

Here \( f_0 \) is the density of \( F_0 \) and \( G_0 \) is specified in (3.4).

The following lemma enables us to characterize the asymptotic distribution of \( N(1-R_N^2) \).

**Lemma 3.3.** Suppose that (A-1)-(A-5) hold with

\[ 0 < p_1 < p_2 < \ldots < p_K \leq A. \]

Then, under \( H_0 \), as \( N \to \infty \),

\[ N(1-R_N^2) \xrightarrow{\text{L}} W^T_{BW}, \]

where \( W = (W_1, \ldots, W_K)^T \) is a \( K \)-dimensional normal random vector with distribution \( N_K(0, \Sigma) \) as stated in Lemma 3.2 and

\[ B = D/S_0^2 = [I_K - uu^T - aa^T]/S_0^2 \]

(3.6)

with \( u^T \) and \( a^T \) as defined in Lemma 3.1.

**Proof:** From the discussion preceding Lemma 3.2 it suffices to find the asymptotic distribution of \( N(1-R_N^2) \) under \( H_0 \) with \( F = F_0 \) and censoring distribution \( G_0 \).

Propositions 2.8 and 2.2 together imply that for \( 1 \leq i \leq K \),

\[ \hat{Q}_{Ni}^p \xrightarrow{\text{P}} q_i, \quad \text{as} \quad N \to \infty. \]
Thus, by Proposition 2.1 (ii), as $N \to \infty$,

$$S^2_N \equiv \sum_{i=1}^{K} (\hat{q}_{Ni} - \bar{q})^2$$

$$P + \sum_{i=1}^{K} (q_i - \bar{q})^2 \equiv S^2_0.$$  

(3.7)

By Lemma 3.2 and Proposition 2.1 (i), we have

$$\frac{Q^{T}DQ}{\sim N} \sim W^{T}D W \sim N \to \infty.$$  

(3.8)

Application of Proposition 2.3 (iii) with (3.7) and (3.8) to

$$\frac{Q^{T}DQ}{\sim N}S^2_N = N(1-R^2_N)$$

in (3.2) proves this lemma. \hfill \Box

Lemma 3.4. The matrix $\Sigma$ defined in (3.5) of Lemma 3.2 is positive definite.

Proof: Denote

$$\beta_j = \frac{1-p_j}{f_0(F^{-1}(p_j))}$$

and

$$\alpha_i = \beta_i \int_0^{p_i} \frac{du}{(1-u)^2[1-G_0 \circ F^{-1}(u)]}, \text{ for } 1 \leq i, j \leq K.$$  

Then, for $i \leq j$, we obtain

$$\sigma_{ij} = \alpha_i \beta_j = \sigma_{ji}.$$  

Hence, $\Sigma = (\sigma_{ij})$ is a Green's matrix. Moreover, since $i < j$ implies that $p_i < p_j$, we have
\[ \frac{\alpha_1}{\beta_1} < \frac{\alpha_2}{\beta_2} < \ldots < \frac{\alpha_K}{\beta_K}. \]

Thus, by Proposition 2.4, \( \Sigma \) is positive definite. \( \square \)

**Lemma 3.5.** With the same notation as in Lemma 3.3,

\[ W^T \Sigma B W = \sum_{i=1}^{K} \lambda_i Z_i^2, \]

where the \( \lambda_i, 1 \leq i \leq K \), are the eigenvalues of \( \Sigma B \) and the \( Z_i, 1 \leq i \leq K \), are independently and identically distributed (i.i.d.) standard normal random variables.

**Proof:** The previous lemma ensures that \( \Sigma \) is nonsingular, so we may make the transformation

\[ V = \Sigma^{-\frac{1}{2}} W. \]

Then, \( V \) is distributed as \( N_K(0, I_K) \) and

\[ W^T \Sigma B W = V^T \Sigma^{\frac{1}{2}} B \Sigma^{\frac{1}{2}} V. \]

Using Proposition 2.5, we can write

\[ V^T \Sigma^{\frac{1}{2}} B \Sigma^{\frac{1}{2}} V = \sum_{i=1}^{K} \lambda_i Z_i^2, \]

where the \( \lambda_i, 1 \leq i \leq K \), are the eigenvalues of \( \Sigma^{\frac{1}{2}} B \Sigma^{\frac{1}{2}} \), and \( z = (z_1, \ldots, z_K)^T = \Gamma^T V \) where \( \Gamma \) is an orthogonal matrix.
having the eigenvectors of $\Sigma^{-\frac{1}{2}}B\Sigma^{-\frac{1}{2}}$ as its column vectors.

Therefore,

$$W^T B W = \sum_{i=1}^{K} \lambda_i Z_i^2,$$

and $Z$ is also distributed as $N_K(0, I_K)$.

With pre-multiplication by $\Sigma^{-\frac{1}{2}}$ and post-multiplication by $\Sigma^{-\frac{1}{2}}$,

$$0 = \det(\Sigma^{-\frac{1}{2}} B \Sigma^{-\frac{1}{2}} - \lambda I_K),$$

$$= \det(\Sigma B - \lambda I_K),$$

so all the eigenvalues of $\Sigma^{-\frac{1}{2}} B \Sigma^{-\frac{1}{2}}$ are the same as those of $\Sigma B$.

The proof is thus completed. □

**Lemma 3.6.** The matrix $B$ defined in (3.6) has eigenvalues 0 of multiplicity 2 and $S_0^{-2}$ of multiplicity (K-2).

**Proof:**

$$0 = \det(B - \lambda I_K),$$

$$= \det(S_0^{-2} [I_K - uu^T - aa^T] - \lambda I_K),$$

$$= \det\left( (S_0^{-2} - \lambda) I_K - S_0^{-2} (u \ a) \begin{pmatrix} u^T \\ a^T \end{pmatrix} \right),$$

20
\[
= (S_0^{-2} - \lambda)^K \cdot \det \left( I_K - [1-S_0^2 \lambda]^{-1}(u \ a) \left( \begin{array}{c} u^T \\ a^T \end{array} \right) \right),
\]
\[
= (S_0^{-2} - \lambda)^K \cdot \det \left( I_2 - [1-S_0^2 \lambda]^{-1} \left( \begin{array}{c} u^T \\ a^T \end{array} \right) (u \ a) \right) \text{ by Proposition 2.6 ,}
\]
\[
= (S_0^{-2} - \lambda)^K \cdot \det (I_2 - [1-S_0^2 \lambda]^{-1} I_2),
\]
\[
= \lambda^2 (S_0^{-2} - \lambda)^{K-2}.
\]

Hence the result holds. \(\square\)

**Theorem 3.7.** Suppose that (A-1)-(A-5) hold with

\[
0 < p_1 < p_2 < \ldots < p_K \leq A.
\]

Then, under \(H_0\), as \(N \to \infty\),
\[
N(1-R_N^2) \Rightarrow \sum_{i=1}^{K-2} \lambda_i Z_i^2,
\]
where the \(\lambda_i, 1 \leq i \leq K-2\), are all the positive eigenvalues of \(\Sigma\), which is defined by (3.5) and (3.6), and the \(Z_i, 1 \leq i \leq K-2\), are i.i.d. standard normal random variables.

**Proof:** Let \(\lambda_i(\Sigma)\) denote the \(i\)th largest eigenvalue of the matrix \(\Sigma\). Then Lemma 3.6 asserts that
\[
\lambda_1(\Sigma) = \ldots = \lambda_{K-2}(\Sigma) = S_0^{-2} \quad \text{and} \quad \lambda_{K-1}(\Sigma) = \lambda_K(\Sigma) = 0.
\]

On the other hand, Lemma 3.4 implies that \(\Sigma\) is positive definite so \(\lambda_i(\Sigma) > 0\), for all \(1 \leq i \leq K\). Clearly, \(\Sigma\) and \(\Sigma\) are symmetric matrices. Hence, by Proposition 2.7,
\[ 0 < S_0^{-2} \lambda_K(S) \leq \lambda_j(SB) \leq S_0^{-2} \lambda_1(S) \text{ for } 1 \leq j \leq K-2, \quad (3.9) \]

and

\[ \lambda_{K-1}(SB) = \lambda_K(SB) = 0 \quad (3.10) \]

Combining Lemmas 3.3 and 3.5 with (3.9) and (3.10) proves the theorem.

This theorem shows that the asymptotic null distribution of \( N(1-R_N^2) \) is a weighted sum of \( K-2 \) i.i.d. \( \chi^2_1 \) distributions with the positive eigenvalues of \( SB \) as the weights. In contrast, the asymptotic alternative distribution of \( N(1-R_N^2) \) is a weighted sum of \( K-2 \) independent noncentral \( \chi^2_1 \) distributions.

To discuss this result more explicitly, let \( H_1 \) be the alternative hypothesis of interest:

\[ H_1: F(x) = F_1\left(\frac{x-F_1(\theta^*)}{\tau^*}\right) \text{ for some unknown } \theta^* \text{ and } \tau^* > 0, \]

where \( F_1 \) is a completely specified distribution function with density \( f_1 \). By reasoning similar to that for \( H_0 \), to examine the asymptotic distribution under \( H_1 \), we may assume \( F = F_1 \) and let

\[ G_1(y) = G(\theta^* + \tau^*y) \]

be the censoring distribution corresponding to \( F_1 \), where \( G \)
is the censoring distribution corresponding to the original \( F \).

Denote

\[
q_i^* = F_1^{-1}(p_i), \text{ for } 1 \leq i \leq K,
\]

\[
\overline{q}^* = \frac{1}{K} \sum_{i=1}^{K} q_i^* / K, \text{ and}
\]

\[
S_0^{*2} = \frac{1}{K} \sum_{i=1}^{K} (q_i^* - \overline{q}^*)^2.
\]

The weak convergence of Proposition 2.8 implies that when \( F = F_1 \), as \( N \to \infty \),

\[
Q_N \equiv \sqrt{N}(\hat{\Delta}_N - q_1^*, \ldots, \hat{\Delta}_N - q_K^*)^T \overset{L}{\to} N_K(\Delta^*, \Sigma^*)
\]

where \( \Delta^T = (q_1^* - q_1, \ldots, q_K^* - q_K) \) and \( \Sigma^* = (\sigma_{ij}^*)_{i,j=1}^{K} \) with elements

\[
\sigma_{ij}^* = \frac{1-p_i}{f_1(F_1^{-1}(p_i))} \cdot \frac{1-p_j}{f_1(F_1^{-1}(p_j))} \int_0^{p_i} \frac{du}{(1-u)^2 [1-G_1 \circ F_1^{-1}(u)]}
\]

\[
= \sigma_{ij}^*, \text{ for } 1 \leq i \leq j \leq K.
\]

Also let

\[
B^* = D / S_0^{*2} = \frac{[I_K - u \tilde{u}^T - \tilde{a} \tilde{a}^T]}{S_0^{*2}}.
\]

Then, the foregoing derivation of the asymptotic null distribution can be directly applied to the present case of the alternative hypothesis, which leads to the next theorem.
Theorem 3.8. Suppose that (A-1)-(A-5) hold with

$$0 < p_1 < p_2 < \ldots < p_K \leq A.$$  Then, under $H_1$, as $N \to \infty$,

$$
N(1-R_N^2) \sum_{i=1}^{K-2} \lambda_i^*[Z_i + (\tilde{\Gamma}_i^* \tilde{\Sigma}_i^{-\frac{1}{2}} \tilde{\Delta}_i)_1]\sum_{i=1}^{K-2} \lambda_i^*[Z_i + (\tilde{\Gamma}_i^* \tilde{\Sigma}_i^{-\frac{1}{2}} \tilde{\Delta}_i)_1]^2,
$$

where the $\lambda_i^*$, $1 \leq i \leq K-2$, are all the positive eigenvalues of $\Sigma^* \Sigma^*$ which is defined by (3.11) and (3.12), $\tilde{\Gamma}_i^*$ is an orthogonal matrix of order $K$ whose ith column vector is a standardized eigenvector corresponding to $\lambda_i^*$ with $\lambda_{K-1}^* = \lambda_K^* = 0$, $\tilde{\Delta}_1 = (q_1^*-q_1, \ldots, q_K^*-q_K)$, $(\tilde{\Gamma}_i^* \tilde{\Sigma}_i^{-\frac{1}{2}} \tilde{\Delta}_i)_1$ is the ith component of the vector $\tilde{\Gamma}_i^* \tilde{\Sigma}_i^{-\frac{1}{2}} \tilde{\Delta}_i$, and the $Z_i$, $1 \leq i \leq K-2$, are i.i.d. standard normal random variables.
IV. Correlation-Type Test for Distribution with Origin Known

In practice, the lifetime variable is positive and starts from zero. For instance, many well-known distributions in survival analysis such as the exponential, gamma, and Weibull distributions have this feature.

Using formal language, the origin \( x_0 \) of a distribution \( H \) is defined to be the supremum of \( \{ x : H(x) = 0 \} \). The origin certainly is a location parameter. In fact, when the origin of the lifetime distribution \( F \) is known beforehand, it is not appropriate to employ the test for location-scale families, which was discussed in the previous section.

Without loss of generality, we may assume that the origin of \( F \) is known to be zero. The composite null hypothesis of goodness-of-fit for this case is

\[
H_0 : F(x) = F_0(x/\tau) \text{ for some unknown } \tau > 0,
\]

where \( F_0 \) is a completely specified distribution function with origin at 0 and density \( f_0 \).

4.1. Test Statistic.

Since the origins of \( F \) and \( F_0 \) are all zero, in a Q-Q plot the curve connecting all \( K \) pairs of empirical values \( \hat{\gamma}_{N_1} = \hat{F}_N^{-1}(p_1) \) and hypothesized values \( q_1 = F_0^{-1}(p_1) \) should always pass through the origin of the two-dimensional plane. Therefore, before proceeding to define a suitable test for this
case, we shall review the basic theory of linear regression through the origin.

In the present problem, the linear regression model of interest is

$$\hat{Q}_{Ni} = b q_i + \text{(error)}_i, \text{ for } 1 \leq i \leq K.$$  

The least-squares estimator for $b$ is

$$\hat{b} = \sum_{i=1}^{K} \hat{Q}_{Ni} q_i / \sum_{i=1}^{K} q_i^2.$$  

The sums of squares can be partitioned as follows: the total sum of squares is

$$\text{SST} = \sum_{i=1}^{K} \hat{Q}_{Ni}^2,$$  

the error sum of squares is

$$\text{SSE} = \sum_{i=1}^{K} [(\text{observed})_i - (\text{estimated})_i]^2$$  

$$= \sum_{i=1}^{K} (\hat{Q}_{Ni} - \hat{b} q_i)^2$$  

$$= \sum_{i=1}^{K} \hat{Q}_{Ni}^2 - \hat{b}^2 \sum_{i=1}^{K} q_i^2,$$  

and the regression sum of squares is
SSR = SST - SSE

\[ = \hat{b}^2 \sum_{i=1}^{K} q_i^2 \]

\[ = \left( \sum_{i=1}^{K} \hat{Q}_{Ni} q_i \right)^2 / \sum_{i=1}^{K} q_i^2 . \]

From the analysis of variance approach, the coefficient of determination in regression terminology is defined by

\[ R_N^2 = \frac{SSR}{SST} \]

\[ = \frac{\left( \sum_{i=1}^{K} \hat{Q}_{Ni} q_i \right)^2}{\left( \sum_{i=1}^{K} \hat{Q}_{Ni}^2 \right) \left( \sum_{i=1}^{K} q_i^2 \right)} \]

(4.1)

\[ = \frac{\left( \sum_{i=1}^{K} \hat{Q}_{Ni} q_i \right)^2}{t_N^2 t_0^2} , \]

where

\[ t_N^2 = \sum_{i=1}^{K} \hat{Q}_{Ni}^2 , \text{ and } \]

\[ t_0^2 = \sum_{i=1}^{K} q_i^2 . \] (4.2)

The statistic (4.1) can be viewed as an analogue of the Shapiro-Francia statistic. The test rejects \( H_0 \) for significantly small \( R_N^2 \).
The equivalent statistic \( N(1-\bar{R}_N^2) \) can be shown to be a quadratic form. Writing

\[
t^2_{N+2} = \left( \sum_{i=1}^{K} \hat{Q}_{Ni} q_i \right)^2
\]

\[
= \left[ \sum_{i=1}^{K} (\hat{Q}_{Ni} - q_i) q_i + \sum_{i=1}^{K} q_i^2 \right]^2,
\]

after algebraic simplification we obtain

\[
t^2_{N+2} (1-\bar{R}_N^2) = t^2_0 \left[ \sum_{i=1}^{K} (\hat{Q}_{Ni} - q_i)^2 (1-h_i^2) \right.
\]

\[
- \sum_{1 \leq i \neq j \leq K} (\hat{Q}_{Ni} - q_i)(\hat{Q}_{Nj} - q_j) h_i h_j \left],
\]

where

\[
h_i = q_i / t_0, \quad 1 \leq i \leq K. \tag{4.3}
\]

Thus, we have the following:

**Lemma 4.1.**

\[
N(1-\bar{R}_N^2) = Q^T_{\bar{N}} D_0 Q_{\bar{N}} / t^2_N ,
\]

where \( Q^T_{\bar{N}} = \sqrt{N}(\hat{Q}_{N1} - q_1, \ldots, \hat{Q}_{NK} - q_K) \) is a \( K \)-dimensional row vector, and

\[
D_0 = I_K - h h^T
\]

with \( h^T = (h_1, \ldots, h_K) \) and \( h_i, 1 \leq i \leq K \), defined in (4.3).
Therefore, \( N(1-R_N^2) \) is taken to be the test statistic and the test rejects \( H_0 \) for significantly large values of this statistic.

4.2. Asymptotic Distributions of the Test Statistic.

The same process of reasoning as discussed in Section 3.2 results in theorems parallel to Theorems 3.7 and 3.8.

These parallel theorems will be exhibited after we introduce some notation. Let

\[
G_0(y) = G(\tau y) ,
\]

where \( G \) is the censoring distribution corresponding to the original \( F(x) = F_0(x/\tau) \); \( \Sigma_0 = (c_{ij})_{i,j=1}^{K} \) with elements

\[
c_{ij} = \frac{1-p_i}{f_0(F_0^{-1}(p_i))} \cdot \frac{1-p_j}{f_0(F_0^{-1}(p_j))} \int_0^{p_i} \frac{du}{(1-u)^2[1-G_0 \circ F_0^{-1}(u)]} ,
\]

\[
= c_{ji} , \quad \text{for } 1 \leq i \leq j \leq K ;
\]

and

\[
\tau_0 = \frac{D_0}{t_0^2} = [I_{2K} - h_{2K}^T]/t_0^2 ,
\]

where \( t_0^2 \) and \( h_{2K}^T = (h_1, \ldots, h_K) \) are defined in (4.2) and (4.3), respectively. Next, notice that \( \tau_0 \) has eigenvalues \( t_0^2 \) of multiplicity \( (K-1) \) and 0 of multiplicity 1.
The theorem concerning the asymptotic null distribution of $N(1-R^2_N)$ can be stated as follows.

**Theorem 4.2.** Suppose that (A-1)-(A-5) hold for $x_0 = 0$ and $0 < p_1 < p_2 < \ldots < p_K \leq A$. Then, under $H_0$,

$$N(1-R^2_N) \overset{d}{\rightarrow} \sum_{i=1}^{K-1} \tilde{\lambda}_i Z_i^2, \text{ as } N \rightarrow \infty,$$

where the $\tilde{\lambda}_i$, $1 \leq i \leq K-1$, are all the positive eigenvalues of $\sum_{0=0}^{B}$ which is defined by (4.5) and (4.6), and the $Z_i$, $1 \leq i \leq K-1$, are i.i.d. standard normal random variables.

On the other hand, one may be interested in finding the asymptotic distribution under an alternative hypothesis:

$$H_1 : F(x) = F_1(x/\tau^*) \text{ for some unknown } \tau^* > 0,$$

where $F_1$ is a completely specified distribution function with origin at 0 and density $f_1$.

Denote

$$G_1(y) = G(\tau^* y);$$

$$\tau^*_0 = (c^*_{ij})_{i,j=1}^K \text{ with elements}$$

$$c^*_{ij} = \frac{1-p_i}{f_1(F_1^{-1}(p_i))} \cdot \frac{1-p_j}{f_1(F_1^{-1}(p_j))} \int_0^{\lambda_i} \frac{du}{(1-u)^2[1-G_1^{-1}(u)]},$$

$$= c^*_{ji}, \text{ for } 1 \leq i \leq j \leq K;$$

30
and

$$B_0^* = [I_K - h\bar{h}_h^T]/t_0^2,$$  \hspace{1cm} (4.8)

where $h^T = (h_1, \ldots, h_K)$ is defined in (4.3) and

$$t_0^2 = \sum_{i=1}^K q_i^2 \quad \text{with} \quad q_i^* = F_1^{-1}(p_i), \quad 1 \leq i \leq K.$$  

Then, we are able to describe the asymptotic alternative distribution of $N(1-R_N^2)$.

Theorem 4.3. Suppose that (A-1)-(A-5) hold for $x_0 = 0$ and $0 < p_1 < p_2 < \ldots < p_K < A$. Then, under $H_1$,

$$N(1-R_N^2) \xrightarrow{L} \sum_{i=1}^{K-1} \lambda_i^*[Z_i + (\Gamma_i T_i^* \Gamma_i^* - \bar{h}_i^2 \Delta_i)]^2,$$ as $N \to \infty$,

where the $\lambda_i^*$, $1 \leq i \leq K-1$, are all the positive eigenvalues of $\Sigma_0^* B_0^*$ which is defined by (4.7) and (4.8), $\Gamma_i^*$ is an orthogonal matrix of order $K$ whose $i$th column vector is a standardized eigenvector corresponding to $\lambda_i^*$ with $\lambda_K^* = 0$, $\Delta_i^T = (q_i^* - q_1^*, \ldots, q_i^* - q_K^*)$, $(\Gamma_i T_i^* \Gamma_i^* - \bar{h}_i^2 \Delta_i)_i$ is the $i$th component of the vector $\Gamma_0 T_0^* \Gamma_0^* - \bar{h}_0^2 \Delta_0$, and the $Z_i$, $1 \leq i \leq K-1$, are i.i.d. standard normal random variables.
V. Correlation-Type Tests for Exponentiality

Since the exponential distribution is a useful, as well as the simplest survival time model, it is customary to first try fitting this distribution to survival data. From now on, we shall concentrate on the testing of exponentiality with the results derived in the previous sections.

The correlation-type tests are not ready for direct application to practical situations, since the unknown parameter(s) τ (and θ) of the lifetime distribution F and the nuisance censoring distribution G are involved in the asymptotic null distributions of the statistics. Some consistent estimators for them could be used, but the test procedures would then become difficult to handle. Hopefully, a few additional assumptions, which are not too restrictive, may be made for computational feasibility.

In the special case of testing exponentiality, we continue following all the notation used for the general case. Furthermore, for testing exponentiality, we assume that

(B-1) F and G have the same origin (location) parameter.

The null hypothesis of one-parameter and two-parameter exponential distributions are expressed, respectively, as

\[ H_0 : F(x) = 1 - e^{-x/\tau} \quad \text{for} \quad x \geq 0 , \]

and
\[ H_0: F(x) = 1 - e^{-(x-\theta)/\tau} \quad \text{for} \quad x \geq \theta. \]

The tests illustrated later are all based on \( K = 6 \) and \( \{p_i\} = \{0.2, 0.35, 0.45, 0.55, 0.65, 0.8\} \). In consideration of economy of computational labor and applicability of the tests to small samples, we have not selected a larger value of \( K \). The choice of the specified proportions \( \{p_i\} \) results from indifference to the extreme tails and attention to the central part of the lifetime distribution. Miller (1981) showed that, under the random censorship model, the asymptotic efficiency of the Kaplan-Meier estimator relative to the maximum likelihood estimator of a parameter lifetime distribution decreases as the censoring proportion increases or the surviving fraction tends to zero or one. According to this finding, our choice of \( \{p_i\} \) seemingly makes sense.

5.1. Asymptotic Robustness of the Tests in the Case of Light Censoring.

Light censoring means that the proportion of censored observations in a sample is less than \( 1/3 \).

Theorems 4.2 and 3.7 represent the asymptotic null distributions for testing \( H_0 \) in the one and two parameter cases, respectively. Since the censoring distribution \( G_0 \) appears in the matrices \( \Sigma_0 \) and \( \Sigma \), it would be interesting to examine how the censoring distribution affects the inference. The following distributions whose densities are quite distinct are considered as censoring distributions in the comparisons.
<table>
<thead>
<tr>
<th>Name</th>
<th>Abbreviation</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) exponential distribution with scale parameter 1/β</td>
<td>Exp(β)</td>
<td>$g_0(y) = \beta e^{-\beta y}$, for $y \geq 0$.</td>
</tr>
<tr>
<td>(ii) gamma distribution with scale parameter 1/λ and shape parameter 2</td>
<td>Gamma (2,λ)</td>
<td>$g_0(y) = \lambda^2 y e^{-\lambda y}$, for $y \geq 0$.</td>
</tr>
<tr>
<td>(iii) uniform distribution on interval [0, U]</td>
<td>Unif [0, U]</td>
<td>$g_0(y) = 1/U$, for $y \in [0, U]$</td>
</tr>
<tr>
<td>(iv) triangular distribution on interval [0, T]</td>
<td>Triang [0, T]</td>
<td>$g_0(y) = 2y/T^2$, for $y \in [0, T]$</td>
</tr>
</tbody>
</table>

Since $F_0$ is the standard exponential distribution, i.e., $\text{Exp}(1)$, the elements $c_{ij}$ of matrix $\Sigma_0$ specified in (4.5) are of the following forms:

$$c_{ij} = \int_0^{p_i} \frac{du}{(1-u)^2[1-G_0(-\ln(1-u))]} = c_{ji}, \quad (5.1)$$

for $1 \leq i \leq j \leq K$, where $G_0$ is defined in (4.4). Similarly, the elements $\sigma_{ij}$ of matrix $\Sigma$ given by (3.5) become

$$\sigma_{ij} = \int_0^{p_i} \frac{du}{(1-u)^2[1-G_0(-\ln(1-u))]} = \sigma_{ji}, \quad (5.2)$$

for $1 \leq i \leq j \leq K$, where $G_0$ is defined in (3.4).

When (5.1) and (5.2) can not be evaluated in closed form, an adaptive, iterative procedure of Simpson's rule is used in the numerical integration. A FORTRAN program for this algorithm is referred to in Shampine and Allen (1973).

The asymptotic null distribution of $N(1-R_N^2)$ or $N(1-R_N^2)$ requires the determination of the eigenvalues of a real matrix.
The path of subroutines BALANC, ORTHES and HQR in the Eigen-system Subroutine Package (EISPACK) by Smith et al. (1976) was used in our computation.

The expected censoring proportion, abbreviated by e.c.p. hereafter, is

$$\text{e.c.p.} = \int_{-\infty}^{\infty} [1-F(x)] \ dG(x)$$

$$= \int_{0}^{\infty} [1-F_0(x)] \ dG_0(x) .$$

The second equality follows from Assumption (B-1). Since $F_0 = \text{Exp}(1)$, the e.c.p. pertaining to each of the four selected censoring distributions $G_0$ is

(i) e.c.p. $= \frac{\beta}{1+\beta}$,

(ii) e.c.p. $= \left(\frac{\lambda}{1+\lambda}\right)^2$,

(iii) e.c.p. $= \frac{1}{U} (1-e^{-U})$, and

(iv) e.c.p. $= \frac{2}{T^2} [1 - (1+T) e^{-T}]$, respectively.

For e.c.p. = 10%, 20%, 30%, and 40%, the corresponding asymptotic null distributions of $N(1-R_N^2)$ are shown with the weights $\tilde{\lambda}_i$, $1 \leq i \leq 5$, in their weighted chi-squares representations in Table 1. The corresponding results to $N(1-R_N^2)$ are presented in Table 2.

In both tables, the triangular censoring distribution is excluded in the case e.c.p. = 40% since, in this case, the
Table 1. Weights in the Weighted Chi-Squares Sum Representation of the Asymptotic Null Distribution of $N(1-R^2_N)$ for Testing Exponentiality with the Origin at 0 when $K = 6$, $p_i = 0.2, 0.35, 0.45, 0.55, 0.65, 0.8$

(A) Expected Censoring Proportion = 10%

<table>
<thead>
<tr>
<th>Censoring Distribution</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp(1/9)</td>
<td>0.3380 0.1106 0.0489 0.0279 0.0176</td>
</tr>
<tr>
<td>Gamma (2, 0.46248)</td>
<td>0.3335 0.1079 0.0475 0.0271 0.0171</td>
</tr>
<tr>
<td>Unif [0, 9.9996]</td>
<td>0.3366 0.1101 0.0487 0.0278 0.0175</td>
</tr>
<tr>
<td>Triag [0, 4.3095]</td>
<td>0.3251 0.1058 0.0468 0.0269 0.0170</td>
</tr>
</tbody>
</table>

(B) Expected Censoring Proportion = 20%

<table>
<thead>
<tr>
<th>Censoring Distribution</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp(1/4)</td>
<td>0.3907 0.1232 0.0536 0.0299 0.0186</td>
</tr>
<tr>
<td>Gamma (2, 0.80902)</td>
<td>0.5919 0.1189 0.0513 0.0284 0.0177</td>
</tr>
<tr>
<td>Unif [0, 4.9651]</td>
<td>0.3841 0.1207 0.0526 0.0294 0.0183</td>
</tr>
<tr>
<td>Triag [0, 2.7609]</td>
<td>0.3693 0.1128 0.0489 0.0275 0.0173</td>
</tr>
</tbody>
</table>

(C) Expected Censoring Proportion = 30%

<table>
<thead>
<tr>
<th>Censoring Distribution</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp(3/7)</td>
<td>0.4749 0.1414 0.0605 0.0326 0.0200</td>
</tr>
<tr>
<td>Gamma (2, 1.2110)</td>
<td>0.5047 0.1376 0.0576 0.0305 0.0186</td>
</tr>
<tr>
<td>Unif [0, 3.1970]</td>
<td>0.4627 0.1354 0.0578 0.0313 0.0193</td>
</tr>
<tr>
<td>Triag [0, 1.9816]</td>
<td>0.5035 0.1266 0.0529 0.0287 0.0177</td>
</tr>
</tbody>
</table>

(D) Expected Censoring Proportion = 40%

<table>
<thead>
<tr>
<th>Censoring Distribution</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp(2/3)</td>
<td>0.6240 0.1703 0.0710 0.0365 0.0218</td>
</tr>
<tr>
<td>Gamma (2, 1.7208)</td>
<td>0.7535 0.1719 0.0687 0.0341 0.0200</td>
</tr>
<tr>
<td>Unif [0, 2.2317]</td>
<td>0.6479 0.1600 0.0659 0.0341 0.0205</td>
</tr>
</tbody>
</table>
Table 2. Weights in the Weighted Chi-Squares Sum Representation of the Asymptotic Null Distribution of \( N(1-R_N^2) \) for Testing Exponentiality with the Origin & Scale Unknown When \( K = 6, p_i = 0.2, 0.35, 0.45, 0.55, 0.65, 0.8 \)

(A) Expected CensoringProportion = 10%

<table>
<thead>
<tr>
<th>Censoring Distribution</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp(1/9)</td>
<td>0.7653 0.2738 0.1394 0.0763</td>
</tr>
<tr>
<td>Gamma (2, 0.46248)</td>
<td>0.7538 0.2658 0.1351 0.0740</td>
</tr>
<tr>
<td>Unif [0, 9.9996]</td>
<td>0.7618 0.2724 0.1387 0.0760</td>
</tr>
<tr>
<td>Triag [0, 4.3095]</td>
<td>0.7322 0.2606 0.1331 0.0733</td>
</tr>
</tbody>
</table>

(B) Expected CensoringProportion = 20%

<table>
<thead>
<tr>
<th>Censoring Distribution</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp(1/4)</td>
<td>0.8986 0.3056 0.1526 0.0817</td>
</tr>
<tr>
<td>Gamma (2, 0.80902)</td>
<td>0.9007 0.2924 0.1448 0.0772</td>
</tr>
<tr>
<td>Unif [0, 4.9651]</td>
<td>0.8816 0.2988 0.1494 0.0803</td>
</tr>
<tr>
<td>Triag [0, 2.7606]</td>
<td>0.8432 0.2760 0.1384 0.0749</td>
</tr>
</tbody>
</table>

(C) Expected CensoringProportion = 30%

<table>
<thead>
<tr>
<th>Censoring Distribution</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp(3/7)</td>
<td>1.1095 0.3525 0.1714 0.0892</td>
</tr>
<tr>
<td>Gamma (2, 1.2110)</td>
<td>1.1811 0.3393 0.1611 0.0825</td>
</tr>
<tr>
<td>Unif [0, 3.1970]</td>
<td>1.0779 0.3353 0.1634 0.0855</td>
</tr>
<tr>
<td>Triag [0, 1.9816]</td>
<td>1.1729 0.3061 0.1480 0.0775</td>
</tr>
</tbody>
</table>

(D) Expected CensoringProportion = 40%

<table>
<thead>
<tr>
<th>Censoring Distribution</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp(2/3)</td>
<td>1.4800 0.4276 0.1999 0.1000</td>
</tr>
<tr>
<td>Gamma (2, 1.7208)</td>
<td>1.7907 0.4282 0.1897 0.0915</td>
</tr>
<tr>
<td>Unif [0, 2.2317]</td>
<td>1.5324 0.3961 0.1846 0.0928</td>
</tr>
</tbody>
</table>
support of triangular distribution is [0, 1.4696] and

\[ F_0^{-1}(p_K) = F_0^{-1}(0.8) \approx 1.609 > 1.4696, \]

which violates Assumption (A-3).

It is observed in Tables 1 and 2 that for a fixed e.c.p. \( \leq 40\% \) the weights are very close to one another for different censoring distributions. Some other calculations show that the weights for different censoring distributions are more dispersed when the e.c.p. is larger than 40%. The weights for various other sets of \( K \) and \( \{p_i, 1 \leq i \leq K\} \) exhibit similar behavior with respect to different censoring distributions for moderate values of e.c.p. pertaining to the exponential lifetime distribution. Therefore, in the presence of light censoring, it is permissible for us to postulate that \( \text{Exp}(\beta) \) is the underlying censoring distribution in the correlation-type tests for exponentiality.

Koziol and Green (1976) introduced the following assumption on the censoring distribution:

\[ (1-G) = (1-F)^\beta, \text{ with } \beta \text{ a positive constant}. \]

They interpreted \( \beta \) as the censoring parameter. This is called the Koziol-Green model of random censorship by Csörgö and Horváth (1981). Our calculations indicate that, in the case of light censoring, the correlation-type tests for exponentiality are
asymptotically robust against departures from the Koziol-Green model of random censorship. For this reason and computational convenience, we assume throughout the remainder of this paper that the Koziol-Green model of random censorship holds.

5.2. Tests Developed with the Koziol-Green Model of Random Censorship.

First, we consider the testing of exponentiality with the origin at 0. Under the Koziol-Green model of random censorship,

\[ \text{e.c.p.} = \Pr(\delta_i = 0) = \frac{\beta}{1+\beta}. \]

For a given e.c.p. = \( \varepsilon \), \( \beta = \frac{\varepsilon}{1-\varepsilon} \), and (5.1) becomes

\[ c_{ij} = (1-\varepsilon) \left[ (1-p_i)^{-(1-\varepsilon)^{-1}} - 1 \right] = c_{ji}, \text{ for } 1 \leq i \leq j \leq K. \]

(5.3)

Then, from Theorem 4.2 the quantiles of the asymptotic null distribution of \( N(1-R_N^2) \) can be computed.

For a fixed value of e.c.p., the asymptotic null upper \( \alpha \) quantiles of \( N(1-R_N^2) \) were computed by Monte Carlo samples. In fact, we generated 40,000 random samples each of five standard normal random variables \( z_i \) to compute \( \sum_{i=1}^{5} \tilde{\lambda}_i z_i^2 \), where \( \tilde{\lambda}_i, 1 \leq i \leq 5 \), are all the positive eigenvalues of \( \Sigma_0 B_0 \) and \( \Sigma_0 \) is evaluated from (5.3) for the fixed e.c.p. The uniform random number generator used in the simulation is an algorithm in Knuth (1981). The normal random variables were then generated by the polar method.
In Table 3 the asymptotic null upper 0.10, 0.05, 0.01, 0.005 and 0.001 quantiles of \( N(1-R_N^2) \) are tabulated with respect to various values of e.c.p. from 0 to 50% with 5% increments. For comparison, in the same table, upper quantiles of the exact finite-sample null distributions of \( N(1-R_N^2) \) for \( N = 30 \) and 100 are also included.

Table 4 gives the corresponding results for the statistic \( N(1-R_N^2) \).

The finite-sample simulations were based on 40,000 generations of lifetime and censoring samples for each fixed value of e.c.p. In these simulations those Monte Carlo samples which contained less than two uncensored observations were excluded. Furthermore, in the construction of Table 4, those Monte Carlo samples which produced all equal \( \hat{Q}_{Ni} \), \( 1 \leq i \leq 6 \), so that \( S_N = 0 \) were also excluded.

The reason for not extending Tables 3 and 4 to values of e.c.p. larger than 50% is that the asymptotic robustness of the tests holds only for light censoring. Also, with heavily censored data one cannot effectively test goodness of fit unless the sample size is tremendously large. Fortunately, heavily censored data are not often encountered in practice.

Roughly speaking, in Tables 3 and 4 with light censoring, the asymptotic null upper \( \alpha \) quantiles corresponding to \( \alpha = .10 \) and .05 are fairly good approximations to the upper quantiles of the finite-sample \( N = 100 \) null distributions.
Table 3. Upper Quantiles of the Null Distribution of \( N(1-R^2_n) \) for Testing Exponentiality with the Origin at 0 when \( k = 6, p_i = 0.2, 0.35, 0.45, 0.55, 0.65, 0.8, \) Based on 40,000 Random Sample Simulations

<table>
<thead>
<tr>
<th>Expected Censoring Proportion</th>
<th>0</th>
<th>5%</th>
<th>10%</th>
<th>15%</th>
<th>20%</th>
<th>25%</th>
<th>30%</th>
<th>35%</th>
<th>40%</th>
<th>45%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper Percentage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finite Sample Distribution</td>
<td>.10</td>
<td>0.96</td>
<td>1.01</td>
<td>1.07</td>
<td>1.16</td>
<td>1.24</td>
<td>1.37</td>
<td>1.53</td>
<td>1.74</td>
<td>1.93</td>
<td>2.16</td>
</tr>
<tr>
<td></td>
<td>.05</td>
<td>1.22</td>
<td>1.31</td>
<td>1.59</td>
<td>1.63</td>
<td>1.84</td>
<td>2.08</td>
<td>2.38</td>
<td>2.54</td>
<td>2.79</td>
<td>2.99</td>
</tr>
<tr>
<td></td>
<td>.01</td>
<td>1.84</td>
<td>2.10</td>
<td>2.26</td>
<td>2.48</td>
<td>2.92</td>
<td>3.28</td>
<td>3.75</td>
<td>4.19</td>
<td>4.32</td>
<td>4.48</td>
</tr>
<tr>
<td>( N = 30 )</td>
<td>.005</td>
<td>2.11</td>
<td>2.51</td>
<td>2.71</td>
<td>3.01</td>
<td>3.54</td>
<td>4.06</td>
<td>4.62</td>
<td>5.02</td>
<td>5.06</td>
<td>5.19</td>
</tr>
<tr>
<td></td>
<td>.001</td>
<td>2.79</td>
<td>3.66</td>
<td>4.01</td>
<td>4.54</td>
<td>4.98</td>
<td>6.00</td>
<td>6.66</td>
<td>6.73</td>
<td>6.89</td>
<td>6.46</td>
</tr>
<tr>
<td></td>
<td>.10</td>
<td>1.01</td>
<td>1.08</td>
<td>1.14</td>
<td>1.21</td>
<td>1.30</td>
<td>1.41</td>
<td>1.58</td>
<td>1.79</td>
<td>2.11</td>
<td>2.53</td>
</tr>
<tr>
<td></td>
<td>.05</td>
<td>1.32</td>
<td>1.42</td>
<td>1.51</td>
<td>1.61</td>
<td>1.74</td>
<td>1.86</td>
<td>2.08</td>
<td>2.42</td>
<td>2.89</td>
<td>3.64</td>
</tr>
<tr>
<td></td>
<td>.01</td>
<td>2.11</td>
<td>2.24</td>
<td>2.46</td>
<td>2.57</td>
<td>2.87</td>
<td>3.10</td>
<td>3.50</td>
<td>4.37</td>
<td>5.76</td>
<td>8.26</td>
</tr>
<tr>
<td>( N = 100 )</td>
<td>.005</td>
<td>2.46</td>
<td>2.66</td>
<td>2.86</td>
<td>3.08</td>
<td>3.43</td>
<td>3.62</td>
<td>4.25</td>
<td>5.53</td>
<td>7.86</td>
<td>10.75</td>
</tr>
<tr>
<td></td>
<td>.001</td>
<td>3.35</td>
<td>3.45</td>
<td>3.91</td>
<td>4.07</td>
<td>4.66</td>
<td>5.20</td>
<td>6.49</td>
<td>9.81</td>
<td>14.23</td>
<td>15.91</td>
</tr>
<tr>
<td>Asymptotic Distribution</td>
<td>.10</td>
<td>1.06</td>
<td>1.10</td>
<td>1.17</td>
<td>1.25</td>
<td>1.33</td>
<td>1.45</td>
<td>1.59</td>
<td>1.78</td>
<td>2.06</td>
<td>2.43</td>
</tr>
<tr>
<td></td>
<td>.05</td>
<td>1.39</td>
<td>1.43</td>
<td>1.56</td>
<td>1.66</td>
<td>1.77</td>
<td>1.94</td>
<td>2.12</td>
<td>2.39</td>
<td>2.77</td>
<td>3.25</td>
</tr>
<tr>
<td></td>
<td>.01</td>
<td>2.27</td>
<td>2.34</td>
<td>2.53</td>
<td>2.66</td>
<td>2.88</td>
<td>3.15</td>
<td>3.44</td>
<td>3.87</td>
<td>4.49</td>
<td>5.38</td>
</tr>
<tr>
<td></td>
<td>.005</td>
<td>2.66</td>
<td>2.74</td>
<td>2.96</td>
<td>3.11</td>
<td>3.43</td>
<td>3.69</td>
<td>4.01</td>
<td>4.61</td>
<td>5.27</td>
<td>6.32</td>
</tr>
<tr>
<td></td>
<td>.001</td>
<td>3.52</td>
<td>3.55</td>
<td>4.02</td>
<td>4.10</td>
<td>4.74</td>
<td>4.94</td>
<td>5.35</td>
<td>6.25</td>
<td>7.13</td>
<td>8.66</td>
</tr>
<tr>
<td>Expected Censoring Proportion</td>
<td>0</td>
<td>5%</td>
<td>10%</td>
<td>15%</td>
<td>20%</td>
<td>25%</td>
<td>30%</td>
<td>35%</td>
<td>40%</td>
<td>45%</td>
<td>50%</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Upper Percentage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finite Sample Distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.10</td>
<td>2.46</td>
<td>2.40</td>
<td>2.52</td>
<td>2.70</td>
<td>2.90</td>
<td>3.18</td>
<td>3.59</td>
<td>4.04</td>
<td>4.59</td>
<td>5.32</td>
<td>6.26</td>
</tr>
<tr>
<td>.05</td>
<td>3.15</td>
<td>3.04</td>
<td>3.20</td>
<td>3.40</td>
<td>3.66</td>
<td>4.02</td>
<td>4.49</td>
<td>5.03</td>
<td>5.62</td>
<td>6.63</td>
<td>8.46</td>
</tr>
<tr>
<td>.01</td>
<td>4.72</td>
<td>4.51</td>
<td>4.71</td>
<td>5.04</td>
<td>5.34</td>
<td>5.88</td>
<td>6.46</td>
<td>7.15</td>
<td>8.32</td>
<td>9.62</td>
<td>11.92</td>
</tr>
<tr>
<td>N = 30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.005</td>
<td>5.27</td>
<td>5.17</td>
<td>5.37</td>
<td>5.67</td>
<td>5.95</td>
<td>6.62</td>
<td>7.28</td>
<td>8.09</td>
<td>9.21</td>
<td>11.53</td>
<td>14.10</td>
</tr>
<tr>
<td>.001</td>
<td>6.66</td>
<td>6.44</td>
<td>6.67</td>
<td>7.33</td>
<td>7.79</td>
<td>8.43</td>
<td>9.10</td>
<td>10.08</td>
<td>12.40</td>
<td>14.40</td>
<td>15.96</td>
</tr>
<tr>
<td>Finite Sample Distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.10</td>
<td>2.34</td>
<td>2.52</td>
<td>2.64</td>
<td>2.82</td>
<td>3.04</td>
<td>3.35</td>
<td>3.71</td>
<td>4.23</td>
<td>4.92</td>
<td>6.06</td>
<td>7.51</td>
</tr>
<tr>
<td>.05</td>
<td>3.06</td>
<td>3.29</td>
<td>3.41</td>
<td>3.69</td>
<td>3.98</td>
<td>4.39</td>
<td>4.80</td>
<td>5.50</td>
<td>6.49</td>
<td>8.04</td>
<td>9.93</td>
</tr>
<tr>
<td>.01</td>
<td>4.71</td>
<td>5.02</td>
<td>5.25</td>
<td>5.62</td>
<td>6.14</td>
<td>6.77</td>
<td>7.48</td>
<td>8.65</td>
<td>10.27</td>
<td>12.47</td>
<td>14.81</td>
</tr>
<tr>
<td>N = 100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.005</td>
<td>5.46</td>
<td>5.69</td>
<td>5.98</td>
<td>6.51</td>
<td>7.09</td>
<td>7.68</td>
<td>8.73</td>
<td>10.20</td>
<td>11.99</td>
<td>14.36</td>
<td>16.66</td>
</tr>
<tr>
<td>.001</td>
<td>7.27</td>
<td>7.60</td>
<td>7.66</td>
<td>8.59</td>
<td>9.12</td>
<td>9.94</td>
<td>11.53</td>
<td>12.85</td>
<td>15.72</td>
<td>18.90</td>
<td>20.55</td>
</tr>
<tr>
<td>Asymptotic Distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.10</td>
<td>2.40</td>
<td>2.52</td>
<td>2.71</td>
<td>2.91</td>
<td>3.10</td>
<td>3.41</td>
<td>3.75</td>
<td>4.27</td>
<td>4.90</td>
<td>5.79</td>
<td>7.06</td>
</tr>
<tr>
<td>.05</td>
<td>3.17</td>
<td>3.35</td>
<td>3.56</td>
<td>3.83</td>
<td>4.12</td>
<td>4.54</td>
<td>4.98</td>
<td>5.71</td>
<td>6.54</td>
<td>7.77</td>
<td>9.56</td>
</tr>
<tr>
<td>.01</td>
<td>5.03</td>
<td>5.43</td>
<td>5.71</td>
<td>6.11</td>
<td>6.62</td>
<td>7.30</td>
<td>8.04</td>
<td>9.36</td>
<td>10.68</td>
<td>12.54</td>
<td>15.64</td>
</tr>
<tr>
<td>.005</td>
<td>5.88</td>
<td>6.33</td>
<td>6.64</td>
<td>7.26</td>
<td>7.83</td>
<td>8.40</td>
<td>9.41</td>
<td>10.78</td>
<td>12.73</td>
<td>14.70</td>
<td>18.31</td>
</tr>
<tr>
<td>.001</td>
<td>7.84</td>
<td>8.39</td>
<td>8.76</td>
<td>9.73</td>
<td>10.52</td>
<td>11.45</td>
<td>12.37</td>
<td>15.10</td>
<td>18.15</td>
<td>20.10</td>
<td>24.84</td>
</tr>
</tbody>
</table>

Table 4. Upper Quantiles of the Null Distribution of $N(1-R^2_N)$ for Testing Exponentiality with the Origin and Scale Unknown When $k = 6$, $p_i = 0.2$, 0.35, 0.45, 0.55, 0.65, 0.8, Based on 40,000 Random Sample Simulations.
In Table 3 for e.c.p. ≥ 35% , the upper quantiles of the finite-sample N = 100 null distribution have larger discrepancies from the asymptotic null upper quantiles than those of the finite-sample N = 30 null distribution. However, the finite-sample null upper quantiles do converge to the asymptotic ones for larger N. An illustration is given in Table 5 for the statistic N(1−RN 2) with e.c.p. = 40%.

To implement the test for H0 , the e.c.p. must be known. One can use

\[ P_c = \frac{N}{\sum_{i=1}^{N} (1-\delta_i)/N} , \]

the proportion of censored observations in the sample, as a consistent estimator of the e.c.p. Substituting Pc for the e.c.p., we can obtain the approximate values of the null upper quantiles of N(1−RN 2). From Table 3, the quantiles of the null distribution are not very sensitive to the e.c.p. for light censoring. Hence, any inaccuracy in substituting Pc for the e.c.p. may not matter much. The test rejects H0 at the nominal α-level if the test statistic is larger than the approximate null upper α quantile.

Similar logic holds for testing H0 with the location-scale family of exponential distributions.

Intuitively, the tests become powerful as K increases. However, for the asymptotic theory to prevail, the sample size N should be substantially larger than K. One may want to use different sets of K and \( \{p_i, 1 \leq i \leq K\} \) in test procedures other than the one we chose.
Table 5. Upper Quantiles of the Null Distribution of $N(1-R_N^2)$ for Testing Exponentiality with the Origin at 0. When $k = 6$, $p_i = 0.2, 0.35, 0.45, 0.55, 0.65, 0.8$, Based on 40,000 Random Sample Simulations

<table>
<thead>
<tr>
<th>Upper Percentage</th>
<th>30</th>
<th>100</th>
<th>200</th>
<th>500</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.10</td>
<td>1.93</td>
<td>2.11</td>
<td>2.04</td>
<td>2.06</td>
<td>2.06</td>
</tr>
<tr>
<td>.05</td>
<td>2.54</td>
<td>2.89</td>
<td>2.77</td>
<td>2.75</td>
<td>2.77</td>
</tr>
<tr>
<td>.01</td>
<td>4.32</td>
<td>5.76</td>
<td>4.87</td>
<td>4.64</td>
<td>4.49</td>
</tr>
<tr>
<td>.005</td>
<td>5.06</td>
<td>7.86</td>
<td>5.92</td>
<td>5.57</td>
<td>5.27</td>
</tr>
<tr>
<td>.001</td>
<td>6.89</td>
<td>14.23</td>
<td>9.49</td>
<td>8.07</td>
<td>7.13</td>
</tr>
</tbody>
</table>
Recall that the asymptotic distributions of the correlation-type test statistics have the representation

\[ Q_m = \sum_{i=1}^{m} \xi_i (Z_i + \psi_i)^2, \]

where the \( Z_i, 1 \leq i \leq m \), are i.i.d. standard normal random variables, \( \xi_i > 0 \) and \( \psi_i \) are bounded constants for all \( 1 \leq i \leq m \), and \( m = K-1 \) or \( K-2 \). Exact quantiles of the distribution of \( Q_m \), for selected values of \( \xi_i \) and all \( \psi_i = 0 \), have been published for \( m = 2, 3, 4, 5, 6, 8, \) and \( 10 \). References can be found in Solomon and Stephens (1977). In the same paper, they proposed two new approximations to the distribution of \( Q_m \) and concluded from their calculations that the three-moment chi-square fit, which approximates the distribution of \( Q_m \) by \( \rho(\chi_p^2)^r \) with the constants \( \rho, p \) and \( r \) being determined by matching the first three moments, has an overall supremacy over other approximations discussed there. Instead of using a Monte Carlo simulation, the asymptotic null upper quantiles of the correlation-type statistics can be approximated by their method.

5.3. Power Studies.

The empirical powers of the correlation-type statistic

\[ N(1-R_N^2) \]

for testing exponentiality with the origin at \( 0 \) against the following alternative distributions whose origins are also at \( 0 \) are calculated:
(i) \( \text{Exp}(0.2) \),

(ii) \( \text{Weibull} \ (0.5, 2) \),

(iii) \( \text{Weibull} \ (2, 0.5) \),

(iv) \( \chi^2_4 \),

(v) \( \chi^2_6 \), and

(vi) \( \text{Unif} [0, 1] \),

where \( \text{Weibull} \ (\lambda, \eta) \) has the distribution function \( F(x) = 1 - e^{-(\lambda x)^\eta} \). Figure 1 displays the plot of \( \ln[1 - F(x)] \) versus \( x \) for the alternatives (ii)-(v) to reveal how un-exponential-like these alternatives \( F \) are.

The two different sample sizes under study are \( N = 30 \) and 100. The empirical powers for \( N = 30 \) and 100 are displayed in Tables 6 and 7, respectively. The censoring distributions are all \( \text{Exp}(\beta) \) with individual \( \beta \) and the corresponding e.c.p. specified in both tables. Each entry of Tables 6 and 7 was based on 1,000 replications. For comparison, the empirical powers using asymptotic and finite-sample null quantiles for the test denoted, respectively, by (a) and (e) in both tables were computed with respect to three nominal significance levels \( \alpha = .10, .05 \) and \( .01 \).

When a randomly censored sample was generated, the empirical proportion of censored observations was calculated. Regardless of the true e.c.p. of the censored sample, the nearest e.c.p. to this empirical censoring proportion which was a multiple of 5%
Figure 1. $\ln[1-F(x)]$ versus $x$ for the Alternative Distribution $F$
Table 6. Empirical Powers of the Correlation-Type Statistic for Testing Exponentiality with the Origin at 0 When $k = 6$, $p_i = 0.2, 0.35, 0.45, 0.55, 0.65, 0.8$, Based on 1,000 Monte Carlo Samples

Sample Size = 30

<table>
<thead>
<tr>
<th>Alternative Distribution</th>
<th>Censoring Distribution</th>
<th>Expected Censoring Proportion</th>
<th>Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>.10</td>
</tr>
<tr>
<td>Exp(.2)</td>
<td>Exp(.05)</td>
<td>20%</td>
<td>(a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(e)</td>
</tr>
<tr>
<td>Weibull (.5, 2)</td>
<td>Exp(.15)</td>
<td>23%</td>
<td>(a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(e)</td>
</tr>
<tr>
<td>Weibull (2, .5)</td>
<td>Exp(.25)</td>
<td>16%</td>
<td>(a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(e)</td>
</tr>
<tr>
<td>$\chi^2_4$</td>
<td>Exp(.08)</td>
<td>26%</td>
<td>(a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(e)</td>
</tr>
<tr>
<td>$\chi^2_6$</td>
<td>Exp(.05)</td>
<td>25%</td>
<td>(a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(e)</td>
</tr>
<tr>
<td>Unif [0, 1]</td>
<td>Exp(.50)</td>
<td>21%</td>
<td>(a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(e)</td>
</tr>
</tbody>
</table>
Table 7. Empirical Powers of the Correlation-Type Statistic for Testing Exponentiality with the Origin at 0 When $K = 6$, $p_1 = 0.2$, 0.35, 0.45, 0.55, 0.65, 0.8, Based on 1,000 Monte Carlo Samples

Sample Size = 100

<table>
<thead>
<tr>
<th>Alternative Distribution</th>
<th>Censoring Distribution</th>
<th>Expected Censoring Proportion</th>
<th>Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>.10</td>
</tr>
<tr>
<td>Exp(.2)</td>
<td>Exp(.05)</td>
<td>20%</td>
<td>(a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(e)</td>
</tr>
<tr>
<td>Weibull (.5, 2)</td>
<td>Exp(.15)</td>
<td>23%</td>
<td>(a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(e)</td>
</tr>
<tr>
<td>Weibull (2, .5)</td>
<td>Exp(.25)</td>
<td>16%</td>
<td>(a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(e)</td>
</tr>
<tr>
<td>$\chi^2_4$</td>
<td>Exp(.08)</td>
<td>26%</td>
<td>(a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(e)</td>
</tr>
<tr>
<td>$\chi^2_6$</td>
<td>Exp(.05)</td>
<td>25%</td>
<td>(a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(e)</td>
</tr>
<tr>
<td>Unif [0, 1]</td>
<td>Exp(.50)</td>
<td>21%</td>
<td>(a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(e)</td>
</tr>
</tbody>
</table>
served as a substitute to obtain the null upper quantiles of $N(1-R_N^2)$ tabulated in Table 3 as the approximate critical values of the test. Those Monte Carlo samples with empirical censoring proportion larger than 50% or containing less than two uncensored observations were excluded in the power calculation.

Tables 6 and 7 show that the correlation-type test for exponentiality with origin known acts very powerfully for $N = 100$, but not quite so for $N = 30$. In the former case, the empirical asymptotic and exact powers are close to each other. In the latter case, the empirical asymptotic and exact powers are not as close, but the difference between them is within 0.05.

Separately, three-point Lagrange interpolation of null upper quantiles for the empirical censoring proportion with respect to the 5% multiples of e.c.p.'s was undertaken in power calculations. The resultant empirical powers are almost the same as the corresponding empirical powers listed in Tables 6 and 7 with differences less than 0.01. Hence, a detailed report of these results is omitted.

If the alternative distribution is given with origin known beforehand and one insists on employing the statistic $N(1-R_N^2)$ derived for a location-scale family to test exponentiality, the resultant power will be considerably lower than that using the statistic $N(1-R_N^2)$. Some Monte Carlo studies were performed for this case, but the results are not reported here.
VI. **Examples**

**Example 1.** The data displayed in Table 2 of Hollander and Proschan (1979) are an updated (March, 1977) version of the data originally analyzed by Koziol and Green (1976). Among 211 stage IV prostatic cancer patients in this data set, 90 died of prostatic cancer, 105 died of other diseases, and 16 were surviving on the closing date. The latter two groups are treated as censored observations.

For testing the simple hypothesis of Koziol and Green (1976) that the survival distribution for the deaths from the prostatic cancer is exponential with mean 100 months, Hollander and Proschan (1979) computed the two-sided p-values of their statistic \( C \), Hyde's (1977) statistic, and the Koziol-Green (1976) statistic \( \psi_N^2 \). They also commented on the sensitivity of these tests with respect to the given data set.

Koziol (1980) considered the Cramér-von Mises-type \( \hat{W}_N^2 \), the Kolmogorov-type \( \hat{D}_N \), and the Kuiper-type \( \hat{V}_N \) statistics corresponding to the Hall-Wellner (1980) transformation of the Kaplan-Meier estimator. Csörgő and Horváth (1981) mentioned in their paper the p-values obtained by Koziol for these three tests applied to the Hollander-Proschan data set. On the other hand, Csörgő and Horváth calculated the p-value of their Cramér-von Mises-type statistic \( \omega_N^2(0,T) \) corresponding to the Efron (1967) transformation of the Kaplan-Meier estimator.

All the p-values of the aforementioned statistics are listed as follows:
<table>
<thead>
<tr>
<th>Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.49</td>
</tr>
<tr>
<td>Hyde</td>
<td>0.86</td>
</tr>
<tr>
<td>$\psi_N^2$</td>
<td>0.14</td>
</tr>
<tr>
<td>$\hat{W}_N$</td>
<td>0.15</td>
</tr>
<tr>
<td>$\hat{D}_N$</td>
<td>0.10</td>
</tr>
<tr>
<td>$\hat{V}_N$</td>
<td>0.04</td>
</tr>
<tr>
<td>$\omega_N^2(0,T)$</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Not all the p-values are consistent with the hypothesized exponential distribution with mean 100 months. As a remark, the MLE for the mean of an exponential distribution estimated from these data is 102.944. Figure 1 of Hollander and Proschan (1979) plots the Kaplan-Meier estimator and the hypothesized exponential distribution. Eyeball analyses of this figure by different individuals might lead to different conclusions on rejection or acceptance of the given simple hypothesis.

For a comparative study considering the composite hypothesis that $F$ is an exponential distribution with origin at 0 with $K = 6$ and $\{p_i\} = \{0.2, 0.35, 0.45, 0.55, 0.65, 0.8\}$, we computed

$$N(1 - \frac{\hat{R}^2}{N}) = 7.651,$$

which corresponds to a p-value about 0.03. For testing the hypothesis that $F$ is an exponential distribution with origin and scale parameters unknown,
\[ N(1-R_N^2) = 23.022, \]

which corresponds to a p-value about 0.02.

The preceding two p-values are found from the asymptotic null upper quantiles of the corresponding test statistics by setting the e.c.p. equal to the empirical proportion of censored observations, 57.3%. These quantiles are displayed in Table 8.

In conclusion, both correlation-type tests reject the composite hypotheses of exponentiality for this data set.

Example 2. When a follow-up study terminates with a large proportion of censored observations which are larger than the maximal uncensored one, it is not appropriate to test the goodness-of-fit of the lifetime distribution \( F \) on the infinite interval \([0, \infty)\). It is only sensible to test the hypothesis that \( F \) follows a specified distributional law in the early life stage. For example, one may test

\[ H_0 : F(x) = 1 - e^{- \frac{x-\theta}{\tau}} \text{ for } x \in [\theta, F^{-1}(A)), \]

with unknown \( \theta \) and \( \tau > 0 \), where \( A \) is a pre-specified proportion.

For this situation the Koziol-Green model of random censorship is assumed to hold only in \([\theta, F^{-1}(A))\), i.e.,

\[ [1-G(x)] = [1-F(x)]^\beta \text{ for } x \in [\theta, F^{-1}(A)) \]
Table 8. Upper Quantiles of the Asymptotic Null Distributions of the Correlation-Type Test Statistics
When $K = 6$, $p_1 = 0.2, 0.35, 0.45, 0.55, 0.65, 0.8,$
Based on 40,000 Random Sample Simulations

Expected Censoring Proportion = 57.3%

<table>
<thead>
<tr>
<th>Upper Percentage</th>
<th>(a) Statistic $N(1-R_N^2)$ for Testing Exponentiality with the Origin at 0</th>
<th>(b) Statistic $N(1-R_N^2)$ for Testing Exponentiality with the Origin and Scale Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>4.45</td>
<td>10.74</td>
</tr>
<tr>
<td>0.09</td>
<td>4.68</td>
<td>11.29</td>
</tr>
<tr>
<td>0.08</td>
<td>4.96</td>
<td>11.90</td>
</tr>
<tr>
<td>0.07</td>
<td>5.30</td>
<td>12.62</td>
</tr>
<tr>
<td>0.06</td>
<td>5.65</td>
<td>13.55</td>
</tr>
<tr>
<td>0.05</td>
<td>6.06</td>
<td>14.56</td>
</tr>
<tr>
<td>0.04</td>
<td>6.59</td>
<td>15.81</td>
</tr>
<tr>
<td>0.03</td>
<td>7.26</td>
<td>17.43</td>
</tr>
<tr>
<td>0.02</td>
<td>8.30</td>
<td>19.88</td>
</tr>
<tr>
<td>0.01</td>
<td>10.16</td>
<td>24.37</td>
</tr>
<tr>
<td>0.005</td>
<td>12.09</td>
<td>28.87</td>
</tr>
<tr>
<td>0.001</td>
<td>16.48</td>
<td>39.48</td>
</tr>
</tbody>
</table>

54
with a positive constant. With this adjustment, we can compute the e.c.p. before $F^{-1}(A)$ by

$$\int_0^{F^{-1}(A)} [1-F(x)] \, dG(x) = \frac{\beta}{1+\beta} [1 - (1-A)^{1+\beta}].$$

In practice, it is plausible to take

$$A = \max\{\hat{F}_N(T_i), 1 \leq i \leq N: \delta_i = 1, \hat{F}_N(T_i) < 1\}$$

and to select

$$p^*_c = N^{-1} \sum_{i=1}^{N} I[\delta_i = 0, T_i < \hat{F}_N^{-1}(A)]$$

as an estimator for the e.c.p. before $F^{-1}(A)$, since

$$N^{-1} \sum_{i=1}^{N} I[\delta_i = 0, T_i < t]$$

is a consistent estimator for the e.c.p. before $t$. We then estimate the censoring parameter $\beta$ by the $\hat{\beta}$ such that

$$p^*_c = \frac{\hat{\beta}}{1+\hat{\beta}} [1 - (1-A)^{1+\hat{\beta}}], \quad (6.1)$$

and use it to find the asymptotic null distribution of the correlation-type test statistics.

The following data set is presented to illustrate the applicability of the preceding arguments.

Begun in February 1978, the Recurrent Coronary Prevention Project (RCPP) at the Mt. Zion Hospital in San Francisco has investigated the effects of behavior modification for myocardial
in farction (MI) patients who have Type A behavior patterns. The data displayed in Table 9 correspond to 606 MI patients in a treatment group, Section II, of the RCPP. The study will run for five years, and the current analysis uses the data up to June 30, 1981. Recurrence of fatal or non-fatal MI is reviewed as a failure. Thirty-one patients failed, and the other 575 patients who were alive but had not failed, had died of other causes, or had withdrawn from the study are all treated as censored observations.

The Kaplan-Meier estimator and the maximum likelihood estimator of an exponential distribution with estimated mean of 16107.99 days for the failure times are plotted in Figure 2.

The largest uncensored observation is 1052 days, so we select $A = \hat{R}_N(1052) = 0.0658$. The total number of censored observations which are smaller than 1052 days is 381. By solving equation (6.1), we obtained the estimated censoring parameter $\hat{\beta} = 15.28$.

Choosing $K = 6$ and $\{p_i\} = \{0.01, 0.02, 0.03, 0.04, 0.05, 0.06\}$ and setting $\beta = 15.28$, we obtained the asymptotic null distributions of the correlation-type statistics evaluated at various ordinates from 40,000 random sample simulations. Table 10 partially displays these results.

To test whether the failure time before 1052 days is exponentially distributed, we computed

$$N(1-R_N^2) = 3.298 \text{ with } \text{p-value} = 0.73$$
Table 9. Recurrence Times of Myocardial Infarction and Withdrawal Times in Days for 606 Patients in Section II of the RCPP (with multiplicities given in parentheses)

<table>
<thead>
<tr>
<th>Recurrence Times of M.I.</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>165</td>
<td>285</td>
<td>391</td>
<td>596</td>
<td>745</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>187</td>
<td>306</td>
<td>395</td>
<td>597</td>
<td>790</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>197</td>
<td>313</td>
<td>499</td>
<td>647(2)</td>
<td>972</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>268</td>
<td>330</td>
<td>564</td>
<td>661</td>
<td>1015</td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>276</td>
<td>390</td>
<td>574</td>
<td>702</td>
<td>1052</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Withdrawal Times</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5(2)</td>
<td>210(2)</td>
<td>378</td>
<td>524</td>
<td>753</td>
<td>1029(10)</td>
<td></td>
</tr>
<tr>
<td>7(3)</td>
<td>212</td>
<td>385(2)</td>
<td>539</td>
<td>782</td>
<td>1034(4)</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>217</td>
<td>392</td>
<td>546</td>
<td>784</td>
<td>1035(8)</td>
<td></td>
</tr>
<tr>
<td>21(4)</td>
<td>219</td>
<td>398</td>
<td>553</td>
<td>786</td>
<td>1040</td>
<td></td>
</tr>
<tr>
<td>28(2)</td>
<td>224(2)</td>
<td>399</td>
<td>567(2)</td>
<td>788(2)</td>
<td>1054(9)</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>231</td>
<td>401</td>
<td>581</td>
<td>811(6)</td>
<td>1056(3)</td>
<td></td>
</tr>
<tr>
<td>35(3)</td>
<td>243(2)</td>
<td>405(2)</td>
<td>594</td>
<td>826</td>
<td>1057(23)</td>
<td></td>
</tr>
<tr>
<td>42(2)</td>
<td>245</td>
<td>406</td>
<td>602(2)</td>
<td>837(14)</td>
<td>1062(7)</td>
<td></td>
</tr>
<tr>
<td>56(2)</td>
<td>249</td>
<td>408</td>
<td>608(2)</td>
<td>847</td>
<td>1079(8)</td>
<td></td>
</tr>
<tr>
<td>63(5)</td>
<td>266(2)</td>
<td>413</td>
<td>609</td>
<td>854(8)</td>
<td>1091(8)</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>273</td>
<td>420</td>
<td>612</td>
<td>860</td>
<td>1098(11)</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>280(2)</td>
<td>426</td>
<td>629</td>
<td>861(5)</td>
<td>1103(7)</td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>292</td>
<td>427(2)</td>
<td>630</td>
<td>862(24)</td>
<td>1114(8)</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>301</td>
<td>434</td>
<td>633</td>
<td>902(9)</td>
<td>1120(6)</td>
<td></td>
</tr>
<tr>
<td>105(2)</td>
<td>309</td>
<td>435</td>
<td>646</td>
<td>907(17)</td>
<td>1121(5)</td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>316</td>
<td>440</td>
<td>651</td>
<td>917</td>
<td>1125(9)</td>
<td></td>
</tr>
<tr>
<td>126</td>
<td>322(4)</td>
<td>448(2)</td>
<td>653</td>
<td>929(5)</td>
<td>1133</td>
<td></td>
</tr>
<tr>
<td>128</td>
<td>329</td>
<td>453</td>
<td>662</td>
<td>943(8)</td>
<td>1136</td>
<td></td>
</tr>
<tr>
<td>133</td>
<td>336</td>
<td>455(3)</td>
<td>665</td>
<td>966(23)</td>
<td>1139(4)</td>
<td></td>
</tr>
<tr>
<td>144</td>
<td>343</td>
<td>462</td>
<td>673</td>
<td>967(5)</td>
<td>1161(29)</td>
<td></td>
</tr>
<tr>
<td>147</td>
<td>350</td>
<td>469</td>
<td>679</td>
<td>971(2)</td>
<td>1165(5)</td>
<td></td>
</tr>
<tr>
<td>154(2)</td>
<td>351</td>
<td>474</td>
<td>700</td>
<td>973(2)</td>
<td>1176(6)</td>
<td></td>
</tr>
<tr>
<td>161</td>
<td>357</td>
<td>490</td>
<td>707(2)</td>
<td>982</td>
<td>1177(7)</td>
<td></td>
</tr>
<tr>
<td>168</td>
<td>364</td>
<td>495</td>
<td>712</td>
<td>1002(2)</td>
<td>1182(3)</td>
<td></td>
</tr>
<tr>
<td>174</td>
<td>370</td>
<td>497</td>
<td>714</td>
<td>1012(3)</td>
<td>1183(7)</td>
<td></td>
</tr>
<tr>
<td>175(2)</td>
<td>371(2)</td>
<td>506</td>
<td>716</td>
<td>1020(13)</td>
<td>1219(3)</td>
<td></td>
</tr>
<tr>
<td>182(2)</td>
<td>374</td>
<td>511(2)</td>
<td>719</td>
<td>1021(25)</td>
<td>1223(5)</td>
<td></td>
</tr>
<tr>
<td>190</td>
<td>377</td>
<td>518</td>
<td>721(4)</td>
<td>1022(21)</td>
<td>1225(19)</td>
<td></td>
</tr>
<tr>
<td>203</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 2. The Kaplan-Meier Estimator and the MLE Exponential Distribution for the RCPP Data
for the hypothesis of exponentiality with the origin at 0, and

\[ N(1-R_N^2) = 13.469 \text{ with } p\text{-value} = 0.61 \]

for the hypothesis of exponentiality with the origin and scale unknown. Here the p-values were calculated by three-point Lagrange interpolation from Table 10.

Just as for an eyeball analysis of Figure 2, the two correlation-type tests conclude that the recurrence time of MI in the early stage for this treatment group follows an exponential distribution.
Table 10. Asymptotic Null Distributions of the Correlation-Type Test Statistics When $K = 6$, $p_i = 0.01, 0.02, 0.03, 0.04, 0.05, 0.06$, Based on 40,000 Random Sample Simulations

Censoring Parameter $\beta = 15.28$

(a) For Testing Exponentiality with the Origin at 0

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>$\Pr{N(1-R_N^2) \leq \nu}$</th>
<th>(b) For Testing Exponentiality with the Origin and Scale Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4964</td>
<td>0.0065</td>
<td>$w$</td>
</tr>
<tr>
<td>0.9929</td>
<td>0.0304</td>
<td>1.5002</td>
</tr>
<tr>
<td>1.4893</td>
<td>0.0704</td>
<td>3.0004</td>
</tr>
<tr>
<td>1.9858</td>
<td>0.1185</td>
<td>4.5005</td>
</tr>
<tr>
<td>2.4822</td>
<td>0.1744</td>
<td>6.0007</td>
</tr>
<tr>
<td>2.9787</td>
<td>0.2331</td>
<td>7.5009</td>
</tr>
<tr>
<td>3.4751</td>
<td>0.2926</td>
<td>9.0011</td>
</tr>
<tr>
<td>3.9715</td>
<td>0.3495</td>
<td>10.5013</td>
</tr>
<tr>
<td>4.4680</td>
<td>0.4021</td>
<td>12.0014</td>
</tr>
<tr>
<td>4.9644</td>
<td>0.4527</td>
<td>13.5016</td>
</tr>
<tr>
<td>5.4609</td>
<td>0.4989</td>
<td>15.0018</td>
</tr>
<tr>
<td>5.9573</td>
<td>0.5433</td>
<td>16.5020</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18.0022</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0107</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0391</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0794</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1278</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1795</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2348</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2911</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3456</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3956</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4434</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4907</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5328</td>
</tr>
</tbody>
</table>
VII. Concluding Remarks

The results presented in the preceding two sections are merely for testing exponentiality. The same ideas can certainly be utilized in testing the goodness of fit to other parametric families - for instance, Weibull distributions or gamma distributions. For other families the robustness of the tests against departures from the Koziol-Green model of random censorship should be investigated.

It has been shown that the correlation-type tests of exponentiality developed with the Koziol-Green model have excellent power for large samples, say \( N \geq 100 \), but, not unexpectedly, the power is not as good for small samples, for example at \( N = 30 \). Even in the application to large samples, there are still some open questions about the correlation-type tests of exponentiality. The effect of substituting a consistent estimator for the censoring parameter \( \beta \) needs further scrutiny. Optimal selection of the fixed number of quantiles, \( K \), and the quantile set \( \{ p_i, 1 \leq i \leq K \} \) is worth close examination.

For future study, it would also be interesting to compare the power of the proposed tests with that of some parametric tests - for example, testing exponentiality by a likelihood ratio statistic within the Weibull family or within the gamma family.

Another possible approach to correlation-type tests could be to choose \( p_i = i/(N+1), 1 \leq i \leq K \), where \( K \) depends on \( N \) and \( G \circ F^{-1}(p_K) < 1 \). Future investigation along this direction may produce a more powerful test for goodness of fit with randomly censored data.
References


