Goodness-of-Fit Tests for Censored Survival Data

by

Robert James Gray

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

June 1983

AN ABSTRACT OF THE THESIS OF

ROBERT JAMES GRAY for the degree of DOCTOR OF PHILOSOPHY in STATISTICS presented on July 8, 1982

Title: GOODNESS-OF-FIT TESTS FOR CENSORED SURVIVAL DATA

Abstract approved: Redacted for Privacy

Donald A. Pierce

The general problem of testing goodness-of-fit with right censored data is considered. The discussion is focused on testing for the underlying distribution in general regression models. Following Cox and Snell (1968) the model is specified in terms of generalized residuals, which have a uniform distribution on (0,1) when the model is true. The problem can then be expressed as testing whether the assumed distribution of the residuals is correct.

A modification of the empirical distribution function for use with censored data, proposed independently by Aitkin and Clayton (1980), is presented. This modified estimate, $\hat{H}_n^*$, is used to estimate the distribution of the residuals. It is proposed that tests be based on the functionals $n \int_0^1 \psi_\ell(u)\, d[\hat{H}_n^*(u) - u]$. It is shown that these functionals are efficient scores from certain parametric alternatives, so that likelihood theory can be used to find the asymptotic distributions. A general discussion of possibilities for estimating the asymptotic variance of the functionals is given. The connection with likelihood theory is exploited to investigate the asymptotic structure of the process $\sqrt{n}\,[\hat{H}_n^*(u) - u]$.

A test of the type considered here that is of special interest is the Neyman smooth test. Possibilities for estimating the asymptotic variance for the Neyman smooth test for exponentiality are considered in detail. The Neyman smooth test is used to test for underlying exponential and Weibull distributions in two data sets taken from the literature.

APPROVED:

Redacted for Privacy

Professor of Statistics

in charge of major

Redacted for Privacy

Chairman of Department of Statistics

Redacted for Privacy

Dean of Graduate School

Date thesis is presented July 8, 1982

Typed by Dorothy Jameson for Robert James Gray

ACKNOWLEDGEMENT

I wish to thank my major professor, Dr. Donald A. Pierce, for his continued and enthusiastic support of my academic career, and for his insightful direction and encouragement, without which this thesis would not have been possible. I also wish to thank the members of the faculty of the Department of Statistics for their inspiration and guidance during my graduate study.

This work was supported in part by United States Public Health Service grant CA27532.

TABLE OF CONTENTS

List of Figures

List of Tables

GOODNESS-OF-FIT TESTS FOR CENSORED SURVIVAL DATA

I. Introduction

The problem considered here is testing whether the underlying distribution in a general regression model has the assumed form. We are particularly interested in models arising in the analysis of survival data, where a major complication is the presence of right censoring in the data.

In survival analysis the data are times to some event. Here the term "failure" is used to represent the event of interest, although the models considered here are also appropriate for other kinds of events. Let T_i denote the failure time of the i-th experimental unit, $i = 1, \ldots, n$, which for convenience we assume to be positive. Throughout the paper, we use capital letters to represent random variables and the corresponding lower case letters for observed values of those variables. We assume the failures are independent, and that T_i has continuous distribution function $F_i(t;\gamma)$ with density $f_i(t;\gamma)$ and survivor function $\overline{F}_i(t;\gamma) = 1 - F_i(t;\gamma)$, where γ is a vector of unknown parameters. Further, we assume $F_i(t;\gamma) = F(t;\lambda_i,\delta)$, $\lambda_i = x_i'\beta$, $\gamma = (\beta,\delta)$, where the x_i are vectors of known constants. The components of the x_i are the covariables in the regression. The interest here is in the adequacy of the underlying distribution F. For example, in the exponential model where $F_i(t;\beta) = 1 - \exp\{-t\cdot\exp(x_i'\beta)\}$, it is the assumption of an exponential distribution we wish to examine.

In general not all the failure times will be observed. We assume that associated with the i-th observation is a censoring time V_i with distribution function $C_i(v)$ and survivor function $\overline{C}_i(v) = 1 - C_i(v)$, and that only $Y_i = \min\{T_i, V_i\}$ and $Z_i = I(T_i \leq V_i)$ are observed, where $I(A)$ is the indicator function of the event A (i.e., $I(A)$ equals 1 if A is true and 0 otherwise). The censoring times are assumed to be independent of each other and of the failure times. The censoring is assumed to be noninformative, in that $C_i(v)$ is assumed not to depend on γ. We refer to the above model as the heterogeneous censoring model. Two special cases of interest are homogeneous random censorship, where $C_i(v) \equiv C(v)$ for all i, and type I censoring where $C_i(v) = I(T \leq v)$ for all i, for a fixed time T. The assumption of strict independence between the censoring times and the failure times, which is used below in computing expected values, excludes such censoring schemes as type II censoring, where the experiment continues until a specified number of failures has occurred.

An important subclass of regression models are the location-scale models, where

$$F_i(t;\gamma) = G([h(t) - x_i'\beta]/\sigma) , \tag{1.1}$$

$\gamma = (\beta,\sigma)$, for some monotone transformation h of T_i and some continuous distribution function G. In this case the question of whether the underlying distribution is correct can be expressed as whether the distribution of the (unobserved) true residuals $E_i = (h(T_i) - x_i'\beta) / \sigma$ is given by G.

This can be extended to the general setting. For the types of models considered here it is always possible to find monotone transformations $r_i(\cdot;\gamma)$ such that the random variables $E_i = r_i(T_i;\gamma)$ are identically distributed. Then these E_i can be regarded as true residuals and the question of the adequacy of the underlying distribution can again be formulated as whether the assumed distribution of the E_i is correct.

Generalized residuals were first considered by Cox and Snell (1968, 1971) for the uncensored case. They discussed how the observed residuals $\hat{e}_i = r_i(t_i;\hat{\gamma})$, where $\hat{\gamma}$ is the maximum likelihood estimate of γ, can be used much as ordinary residuals in location-scale regression models to examine the adequacy of the model. With censored data the use of plots of the $\hat{e}_i = r_i(y_i;\hat{\gamma})$ to examine distributional form has been considered by Crowley and Hu (1977) and Aitkin and Clayton (1980). Crowley and Hu estimate the distribution of the E_i with the Kaplan-Meier estimate computed from the $\hat{e}_i$. Aitkin and Clayton use a different estimate, which is discussed below.

The formulation of models in terms of generalized residuals is not unique. Since F_i is continuous, the random variables $U_i = F_i(T_i;\gamma)$ will always have a uniform distribution on $(0,1)$ when the model is true. In the general discussion it will be convenient to restrict attention to this form of generalized residuals.

There are many other types of departures from the assumed model that could be present. Some examples are failure to include important covariables, incorrect specification of the form of the relationship between the data and the covariables, lack of independence, and the presence of outliers. In general the presence of these other types of departures will have an effect on the techniques discussed below for examining distributional form, so in practice it will be necessary to combine other forms of residual analysis with these techniques before any firm conclusions can be reached. However, in the discussion here it is generally assumed that the only departure that might be present is that the underlying distribution is not correctly specified. Under this alternative, the U_i remain independent and identically distributed, but with some distribution other than the uniform. Thus the problem with which we are concerned can be expressed as follows. Given that the U_i are independent and identically distributed, test

$$H_o: \ U_i \text{ has a uniform distribution on } (0,1) \,, \tag{1.2}$$

against the alternative that U_i has some other distribution.

In general the basic approach used in testing goodness-of-fit is to compare an empirical estimate of the distribution to the hypothesized distribution. To use this approach for the hypothesis (1.2), we need an estimate of the distribution of the U_i. Let H be the true distribution function of U_i. First consider the uncensored case, where the failure times t_i are known for all observations. In this case, H can be estimated with the empirical distribution function,

$$\hat{H}_n(u) = n^{-1} \sum_{i=1}^{n} I(u \geq \hat{u}_i) \,, \qquad 0 \leq u \leq 1 \,, \tag{1.3}$$

which estimates the distribution of U_i by placing a mass of n^{-1} on each of the observed residuals $\hat{u}_i = F_i(t_i \, ; \hat{\gamma})$. (When attention is restricted to the simple null hypothesis case the "^" will usually be dropped.) Tests for the hypothesis (1.2) in the uncensored case are in general based on functionals of the stochastic process

$$\hat{y}_n(u) = \sqrt{n}\,[\hat{H}_n(u) - u]\ .$$

With censoring, we only have available the observed (censored) residuals $\hat{u}_i = F_i(y_i\,;\hat{\gamma})$. For a censored observation, the value of $F_i(t_i\,;\hat{\gamma})$ can lie anywhere in the interval $(\hat{u}_i\,, 1]$, since t_i is only known to be greater than y_i. Thus for censored observations, the mass of n^{-1} can no longer be placed on $F_i(t_i\,;\hat{\gamma})$, as in (1.3). One possibility for an estimate of H is the Kaplan-Meier (1958) estimator,

$$\hat{K}_n(u) = 1 - \prod_{i:\hat{u}_i \leq u} [(n - r_i) / (n - r_i + 1)]^{z_i}\ ,$$

where r_i is the rank of $(\hat{u}_i\,, 1 - z_i)$ in the lexicographic ordering of the sequence $(\hat{u}_1\,, 1 - z_1)\,, \cdots, (\hat{u}_n\,, 1 - z_n)$. The Kaplan-Meier estimate places a mass of n^{-1} on $\hat{u}_i$ for uncensored observations, and distributes a mass of n^{-1} over the failures in the interval $(\hat{u}_i\,, 1]$ for those $\hat{u}_i$ corresponding to censored observations. Tests could again be based on functionals of the process

$$\hat{w}_n(u) = \sqrt{n}\ [\hat{K}_n(u) - u]\ .$$

In this paper, instead of using the Kaplan-Meier estimator, we use an approach suggested by Aitkin and Clayton (1980). In this approach, we place a mass of n^{-1} on $\hat{u}_i$ for uncensored observations, and for censored observations we distribute the mass of n^{-1} over the interval $(\hat{u}_i, 1]$ to be consistent with the null hypothesis (uniform) distribution. This gives

$$\hat{H}_n^*(u) = n^{-1} \sum_{i=1}^{n} I(u \geq \hat{u}_i)\{z_i + (1 - z_i)(u - \hat{u}_i)/(1 - \hat{u}_i)\} \tag{1.4}$$

for the estimate of H. We call $\hat{H}_n^*$ the modified empirical distribution function (MEDF). This estimate is discussed in more detail in Sections III and V. We will base our tests on functionals of the form

$$\sqrt{n} \int_0^1 \psi_\ell(u) \; d\, \hat{y}_n^*(u) \, , \tag{1.5}$$

for arbitrary functions ψ_ℓ, where

$$\hat{y}_n^*(u) = \sqrt{n} \; [\, \hat{H}_n^*(u) - u \,] \, .$$

In the following section we review previous results for the regression case with no censoring and the identically distributed case with censoring. In Section III we begin by examining efficient scores from a certain class of parametric alternatives, and show these scores are the same as the functionals (1.5). Because of this connection, the asymptotic distributions are given by likelihood theory. In Section IV we discuss the asymptotic distribution of the scores and general issues in estimating their asymptotic variances. In Section V the asymptotic distribution of $\hat{y}_n^*$ is considered. In Section VI chi-square and Neyman smooth tests are examined. In Section VII the Neyman smooth test is applied to two data sets taken from the literature.

II. Review of Previous Results

In this section we discuss previous results on goodness-of-fit tests, first for regression models in the uncensored case, and then for the identically distributed case (where $F_i(\cdot;\gamma) \equiv F(\cdot;\gamma)$ for all i) with censoring.

For general regression models in the uncensored case, it was indicated by Pierce and Kopecky (1979) and proven in detail by Loynes (1980), that, subject to regularity conditions, $\hat{y}_n$ converges weakly to a mean 0 Gaussian process $\hat{y}$ under the null hypothesis. The main result of the Pierce-Kopecky paper is that for the location-scale models (1.1), the covariance structure of $\hat{y}$ is the same for regression models as in the identically distributed case, and thus known (Durbin, 1973) not to depend on the true values of the parameters. Consequently, for location-scale regression models in the uncensored case, the asymptotic distribution of suitably continuous functionals of $\hat{y}_n(u)$ will be the same as in the identically distributed case, and so will not depend on the covariables or the true values of the parameters, but only on the type of distribution. In the general uncensored case the result of Loynes shows that the asymptotic distributions of functionals of $\hat{y}_n(u)$ do in general depend on the covariables and the true values of the parameters. Since for many common goodness-of-fit statistics these asymptotic distributions are difficult to find, the use of such statistics to test the hypothesis (1.2) will not be easy in practice. Loynes suggests using a random adjustment of the parameter estimates, and shows that if the adjustment is of the proper form, the asymptotic process reduces to a standard brownian bridge. However, the introduction of a random adjustment is conceptually unappealing. The methods discussed in the following sections for use with censored data should also be useful in the general uncensored case.

Loynes also gives the asymptotic distribution of $\hat{y}_n(u)$ under a sequence of local alternatives. In the appendix, his result is used to show that for the location-scale models (1.1), functionals of $\hat{y}_n(u)$ have "no local power" against omitting covariables from the model, in the sense that the asymptotic distribution is the same under the sequence of local alternatives $T_{in} \sim G[(h(t) - x_i'\beta - w_i'\alpha_n)/\sigma]$, $i = 1, \ldots, n$, where $\alpha_n = \lambda/\sqrt{n}$, as it is under the null hypothesis $(\lambda = 0)$. This result does not extend to general regression models or to location-scale models with censoring, and an example in Section VII shows the effect can be substantial.

With censored data the problem of testing goodness-of-fit becomes considerably more complex. In the identically distributed, simple null hypothesis case the asymptotic distribution of the process $w_n(u)$ was established by Breslow and Crowley (1974) under homogeneous random censorship, and was found to depend on the censoring distribution C. Related results, including a transformation of $w_n(u)$ that asymptotically does not depend on the censoring, were given by Burke, Csörgö and Horváth (1981). There are a number of results on specific tests available in the simple null hypothesis case with randomly censored data. A Cramér-von Mises statistic was given by Koziol and Green (1976) and Kolomogorov-Smirnov tests by Fleming _et_. _al_. (1980) and Fleming and Harrington (1981). Koziol (1980) gave a different approach to both types of statistics, as well as Kuiper statistics. Hyde (1977) and Hollander and Proschan (1979) gave tests that have their power focused more toward particular alternatives than these omnibus procedures. Confidence bands for an unknown distribution function (and associated tests) were considered by Gillespie and Fisher (1979), Hall and Wellner (1980) and Nair (1981).

With a composite null hypothesis, the transformations considered by several of the above authors no longer remove the dependence on the censoring distribution, and there are only a few results on tests available, even in the identically distributed case, mostly for special censoring models. Smith and Bain (1976) gave correlation type statistics for several distributions with type II censoring, and Chen (1981) considered similar tests under several random censoring distributions. Pettit (1976, 1977) considered Cramér-von Mises tests for normal and exponential distributions with type I and type II censored data. Turnbull and Weiss (1978) gave a likelihood ratio test for grouped data. Chi-square tests were studied by Mihalko and Moore (1980) for type II censoring and by Habib (1981) for homogeneous random censoring. Bargal (1981) considered several procedures for testing for Weibull and Gamma distributions with type I censoring.

Habib showed the asymptotic normality of the finite dimensional distributions of $\hat{w}_n(u)$, although he did not prove weak convergence of the process. Even for location-scale families, these asymptotic distributions depend on the censoring distribution and on the true values of the parameters. This dependence is also seen in the other results listed above, except that of Turnbull and Weiss (as discussed in Section IV, below, the likelihood function does not depend on the censoring distribution).

In the censored regression case there are no results on functionals of $\hat{w}_n$, although there are some results for examining goodness-of-fit in specialized situations. For example, Farewell and Prentice (1977) considered use of the generalized gamma distribution for discriminating between several simpler distributions. Since in the identically distributed, location-scale case, the asymptotic distributions of functionals of $\hat{w}_n(u)$ depend on the true values of the parameters and the censoring, this dependence, as well as dependence on the covariables, will also be seen in the regression case, even for location-scale models. As in the uncensored case with general regression models, this will make many common types of goodness-of-fit statistics very difficult to use. It will be seen below that the connection between the functionals (1.5) and the efficient scores allows asymptotic likelihood theory to be used here, so to some extent the complex distributional problems associated with many other methods can be avoided.

III. Efficient Scores and the MEDF

In this section we begin by considering efficient scores for a certain class of parametric alternatives. Then we show the scores are the same as the functionals (1.5). The section is concluded by further exploring the relationship between the MEDF and the Kaplan-Meier estimator.

Under the null hypothesis the U_i are uniformly distributed on $(0,1)$. We consider alternative distributions with densities of the form $\exp\{\sum_{\ell=1}^{m} \theta_\ell \psi_\ell(u) - K(\theta)\}$, $0 < u < 1$, where the ψ_ℓ are known functions and $K(\theta)$ is a normalizing constant. This gives a family of densities for T_i of

$$f_i(t;\gamma) \exp\{\sum_{\ell=1}^{m} \theta_\ell \psi_\ell[F_i(t;\gamma)] - K(\theta)\} . \tag{3.1}$$

These reduce to the null hypothesis density when $\theta = 0$. The principal advantage of this approach to goodness-of-fit tests is that the complex distributional problems of other approaches can to some extent be avoided by using asymptotic likelihood theory. Another advantage is that the power of the tests can be focused against alternatives of particular interest. A third advantage is that consideration of (3.1) leads to some general principles of interest, independently of the particular choice of the ψ_ℓ.

Under the censoring models introduced in Section I, the log likelihood of the data under the alternative densities is

$$\ell(\gamma,\theta) = \sum_{i=1}^{n} \{z_i \left(\log f_i(y_i;\gamma) + \sum_{\ell=1}^{m} \theta_\ell \, \psi_\ell[F_i(y_i;\gamma)]\right)$$

$$+ (1-z_i) \log \int_{F_i(y_i;\gamma)}^{1} \exp\{\sum_{\ell=1}^{m} \theta_\ell \, \psi_\ell(u)\} \, du - K(\theta)\} \ . \qquad (3.2)$$

(This actually holds under less restrictive assumptions on the censoring. For a discussion see Kalbfleisch and Prentice, 1980, Section 5.2.) We will base our tests on the efficient scores

$$\partial\ell(\hat{\gamma}_o,0)/\partial\theta_\ell = \sum_{i=1}^{n} \{z_i \, \psi_\ell(\hat{u}_i) + (1-z_i) \int_{\hat{u}_i}^{1} \psi_\ell(u) \, du \,/\, (1-\hat{u}_i) - \partial K(0)/\partial\theta_\ell\}$$

$$= \sum_{i=1}^{n} \{z_i \, \psi_\ell(\hat{u}_i) + (1-z_i) \, E[\psi_\ell(\mathcal{U}) \mid \mathcal{U} > \hat{u}_i] - E[\psi_\ell(\mathcal{U})]\}, \quad (3.3)$$

where $\mathcal{U}$ has a uniform distribution on $(0,1)$, $\hat{\gamma}_o$ is the maximum likelihood estimate for γ when $\theta = 0$, and as before $\hat{u}_i = F_i(y_i;\hat{\gamma}_o)$. (Other estimates besides the maximum likelihood estimate could be considered, but some of the details of the tests would be different.) The expression for $\partial K(0)/\partial\theta_\ell$ comes from differentiating the equation $\int_0^1 \exp\{\sum_{\ell=1}^{m} \theta_\ell \, \psi_\ell(u) - K(\theta)\} du = 1$.

Uncensored observations contribute a term of $\psi_\ell[F_i(t_i;\hat{\gamma}_o)]$ to (3.3). For censored observations the value of t_i is not observed. Under the assumptions made in Section I about the censoring, all that is known about t_i for an observation censored at y_i is that $t_i > y_i$. For censored observations the score (3.3) takes the intuitively reasonable approach of replacing the unobserved contribution $\psi_\ell[F_i(t_i;\hat{\gamma}_o)]$

with its conditional expectation $E_{H_0}\{\psi_\ell[F_i(T_i;\gamma)] \mid T_i > y_i\}\big|_{\gamma=\hat{\gamma}_0}$, where E_{H_0} denotes the expectation under the null hypothesis. Another way of looking at (3.3) is that it is the expectation of the score for uncensored data conditional on the observed censored data. This is similar to the idea of the EM algorithm of Dempster, Laird and Rubin (1977).

Instead of using the scores (3.3), we could consider other likelihood based methods, specifically the likelihood ratio test or tests based on the maximum likelihood estimate $\hat{\theta}$. All three methods are asymptotically equivalent (see, for example, Cox and Hinkley, 1974, pages 323-4). However, for these other methods, it is necessary to fit the alternative model, which can be much more difficult than fitting the original model. Also, only the scores (3.3) have a direct interpretation as functionals of $\hat{y}_n^*$.

The class of tests arising from the scores (3.3) is very broad. It is shown below that any functional of $\hat{y}_n^*$ of the form (1.5) is a score of the form (3.3). Two omnibus tests included in this class in the uncensored, identically distributed case are the Rao-Robson (1974) modified χ^2 test and the Neyman smooth test. The extensions of these tests to the censored regression setting given by (3.3) and (1.5) are considered in Section VI. This class also includes efficient scores from some alternatives not of the form (3.1). For example, the efficient score for the Weibull $(F(t;\alpha,\sigma) = 1 - \exp\{-(t/\alpha)^\sigma\})$ alternative to the exponential $(\sigma = 1)$ is given by (3.3) with

$$\psi_\ell(u) = 1 + \log[-\log(1-u)][1 + \log(1-u)] .$$

To see that the scores (3.3) and the functionals (1.5) are the same, note that

$$\sqrt{n} \int_0^1 \psi_\ell(u)\, d\,\hat{y}_n^*(u) = n\left\{ \int_0^1 \psi_\ell(u)\, d\,\hat{H}_n^*(u) - \int_0^1 \psi_\ell(u)\, du \right\} \tag{3.4}$$

$$= \sum_{i=1}^{n} \left\{ z_i\, \psi_\ell(\hat{u}_i) + (1-z_i) \int_{\hat{u}_i}^1 \psi_\ell(u)\, du/(1-\hat{u}_i) - E[\psi_\ell(U)] \right\}$$

$$= \sum_{i=1}^{n} \left\{ z_i\, \psi_\ell(\hat{u}_i) + (1-z_i) E[\psi_\ell(U) \mid U > \hat{u}_i] - E[\psi_\ell(U)] \right\} ,$$

which is the same as (3.3). From (3.4) we see that the scores (3.3) (or the functionals (1.5)) compare the MEDF distribution to the null hypothesis distribution by taking the difference of the expectations of various functions ψ_ℓ under the two distributions.

Instead of the MEDF, we could use the Kaplan-Meier estimate, and consider statistics of the form $\int_0^1 \psi_\ell(u)\, d\,\hat{w}_n(u)$. However, such statistics would not be efficient scores from parametric alternatives, so likelihood theory could not be exploited to find the asymptotic distributions. Below we explore in some detail the differences between the Kaplan-Meier estimate and the MEDF in the simple null hypothesis case.

As before, let $u_i = F_i(y_i)$ and assume the $U_i = F_i(T_i)$ are identically distributed with distribution function H . As discussed in Section I, for censored observations, the unobserved value $F_i(t_i)$ can be anywhere in the interval $(u_i , 1]$. Thus for an observation censored at y_i , we have $U_i \leq u$ with conditional probability

$$P[U_i \leq u \mid U_i > u_i] = [H(u) - H(u_i)] / [1 - H(u_i)] . \tag{3.5}$$

The contribution of a censored observation to the estimate of $H(u)$ should represent n^{-1} times the conditional probability (3.5). For the MEDF, H_n^*, this probability is estimated by computing it under the null hypothesis distribution, giving $n^{-1} I(u \geq u_i) \cdot (u - u_i)/(1 - u_i)$ for the contribution of a censored observation (see (1.4)). Efron (1967) showed that for the Kaplan-Meier estimator K_n, (3.5) is actually computed under the Kaplan-Meier estimator of the distribution; that is, the Kaplan-Meier estimator satisfies

$$K_n(u) = n^{-1} \sum_{i=1}^{n} I(u \geq u_i) \left\{ z_i + (1-z_i)[K_n(u) - K_n(u_i)]/[1 - K_n(u_i)] \right\} .$$

Efron called this property self-consistency.

The MEDF is always defined on the entire line, whereas the Kaplan-Meier estimate is not clearly defined beyond the largest censoring time when there are no larger failure times. The MEDF is exactly unbiased in finite samples under the null hypothesis (see (5.2)), whereas the Kaplan-Meier estimate is slightly biased. In Section V it is shown (see (5.4)) that

$$\mathrm{Var}_{H_o} [\sqrt{n}\, H_n^*(u)] = (1-u)^2\, n^{-1} \sum_{i=1}^{n} \int_o^{d_i} \overline{C}_i(y)\,[\overline{F}_i(y)]^{-2}\, d\,F_i(y)$$

where $d_i = \sup\{t : F_i(t) = u\}$, which reduces to

$$(1-u)^2 \int_o^{d} \overline{C}(y)\,[\overline{F}(y)]^{-2}\, d\,F(y)$$

in the identically distributed, homogeneous censorship case. In comparison, the asymptotic variance of the Kaplan-Meier estimate in this case, as given by Efron (1967) and Breslow and Crowley (1974), is

$$(1-u)^2 \int_o^{d} [\overline{C}(y)\ \overline{F}^{\,2}(y)]^{-1}\, d\,F(y) .$$

Since in general $\overline{C}(y) \longrightarrow 0$ as $y \longrightarrow \infty$, the Kaplan-Meier estimate will be much more variable in the right tail of the distribution.

On the other hand, since the mass for censored observations is spread out over the interval $(u_i, 1]$ to be consistent with the null hypothesis, the MEDF is not a true nonparametric estimate, and will not be a consistent estimate of the true distribution function when the null hypothesis is not true, unlike the Kaplan-Meier estimate. Thus the MEDF could not be used to construct confidence intervals on an unknown distribution function, for example. However, these facts do not make use of the MEDF unreasonable in testing goodness-of-fit. Because of the censoring, there is a loss of information about goodness-of-fit (as compared with a complete sample of the same size). The Kaplan-Meier estimate reflects this with a large variability in the right tail of the distribution. To the extent that information has been lost, the MEDF has been smoothed to be consistent with the null hypothesis, and has a corresponding reduction in its variability. Thus in the right tail, the MEDF has a small variance but is consistent with the null hypothesis, and the Kaplan-Meier estimate has a large variance and is therefore compatible with a wide class of distributions. It is difficult to say if one estimator better reflects the information in the sample about lack of fit. The variability of the estimators should be kept in mind when comparing plots of them, however.

IV. Quadratic Score Statistics

IV.1. The Asymptotic Distribution of the Scores, and Possibilities for Estimating Their Variance

The standard asymptotic likelihood result for the scores (3.3) is that subject to mild regularity conditions, under the null hypothesis (1.2),

$$n^{-\frac{1}{2}} \partial\ell(\hat{\gamma}_o, 0)/\partial\theta \xrightarrow{\mathscr{L}} N\left(0, I_{\theta\theta|\gamma}\right), \tag{4.1}$$

where $\xrightarrow{\mathscr{L}}$ denotes convergence in distribution, $I_{\theta\theta|\gamma} = I_{\theta\theta} - I_{\theta\gamma} I_{\gamma\gamma}^{-1} I_{\gamma\theta}$, and $I = \begin{pmatrix} I_{\gamma\gamma} & I_{\gamma\theta} \\ I_{\theta\gamma} & I_{\theta\theta} \end{pmatrix}$ is the average Fisher information per observation (that is, $I_{\theta_\ell\theta_k} = \lim_{n\to\infty} I^{(n)}_{\theta_\ell\theta_k}$, where $I^{(n)}_{\theta_\ell\theta_k} = n^{-1}\,\mathrm{Cov}[\partial\ell(\gamma,0)/\partial\theta_\ell,\ \partial\ell(\gamma,0)/\partial\theta_k]$, and so on). For a discussion of asymptotic likelihood results with censored data and references on the general theory see Kalbfleisch and Prentice, (1980), Section 3.4. To actually carry out the test, we can use the quadratic score statistic

$$Q_m = n^{-1}\,[\partial\ell(\hat{\gamma}_o,0)/\partial\theta]'\ \hat{V}^{-1}\,[\partial\ell(\hat{\gamma}_o,0)/\partial\theta]$$

which $\xrightarrow{\mathscr{L}} \chi^2_m$ under the null hypothesis when $\hat{V}$ is any consistent estimate of $I_{\theta\theta|\gamma}$.

A natural way to estimate the asymptotic variance $I_{\theta\theta|\gamma}$ is with the average expected information in the sample, $I^{(n)}_{\theta\theta|\gamma} = I^{(n)}_{\theta\theta} - I^{(n)}_{\theta\gamma}\, {I^{(n)}_{\gamma\gamma}}^{-1}\, I^{(n)}_{\gamma\theta}$, evaluating (γ,θ) at $(\hat{\gamma}_o, 0)$.

Formulas for the components of $I^{(n)}_{\theta\theta|\gamma}$ under the heterogeneous censoring model are

$$n\, I^{(n)}_{\theta_\ell \theta_k} = n\, \mathrm{Cov}[\psi_\ell(\mathcal{U})\,,\, \psi_k\,(\mathcal{U})]$$

$$- \sum_{i=1}^{n} \int \mathrm{Cov}[\psi_\ell(\mathcal{U}),\psi_k(\mathcal{U})\,|\,\mathcal{U} > F_i(y;\gamma)]\, \overline{F}_i(y;\gamma)\, d\, C_i(y)$$

$$= \sum_{i=1}^{n} \int s_{i\ell}\,(y;\gamma)\; s_{ik}\,(y;\gamma)\; \overline{C}_i(y)\, d\, F_i\,(y;\gamma)\;, \qquad (4.2)$$

$$n\, I^{(n)}_{\theta_\ell \gamma_j} = \sum_{i=1}^{n} \Big\{ \mathrm{Cov}[\partial \log f_i(T_i;\gamma)/\partial\gamma_j\;,\; \psi_\ell[F_i(T_i\;;\gamma)]]$$

$$- \int \mathrm{Cov}[\partial \log f_i(T_i;\gamma)/\partial\gamma_j,\; \psi_\ell[F_i(T_i;\gamma)]\,\big|\,T_i > y]\,\overline{F}_i(y;\gamma)\, dC_i(y) \Big\}$$

$$= \sum_{i=1}^{n} \int s_{i\ell}\,(y;\gamma)\; q_{ij}(y;\gamma)\; \overline{C}_i(y)\; d\, F_i(y;\gamma)\;, \qquad (4.3)$$

and

$$n\, I^{(n)}_{\gamma_j \gamma_k} = \sum_{i=1}^{n} \Big\{ \mathrm{Cov}[\partial \log\; f_i(T_i;\gamma)/\partial\gamma_j\;,\; \partial \log\; f_i(T_i;\gamma)/\partial\gamma_k]$$

$$- \int \mathrm{Cov}[\partial \log\; f_i(T_i;\gamma)/\partial\gamma_j\;,\; \partial \log\; f_i(T_i;\gamma)/\partial\gamma_k\;|T_i > y]$$

$$\overline{F}_i(y;\gamma)\; d\, C_i(y) \Big\}$$

$$= \sum_{i=1}^{n} \int q_{ij}(y;\gamma)\; q_{ik}(y;\gamma)\; \overline{C}_i(y)\; d\, F_i(y;\gamma)\;, \qquad (4.4)$$

where $s_{i\ell}(y;\gamma) = \psi_\ell[F_i(y;\gamma)] - E[\psi_\ell(U) \mid U > F_i(y;\gamma)]$, $q_{ij}(y;\gamma) = \partial\{\log[f_i(y;\gamma)/\overline{F}_i(y;\gamma)]\}/\partial\gamma_j$, and as before U has a uniform distribution on $(0,1)$. Derivations of these formulas are given in the next section. In each case the first expression is in terms of the expected information when there is no censoring minus a correction for the censoring, and the second expression may be easier to use for computational purposes. This is particularly true of survival models, since they are often formulated in terms of the hazard function, $f_i/\overline{F}_i$, which appears in q_{ij} .

These expected information terms depend on the censoring distributions C_i . In some cases censoring arises in some systematic way and there is some prior knowledge of the distributional form, so there are censoring distributions, possibly with parameters estimated from the data, available to use in computing the expected information. Also, if there is little censoring, any reasonable censoring model should give similar results. In this case choice of a parametric censoring distribution could be guided by ease of computation. Unfortunately, outside of these special cases assumption of a particular parametric form for the censoring distribution will not be reasonable.

Censoring can arise from a variety of sources. For example, in clinical trials where the failure times are times to death from a particular cause, some possible sources of censoring are the subject dying from an unrelated cause, the subject moving away from the study area, and the experiment being stopped before the subject has died. In general, little is known about the distributions of these events, and adequately representing them can be a difficult problem. Further, the chances of these events might not be the same for all subjects in the study, which would require different distributions for different subjects (although it is assumed here that factors affecting the censoring have no influence on the failures). Because of the difficulties in representing the censoring, it is desirable to find a conditional reference set that eliminates the need to specify censoring distributions. In some special cases the potential censoring times $v_1, \ldots, v_n$ are known for all subjects. For example, if the only source of censoring is the experiment being stopped before all subjects have died, then the potential censoring times are just the times from entry into the experiment until the end of the experiment, which are known for all subjects. In this case we can condition on these potential censoring times by taking $C_i(y) = I(y \geq v_i)$ in computing the expected information. If the potential censoring times for uncensored observations are not known, one possibility for a conditional approach is to compute the expected information using $C_i(y) = I(y \geq v_i)$ for censored observations and $C_i(y) \equiv 0$ for uncensored observations. In effect this assumes that uncensored observations could not have been censored. We call this the "totally conditional" model. It is shown in Section VI that this can be very conservative.

Even if the censoring distributions are not exactly the same for all subjects, it might still be a reasonable approximation to assume the censoring is homogeneous. In this case (or if the censoring actually is homogeneous), rather than specify the censoring distribution we can use the Kaplan-Meier estimator to estimate the unknown censoring distribution from the data. One minor difficulty with this approach is that the Kaplan-Meier estimate of the censoring distribution does not reduce to 0 if the largest observation is a failure time. We suggest the slightly conservative approach of leaving the unassigned mass at $+\infty$.

In addition to the difficulty in choosing a censoring model, another problem with the use of the expected information is that the integrals in (4.2)-(4.4) can be very difficult to evaluate. An alternate approach is to use simulations, generating failures under the parametric maximum likelihood estimate of the failure distributions and the censoring times from any of the above censoring models, to compute the actual finite sample variance of the scores.

In using the expected information, there is also the deeper issue of whether the inference should depend on the possibility that uncensored observations could have been censored. To evaluate the expected information some assumption about the nature of the potential censoring for uncensored observations must be made. Even in the totally condition model, where the assumption is that uncensored observations would never be censored, an assumption is still being made about the nature of the potential censoring for uncensored observations. A procedure which avoids making such assumptions, and also avoids the difficulty involved in giving a reasonable parametric model for the censoring, is to use the observed information (the negative of the second derivatives of the log likelihood function) to get an estimate of $I_{\theta\theta|\gamma}$.

The log likelihood (3.1) is the same under any censoring scheme that is independent and noninformative (see Kalbfleisch and Prentice, 1980, section 5.2), and does not involve the censoring distributions.

Thus in computing the second derivatives of the log likelihood, it is not necessary to specify censoring distributions, or even to make assumptions about whether uncensored observations could have been censored. Under the null hypothesis the observed information does provide a consistent estimate of the expected information, subject to mild regularity conditions, so the adjusted observed information

$$- \ddot{\ell}_{\theta\theta|\gamma} = - \ddot{\ell}_{\theta\theta} + \ddot{\ell}_{\theta\gamma} \ddot{\ell}_{\gamma\gamma}^{-1} \ddot{\ell}_{\gamma\theta} \quad ,$$

where $\ddot{\ell}_{\theta_\ell \theta_k} = \partial^2 \ell(\hat{\gamma}_o , 0)/\partial\theta_\ell \partial\theta_k$ and so on, could be used for $n\hat{V}$ in Q_m. This gives a procedure where the inference will be the same under any independent and noninformative censoring model that is compatible with the data.

Although from a conceptual standpoint use of the observed information is ideal, in practice there can be serious difficulties. In computing the observed information, the derivatives are evaluated at $(\hat{\gamma}_o , 0)$. This point is usually not the global maximum of $\ell(\gamma,\theta)$, so if $\ell(\gamma,\theta)$ is not concave everywhere, the observed information need not even be positive definite at $(\hat{\gamma}_o ,0)$. In Section VI.2 it is shown that this does happen with the Neyman smooth test.

Another possible approach to estimating the variance would be to jackknife the scores. However, nonparametric estimation of the variance through such methods as the jackknife and nonparametric bootstrap does not seem appropriate here. These methods estimate the variance under the empirical distribution of the data, but it is preferable to use the variance under the null hypothesis distribution. There are other difficulties with these methods.

In the ordinary jackknife observations are deleted one at a time, or in groups, and the statistic recomputed from the remaining observations. The variability between these recomputed statistics is then used to estimate the variance of the original statistic. In the uncensored, identically distributed case the jackknife has been discussed in considerable detail. See Miller (1974) for a review. With regression data the jackknife can perform poorly (see Hinkley, 1977). It is thought that by deleting observations in groups, where there is a certain degree of balance in the covariables between the groups, will improve the performance somewhat. It is not known what effect censoring might have on the jackknife, but it may again be necessary to delete the observations in groups, where the censoring is balanced between the groups.

In the bootstrap, introduced by Efron (1979), the variance of the statistic is actually computed under the empirical distribution of the data, often by using simulations. Efron (1979) also gives a method of bootstrapping regression data (in the uncensored case), and explains why it is generally better than the jackknife. This method, in the context of the general regression models considered here, is to use the empirical distribution $\hat{H}_n(u)$ of the residuals $\hat{u}_i = F_i(t_i;\hat{\gamma})$, and compute the variance assuming the distribution of T_i is that of $F_i^{-1}(U;\hat{\gamma})$, where $U \sim \hat{H}_n$. Efron (1981) has considered methods for bootstrapping censored data in the identically distributed case. His approach with randomly censored data is to assume the censoring is homogeneous and then use the Kaplan-Meier estimates of both the failure and censoring distributions. Thus to use this method with censored data it is necessary to make some assumptions about the nature of the censoring. Also, since it would be necessary to use simulations in using the bootstrap to estimate the variance of the scores (3.3), it would be no simpler than the conceptually preferred method of using the parametric estimates of the failure distributions in the simulations.

Even if all these difficulties could be overcome, there can still be a serious problem with the size of the test using the jackknife and the bootstrap, even in the uncensored, identically distributed case. In Section VI.2 we show that the size of the Neyman smooth test for exponentiality is substantially larger than the nominal level when the jackknife or the bootstrap is used to estimate the variance, in the uncensored, identically distributed case.

IV.2. Expected Information Calculations

In this section details are given for the derivation of the expected information formulas (4.2)-(4.4). For convenience we suppress the dependence of F_i on γ .

Under the heterogeneous censoring model introduced in Section I, if $E \mid g(Y_i , Z_i) \mid < \infty$, then

$$E[g(Y_i , Z_i)] = E[g(Y_i , 0) \mid Z_i = 0] \, P[Z_i = 0] + E[g(Y_i , 1) \mid Z_i = 1] \, P[Z_i = 1] \tag{4.5}$$

$$= \int_0^\infty g(y,0) \, \overline{F}_i(y) \, d\, C_i(y) + \int_0^\infty g(y,1) \, \overline{C}_i(y) \, d\, F_i(y) \, .$$

Setting $a_{i\ell}(y) = \psi_\ell[F_i(y)]$, $b_{i\ell}(y) = E[\psi_\ell(U) \mid U > F_i(y)]$ and

$r_{i\ell k}(y) = E[\psi_\ell(U)\ \psi_k(U) \mid U > F_i(y)]$, then from (3.3) and (4.5) we have

$$n\, I^{(n)}_{\theta_\ell \theta_k} = \mathrm{Cov}[\partial\mathcal{L}(\gamma,0)/\partial\theta_\ell,\ \partial\mathcal{L}(\gamma,0)/\partial\theta_k]$$

$$= \sum_{i=1}^{n} \Big\{ E[Z_i\ a_{i\ell}(Y_i)\ a_{ik}(Y_i) + (1-Z_i)\ b_{i\ell}(Y_i)\, b_{ik}(Y_i)] - E[\psi_\ell(U)]\ E[\psi_k(U)] \Big\} \qquad (4.6)$$

$$= \sum_{i=1}^{n} \Big(E\Big\{ Z_i\ a_{i\ell}(Y_i)\ a_{ik}(Y_i) + (1-Z_i)\ r_{i\ell k}(Y_i) - (1-Z_i)[r_{i\ell k}(Y_i) - b_{i\ell}(Y_i)\, b_{ik}(Y_i)] \Big\} - E[\psi_\ell(U)]\ E[\psi_k(U)] \Big)$$

$$= \sum_{i=1}^{n} \Big\{ \int a_{i\ell}(y)\ a_{ik}(y)\overline{C}_i(y)\, d\,F_i(y) + \int r_{i\ell k}(y)\overline{F}_i(y)\, d\,C_i(y) - E[\Psi_\ell(U)]\, E\,[\psi_k(U)] - \int [r_{i\ell k}(y) - b_{i\ell}(y)\, b_{ik}(y)]\,\overline{F}_i(y)\, d\,C_i(y) \Big\}. \qquad (4.7)$$

Now

$$\int_0^\infty r_{i\ell k}(y)\ \overline{F}_i(y)\, d\,C_i(y) = \int_0^\infty \Big\{ [\overline{F}_i(y)]^{-1} \int_{F_i(y)}^{1} \psi_\ell(u)\psi_k(u)\, du \Big\}\ \overline{F}_i(y)\, d\,C_i(y)\ .$$

Applying change of variables to the inner integral and Fubini's theorem gives

$$= \int_0^\infty a_{i\ell}(t)\; a_{ik}(t)\; C_i(t)\, d\, F_i(t)\ .$$

Thus the sum of the first two integrals in (4.7) is $\int a_{i\ell}(y)\; a_{ik}(y)\, d\, F_i(y) = E[\psi_\ell(\mathcal{U})\; \psi_k(\mathcal{U})]$. Also $r_{i\ell k}(y) - b_{i\ell}(y)\, b_{ik}(y)$ $= \mathrm{Cov}[\psi_\ell(\mathcal{U})\; \psi_k(\mathcal{U}) \mid \mathcal{U} > F_i(y)]$. Making these substitutions in (4.7) gives the first expression in (4.2).

On the other hand, from (4.5) and (4.6),

$$n\, I^{(n)}_{\theta_\ell \theta_k} = \sum_{i=1}^{n} \left\{ \int a_{i\ell}(y)\; a_{ik}(y)\; \overline{C}_i(y)\, d\, F_i(y) \right. \tag{4.8}$$
$$\left. + \int b_{i\ell}(y)\; b_{ik}(y)\; \overline{F}_i(y)\, d\, C_i(y) - E[\psi_\ell(\mathcal{U})]\; E[\psi_k(\mathcal{U})] \right\}\ .$$

Now

$$\int b_{i\ell}(y)\; b_{ik}(y)\; \overline{F}_i(y)\, d\, C_i(y)$$
$$= \int\; [\overline{F}_i(y)]^{-1} [\int_{F_i(y)}^{1} \psi_\ell(u)\, du\ ]\ [\int_{F_i(y)}^{1} \psi_k(u)\, du\ ]\ d\, C_i(y)$$
$$= E[\psi_\ell(\mathcal{U})]E[\psi_k(\mathcal{U})] - \int [a_{i\ell}(y) b_{ik}(y) + a_{ik}(y) b_{i\ell}(y) - b_{ik}(y) b_{i\ell}(y)]$$
$$\overline{C}_i(y)\, d\, F_i(y)\ ,$$

after an integration by parts. Substituting this in (4.8) gives

$$n\, I^{(n)}_{\theta_\ell \theta_k} = \sum_{i=1}^{n} \int [a_{i\ell}(y) - b_{i\ell}(y)]\ [a_{ik}(y) - b_{ik}(y)]\ \overline{C}_i(y)\, d\, F_i(y)\ ,$$

which is the same as the second expression in (4.2), since $s_{i\ell}(y) = a_{i\ell}(y) - b_{i\ell}(y)$.

The derivations of (4.3) and (4.4) are very similar to the above, and will not be given here.

V. Further Properties of the MEDF

In this section further consideration is given to the properties of the MEDF, $\hat{H}_n^*$, defined in (1.4). Note that if we take $\psi_\ell(x) = I(a_\ell \geq x)$ in (3.1), then from (3.3),

$$n^{-\frac{1}{2}} \partial \mathcal{L}(\hat{\gamma}_o, 0)/\partial\theta_\ell = n^{-\frac{1}{2}} \sum_{i=1}^{n} \Big\{ z_i \; I(a_\ell \geq \hat{u}_i)$$

$$+ (1-z_i) \; E[I(a_\ell \geq U) | U > \hat{u}_i] - E[I(a_\ell \geq U)] \Big\}$$

$$= n^{-\frac{1}{2}} \sum_{i=1}^{n} \Big\{ I(a_\ell \geq \hat{u}_i)[z_i + (1-z_i)(a_\ell - \hat{u}_i)/(1-\hat{u}_i)] - a_\ell \Big\}$$

$$= \sqrt{n} \; [\hat{H}_n^* (a_\ell) - a_\ell] = \hat{y}_n^*(u) . \tag{5.1}$$

This connection between the scores (3.3) and the MEDF will be exploited in the following, where we will examine the distribution of the MEDF under the null hypothesis, first giving the mean and covariance of H_n^* in the simple null hypothesis case, and then the asymptotic structure of $\hat{y}_n^*$ in the composite null hypothesis case. The dependence of the distribution function on γ will usually be suppressed.

Let a_1 and a_2 be fixed constants, with $0 \leq a_1 \leq a_2 \leq 1$, and set $d_{i\ell} = \sup\{y : F_i(y) \leq a_\ell\}$. Note that since F_i is continuous, $F_i(d_{i\ell}) = a_\ell$. Then using (4.5), the expected value of $H_n^*(a_\ell)$ under the null hypothesis is

$$E[H_n^*(a_\ell)] = n^{-1} \sum_{i=1}^{n} E\left\{ I(a_\ell \geq u_i)[z_i + (1-z_i)(a_\ell - u_i)/(1-u_i)] \right\}$$

$$= n^{-1} \sum_{i=1}^{n} \left\{ \int I(a_\ell \geq F_i(y))\, \overline{C}_i(y)\, d\, F_i(y) \right.$$

$$\left. + \int I(a_\ell \geq F_i(y))[a_\ell - F_i(y)]\, d\, C_i(y) \right\} .$$

Now $I(a_\ell \geq F_i(y)) = I(d_{i\ell} \geq y)$, so

$$E[H_n^*(a_\ell)] = n^{-1} \sum_{i=1}^{n} \left\{ \int_0^{d_{i\ell}} \overline{C}_i(y)\, d\, F_i(y) + \int_0^{d_{i\ell}} [a_\ell - F_i(y)]\, d\, C_i(y) \right\}.$$

Applying an integration by parts to the second integral gives

$$E[H_n^*(a_\ell)] = n^{-1} \sum_{i=1}^{n} \left\{ \int_0^{d_{i\ell}} \overline{C}_i(y)\, d\, F_i(y) + a_\ell - \int_0^{d_{i\ell}} \overline{C}_i(y)\, d\, F_i(y) \right\}$$

$$= a_\ell . \tag{5.2}$$

Thus, under the null hypothesis, H_n^* is exactly unbiased.

For the covariance, note that $\mathrm{Cov}[\sqrt{n}\, H_n^*(a_1),\ \sqrt{n}\, H_n^*(a_2)] = \mathrm{Cov}[y_n^*(a_1)\ ,\ y_n^*(a_2)] = n^{-1}\, \mathrm{Cov}[\partial\ell(\gamma,0)/\partial\theta_1\ ,\ \partial\ell(\gamma,0)/\partial\theta_2]$, from (5.1), where again $\psi_\ell(x) = I(a_\ell \geq x)$ in the scores. Then from (4.2) we have

$$\mathrm{Cov}[\sqrt{n}\, H_n^*(a_1),\ \sqrt{n}\, H_n^*(a_2)] = n^{-1} \sum_{i=1}^{n} \int s_{i1}(y)\, s_{i2}(y)\, \overline{C}_i(y)\, d\, F_i(y) ,$$

where

$$s_{i\ell}(y) = I(a_\ell \geq F_i(y)) - E[I(a_\ell \geq U) | U > F_i(y)]$$

$$= I(a_\ell \geq F_i(y))\ (1-a_\ell)/\,\overline{F}_i(y) . \tag{5.3}$$

Also, $I(a_1 \geq F_i(y))\ I(a_2 \geq F_i(y)) = I(a_1 \geq F_i(y))$, since $a_1 \leq a_2$. Thus

$$\mathrm{Cov}[\sqrt{n}\ H_n^*(a_1)\ ,\ \sqrt{n}\ H_n^*(a_2)] = (1-a_1)(1-a_2)n^{-1} \sum_{i=1}^{n} \int_0^{d_{i1}} [\overline{F}_i(y)]^{-2}$$

$$\overline{C}_i(y)\, d\, F_i(y)\ . \tag{5.4}$$

To investigate the asymptotic structure of the process $\hat{y}_n^*(u)$ in the composite null hypothesis case, we use the relationship (5.1) and asymptotic likelihood theory, but give no proofs. For fixed constants $a_1, \ldots, a_n$, $0 \leq a_i \leq 1$, we have, from (5.1), $\hat{y}_n^*(a_\ell) = n^{-\frac{1}{2}}\ \partial\ell(\hat{\gamma}_o, 0)/\partial\theta_\ell$, where again $\psi_\ell(x) = I(a_\ell \geq x)$ in the scores. So if conditions are such that the asymptotic likelihood theory holds, then the finite dimensional distributions of $\hat{y}_n^*(u)$ converge weakly to $N\left(0\ ,\ I_{\theta\theta|\gamma}\right)$ random vectors (see (4.1)), under the null hypothesis. Thus if conditions are also such that the sequence of measures is tight (see Billingsley, 1968), then $\hat{y}_n^*$ converges weakly to a mean 0 Gaussian process $\hat{y}^*$ with $\mathrm{Cov}[\hat{y}^*(a_1)\ ,\ \hat{y}^*(a_2)] = I_{\theta_1\theta_2|\gamma} = \lim_{n\to\infty} \left[I_{\theta_1\theta_2}^{(n)} - I_{\theta_1\gamma}^{(n)}\left(I_{\gamma\gamma}^{(n)}\right)^{-1} I_{\gamma\theta_2}^{(n)}\right]$. For $a_1 \leq a_2$, $I_{\theta_1\theta_2}^{(n)}$ is given by (5.4). From (4.3) and (5.3) we have

$$I_{\theta_\ell\gamma_j}^{(n)} = (1-a_\ell)\, n^{-1} \sum_{i=1}^{n} \int_0^{d_{i\ell}} q_{ij}(y)\ \overline{C}_i(y)\ [\overline{F}_i(y)]^{-1}\, d\, F_i(y)\ , \tag{5.5}$$

$$= n^{-1} \sum_{i=1}^{n} \left\{ -\overline{C}_i(d_{i\ell})\,\partial\,\overline{F}_i(d_{i\ell})/\partial\gamma_j \right.$$

$$\left. -\ (1-a_\ell) \int_0^{d_{i\ell}} \partial[\log \overline{F}_i(y)]/\partial\gamma_j\ d\, C_i(y) \right\}\ ,$$

after an integration by parts. $I^{(n)}_{\gamma_j\gamma_k}$ is as given in (4.4). In the uncensored case $\overline{C}_i(y) \equiv 1$, and these expressions reduce to the corresponding expressions for $\hat{y}$ as given by Pierce and Kopecky (1979) and Loynes (1980).

We could use other functionals of $\hat{y}^*_n$ than the scores (3.3) to test the hypothesis (1.2), but likelihood theory could not be used to find their asymptotic distributions. These asymptotic distributions would be difficult to find, and the following simple example shows that in general the covariance of $\hat{y}^*$ does depend on the covariables, the true values of the parameters and the censoring distributions, so the asymptotic distributions would essentially be different for every problem.

Suppose $T_1, \ldots, T_n$ are independent with the distribution function of T_i given by $1 - \exp\{-t \cdot \exp(\beta_0 x_{i0} + \beta_1 x_{i1})\}$, where $x_{i0} = 1$ for all i, $x_{i1} = 0$ for $i = 1, \ldots, n_1$, and $x_{i1} = 1$ for $i = n_1 + 1, \ldots, n$ (the two sample exponential problem). Also suppose the data are type I censored at T. Note that since this is a location family in $\log T_i$, the Pierce-Kopecky result would hold here if there were no censoring. We will compute $\mathrm{Cov}[\hat{y}^*(a_1), \hat{y}^*(a_2)]$, where $a_1 \leq a_2 < 1 - \exp\{-T \cdot \exp(x_i'\beta)\}$ for all i, where $x_i'\beta = x_{i0}\beta_0 + x_{i1}\beta_1$. From this restriction we have that $d_{i\ell} = [-\log(1-a_\ell)] / \exp(x_i'\beta) < T$ for all i and ℓ, so $\int_0^{d_{i\ell}} g(y)\, \overline{C}_i(y)\, dF_i(y) = \int_0^{d_{i\ell}} g(y)\, dF_i(y)$, since $\overline{C}_i(y) = 1$ for $y < T$. Thus from (5.4), making the change of variables $u = F_i(y)$, we have

$$I^{(n)}_{\theta_1\theta_2} = (1-a_1)(1-a_2)\ n^{-1} \sum_{i=1}^{n} \int_0^{a_1} (1-u)^{-2}\, du = a_1(1-a_2) \tag{5.6}$$

For this exponential model the hazard function $f_i/\overline{F}_i = \exp(x_i'\beta)$, so $q_{ij}(y) = \partial \log[f_i(y) / \overline{F}_i(y)]/\partial\beta_j = x_{ij}$, $j = 0,1$. Thus from (5.5), for $j = 0,1$,

$$I^{(n)}_{\theta_\ell\beta_j} = (1-a_\ell)\, n^{-1} \sum_{i=1}^{n} \int_0^{d_{i\ell}} x_{ij}\, [\,\overline{F}_i(y)\,]^{-1}\, d\, F_i(y)$$

$$= (1-a_\ell)\ [\,-\log(1-a_\ell)\,]\ \ n^{-1} \sum_{i=1}^{n} x_{ij}\ .$$

Note that $\sum_{i=1}^{n} x_{i0}/n = 1$ and $\sum_{i=1}^{n} x_{i1}/n = n_2/n$, where $n_2 = n-n_1$, so

$$I^{(n)}_{\theta_\ell\beta} = -\ (1-a_\ell)\ \log\ (1-a_\ell)\ (\,1\ ,\ n_2/n)\ . \tag{5.7}$$

Also,

$$I^{(n)}_{\beta_j\beta_k} = n^{-1} \sum_{i=1}^{n} \int_0^{T} x_{ij}x_{ik}\, d\, F_i(y)$$

$$= n^{-1} \sum_{i=1}^{n_1} x_{ij}x_{ik}b_0 + n^{-1} \sum_{i=n_1+1}^{n} x_{ij}x_{ik}b_1\ ,$$

where $b_0 = 1 - \exp\{-T\cdot\exp(\beta_0)\}$ and $b_1 = 1 - \exp\{-T\cdot\exp(\beta_0+\beta_1)\}$. The expression on the left is 0 unless $j = k = 0$, in which case it is n_1b_0/n . The expression on the right is n_2b_1/n for any j and k . Thus

$$I^{(n)}_{\beta\beta} = n^{-1} \begin{bmatrix} n_1b_0 + n_2b_1 & n_2b_1 \\ n_2b_1 & n_2b_1 \end{bmatrix}\ ,$$

and

$$I_{\beta\beta}^{(n)^{-1}} = n \begin{bmatrix} (n_1 b_0)^{-1} & -(n_1 b_0)^{-1} \\ -(n_1 b_0)^{-1} & (n_1 b_0)^{-1} + (n_2 b_1)^{-1} \end{bmatrix} \tag{5.8}$$

Thus from (5.7) and (5.8),

$$I_{\theta_1\beta}^{(n)} \left(I_{\beta\beta}^{(n)} \right)^{-1} I_{\beta\theta_2}^{(n)} =$$

$$(1-a_1)\ (1-a_2)\ \log(1-a_1)\ \log(1-a_2)\ [n_1/(nb_0) + n_2/(nb_1)]. \tag{5.9}$$

If $p = \lim_{n\to\infty} n_1/n$, then from (5.6) and (5.9) ,

$$\mathrm{Cov}[\hat{y}^*(a_1)\ ,\ \hat{y}^*(a_2)] = \lim_{n\to\infty} I_{\theta_1\theta_2|\beta}^{(n)} =$$

$$a_1(1-a_2) - (1-a_1)(1-a_2)\ \log(1-a_1)\ \log(1-a_2)[p/b_0 + (1-p)/b_1] \tag{5.10}$$

Clearly this does depend on the parameters, the covariables, and on the censoring time T . As $T \to \infty$ (that is, as the probability of censoring $\to 0$) , (5.10) approaches

$$a_1(1-a_2) - (1-a_1)(1-a_2)\log(1-a_1)\log(1-a_2) ,$$

which can be seen from Loynes' (1980) result to be the covariance of $\hat{y}(u)$ for this example when there is no censoring. Since this does not depend on the covariable x_{i1} or on the parameters β , this confirms the Pierce-Kopecky result for this special case.

In this example, with type I censoring, and a_1 and a_2 chosen so that $d_{i\ell} < T$ for all i and ℓ, all the $\hat{u}_i$ that are $\leq a_2$ correspond to uncensored observations. Thus at a_1 and a_2, the only difference between $\hat{y}_n^*$ and what we would have had without any censoring is that the maximum likelihood estimate, $\hat{\gamma}_o$, is computed from the censored data. Comparing (5.6), (5.7) and (5.8) with the corresponding expressions for the uncensored case, which can easily be computed from Loynes' result, the only difference is in (5.8), the inverse of the information for the estimation of β. Thus if we had computed the covariance in this example as given by Loynes' result, replacing the expected information for the estimation of the parameters by its censored version, we would still get the same result. This agrees with a result of Pettit (1976), generalizing Durbin's (1973) result on the asymptotic distribution of $\hat{y}_n(u)$ in the identically distributed, uncensored case to type I and type II censoring. The only difference that Pettit finds between the censored and uncensored cases is also just in the expected information for the estimation of the parameters.

Over the range of a_1 and a_2 values considered in computing (5.10), the Kaplan-Meier estimator and the MEDF are identical, so (5.10) should also give the asymptotic covariance of the Kaplan-Meier estimator for this example. If this is true, then this example also confirms the asymptotic dependence of the Kaplan-Meier estimator on the covariables, the true values of the parameters and the censoring distribution.

If the censoring is more general than the type I model used in this example, the terms $I_{\theta_1\theta_2}$ and $I_{\theta_\ell\gamma}$ will also involve the censoring distributions, and expressions for $\mathrm{Cov}[\hat{y}^*(a_1), \hat{y}^*(a_2)]$ become more complex.

VI. Some Specific Score Tests

VI.1. Chi-Square Tests

Rao and Robson (1974) develop a modification of the classical χ^2 test in the uncensored, identically distributed case which does have a limiting χ^2 distribution. If $0 = a_o < a_1 < \cdots < a_m < 1$, and $\psi_\ell(x) = I(a_{\ell-1} \leq x < a_\ell)$, $\ell = 1, \ldots, m$, then the score $\partial \mathcal{L}(\hat{\gamma}_o, 0)/\partial\theta_\ell$ in the uncensored case is simply the number of $\hat{u}_i$ in the interval $[a_{\ell-1}, a_\ell)$ minus the expected number in the interval under the null hypothesis. If the expected information calculations are carried out for the scores with these ψ_ℓ functions, then for the uncensored, identically distributed case the resulting quadratic score statistic Q_m is the same as the Rao-Robson test, provided we take $a_\ell = \ell/(m+1)$.

Using these same ψ_ℓ functions in the general setting provides a way of extending this test to both censored data and regression problems. With censoring the scores are

$$\partial \mathcal{L}(\hat{\gamma}_o, 0)/\partial\theta_\ell = \sum_{i=1}^{n} \Big\{ z_i I(a_{\ell-1} \leq \hat{u}_i < a_\ell) - (a_\ell - a_{\ell-1}) \\ + (1-z_i) I(a_\ell > \hat{u}_i)\ (a_\ell - \max\{a_{\ell-1}, \hat{u}_i\})/(1-\hat{u}_i) \Big\},$$

which is the number of $\hat{u}_i$ corresponding to failures in the interval $[a_{\ell-1}, a_\ell)$ plus the part of each censored observation that lies in the interval $[a_{\ell-1}, a_\ell)$ when its mass is spread uniformly over the interval $(\hat{u}_i, 1]$ minus the expected number of observations in the interval $[a_{\ell-1}, a_\ell)$. The information calculations are very similar to (5.4) and (5.5) (the scores there are based on $\psi_\ell(x) = I(a_\ell \geq x))$.

It takes a larger value of m to adequately model smooth alternatives with this test than with the Neyman smooth test (where $\psi_\ell(x) = x^\ell)$. Because of this, when both tests have modelled a reasonably smooth alternative equally well, so that the values of the test statistics are roughly the same, the above test will have a larger number of degrees of freedom than the Neyman smooth test. Thus the Neyman smooth test will generally be more powerful against smooth alternatives. We will not discuss the generalization of the Rao-Robson test further here, but instead move on to a discussion of the Neyman smooth test.

VI.2. Neyman Smooth Tests

Neyman (1937) proposed a goodness-of-fit test for a completely specified distribution which focuses power against "smooth" alternatives. Neyman's test is a score test of the type considered here, using a system of orthogonal polynomials for the ψ_ℓ functions. Durbin and Knott (1972) noted that the first m components in their expansion for the Anderson-Darling statistic give the m degree of freedom Neyman smooth test. Barton (1956) studied the limiting distribution of Neyman's statistic when the hypothesis is composite. Thomas and Pierce (1979) used the approach adopted here of modifying the test statistic to allow for the estimation of parameters, so that it still has a limiting χ^2 distribution. Bargal (1981) considered extensions to censored data.

Rather than use orthogonal polynomials, it is somewhat simpler here to take $\psi_\ell = x^\ell$, $\ell = 1, \ldots, m$. The value of m is to be chosen large enough to give a reasonably broad class of alternatives, but small enough to give good power against reasonably smooth alternatives. Thomas and Pierce suggest $m = 2$ may be best in the composite hypothesis setting. Their limited power studies also suggest the 2 degree of freedom test for normality may be reasonably competitive with other, possibly more specialized, commonly used tests.

It is slightly simpler here to take $U_i = \overline{F}_i(T_i;\gamma) = 1 - F_i(T_i;\gamma)$ in defining the alternatives (3.1) (these new U_i are still uniform on $(0,1)$ under the null hypothesis). The only change made in the development is that the condition $T_i > y_i$ is now equivalent to $\mathcal{U} = \overline{F}_i(T_i;\gamma) < \overline{F}_i(y_i;\gamma)$, so "$\mathcal{U} > \hat{u}_i$" in (3.3) needs to be replaced by "$\mathcal{U} < \overline{F}_i(y_i;\hat{\gamma}_o)$", and "$\mathcal{U} > F_i(y;\gamma)$" in $s_{i\ell}(y;\gamma)$ in (4.2) and (4.3) needs to be replaced by "$\mathcal{U} < \overline{F}_i(y;\gamma)$". Then from (3.3) the scores for the Neyman smooth test are

$$\partial\ell(\hat{\gamma}_o,0)/\partial\theta_\ell = \sum_{i=1}^{n} \left\{ z_i\, \overline{F}_i^{\,\ell}(y_i;\hat{\gamma}_o) + (1-z_i)\, \overline{F}_i^{\,\ell}(y_i;\hat{\gamma}_o)/(\ell+1) - (\ell+1)^{-1} \right\}, \tag{6.1}$$

since $E[\mathcal{U}^\ell \mid \mathcal{U} < k] = k^\ell/(\ell+1)$ and $E[\mathcal{U}^\ell] = (\ell+1)^{-1}$. Also, in the expected information components (4.2) and (4.3), the terms

$$\begin{aligned} s_{i\ell}(y;\gamma) &= \psi_\ell[\overline{F}_i(y;\gamma)] - E[\psi_\ell(\mathcal{U}) \mid \mathcal{U} < \overline{F}_i(y;\gamma)] \\ &= \ell\, \overline{F}_i^{\,\ell}(y;\gamma)/(\ell+1) . \end{aligned} \tag{6.2}$$

In examining possibilities for estimating $I_{\theta\theta|\gamma}$ for the Neyman smooth test, only the exponential model with $\bar{F}_i(t;\beta) = \exp\{-t\exp(x_i'\beta)\}$ will be considered in detail here. We use W_1 and W_2 to represent the one and two degree of freedom Neyman smooth tests for exponentiality (i.e., Q_m with $m = 1,2$), regardless of what estimate of $I_{\theta\theta|\gamma}$ is used. Below we discuss using the observed information, nonparametric methods, and the expected information computed under the censoring models discussed in Section IV, for computing the W_1 and W_2 tests.

To illustrate the difficulties with using the observed information, it is only necessary to examine the uncensored, identically distributed case. We generated 200 uncensored samples of size 50 from a unit exponential distribution and computed the observed information components for W_1 and W_2. In 70 of the samples, $-\ddot{\ell}_{\theta\theta|\gamma}$ for W_2 was not positive definite. In 3 of the samples $-\ddot{\ell}_{\theta_1\theta_1|\gamma}$, the variance estimate of the θ_1 score, was negative. In addition, for 13 of the 197 samples where $-\ddot{\ell}_{\theta_1\theta_1|\gamma} > 0$, the value of the W_1 statistic computed from this quantity was greater than the χ^2 critical value for a 1% level test. Similar results were found for the Neyman smooth test for normality.

The difficulty with the size of the test using the jackknife and the bootstrap can also be seen in the identically distributed, uncensored case. Table 1 gives empirical sizes for W_1 and W_2, using both the bootstrap variance estimate and the expected information, based on 500 samples of size 50 generated from a unit exponential distribution. The bootstrap estimates were determined by simulations, with 100 bootstrap replications for each sample. The binomial standard errors for comparing the empirical sizes to the nominal levels are given in parentheses after the nominal levels. In all cases with the bootstrap the empirical sizes are more than 5.8 standard errors above the nominal level. In all cases with the expected information the empirical sizes are within 1.9 standard errors of the nominal levels.

Table 1: Empirical Sizes (in %) of the W_1 and W_2 Tests Using the Bootstrap (B) and the Expected Information (EI)

		W_1		W_2	
Nominal Level	S.E.	B	EI	B	EI
10%	(1.34)	17.8	12.4	21.4	11.2
5%	(.97)	11.0	6.8	15.8	4.8
1%	(.44)	4.8	1.6	6.8	1.4

The bootstrap does a fairly good job of estimating the components of $I_{\theta\theta|\gamma}$ on the average. However, these estimates are quite variable, and there is a correlation between the variance estimates and the values of the scores. In particular, the correlation between the θ_1 score and its variance estimate is .639 . Thus samples for which the θ_1 score is a large negative quantity have the smallest bootstrap estimate of its variance, and samples where the θ_1 score is positive and large have the largest bootstrap variance estimate. This is reflected in Table 2 where the average value of the θ_1 score is given for the samples where the bootstrap estimate of variance falls in certain intervals, where the intervals were chosen so that each group has 100 samples. Thus the excessively large sizes for W_1 and W_2 using the bootstrap estimate of variance are the result of the excessively small variance estimate the bootstrap consistently gives for samples with a large negative value of the θ_1 score.

Table 2: Average Values of the θ_1 Score for Certain Ranges of the Bootstrap Variance Estimate

Interval of the Bootstrap Variance	Average Value of the θ_1 Score
(.3918, .6511]	-1.262
(.6511, .8203]	- .541
(.8203, .9511]	- .083
(.9511, 1.202]	.227
(1.202 , 2.558]	.777

The jackknife gives very similar results here. Empirical sizes of the W_1 and W_2 tests using the jackknife estimate of variance (computed by deleting observations one at a time) and the expected information, based on 300 samples of size 50 generated from a unit exponential distribution, are given in Table 3, with standard errors in parentheses after the nominal levels. In all cases the empirical sizes are within 1.6 standard errors of the nominal levels when the expected information is used. Using the jackknife the empirical sizes are more than 2.6 standard errors larger than the nominal level for the W_1 test, and more than 4.8 standard errors larger for the W_2 test.

Table 3: Empirical Sizes (in %) of the W_1 and W_2 Tests Using the Jackknife (J) and the Expected Information (EI)

		W_1		W_2	
Nominal Level	S.E.	J	EI	J	EI
10%	(1.73)	14.67	9.67	18.33	7.33
5%	(1.26)	9.0	5.0	12.67	3.67
1%	(.57)	2.67	1.0	5.33	1.33

The expected information is known to perform well in the uncensored case (see Thomas and Pierce, 1979, and Tables 1 and 3 here). Since censored observations contribute a conditional expectation term to the scores (3.3), we would guess that, for a fixed sample size, as the degree of censoring increases the variability of the scores will decrease. For the Neyman smooth test for exponentiality in the uncensored case, $I_{\theta_1\theta_1|\gamma} = 1/48 = .02083$. If the failures are exponential with rate λ_f and the (homogeneous) censoring is exponential with rate λ_c , then

$$I_{\theta_1\theta_1|\gamma} = (1-p_c)^3/[4(3-2\,p_c)(2-p_c)^2] ,$$

where p_c is the probability of an observation being censored, $p_c = \int_0^\infty \bar{F}\, dC = \lambda_c / (\lambda_f + \lambda_c)$. When $p_c = .2$, $I_{\theta_1\theta_1|\gamma} = .01519$, and when $p_c = .5$, $I_{\theta_1\theta_1|\gamma} = .00694$. Thus $I_{\theta_1\theta_1|\gamma}$ does decrease rapidly as the degree of censoring increases, so using the uncensored value of $I_{\theta_1\theta_1|\gamma}$ would only be appropriate if there is very little censoring.

Table 4 contains the results of simulations to examine the size of the W_1 and W_2 tests in the identically distributed case, using the expected information computed under different censoring models. Binomial standard errors are given in parentheses after the nominal levels for the first case. The same standard errors can be used for the other cases. In the simulations a potential censoring time is generated for each observation. In the "conditional" model the expected information is computed conditional on all the potential censoring times. In the "homogeneous exponential" model the censoring is assumed to be homogeneous with $\bar{C}(v) = \exp(-\lambda v)$. The value of λ is estimated from the data. Under the "Kaplan-Meier" model the censoring is assumed to be homogeneous and the censoring distribution is estimated with the Kaplan-Meier estimator. In the "totally conditional" model the expected information is computed as discussed in Section IV.1. Details of the expected information calculations can be found in Section VI.4.

In the simulations, failure times were generated from the unit exponential distribution. Homogeneous censoring times were generated from different Weibull distributions with $\bar{C}(v) = \exp\{-(v/\alpha)^{\sigma}\}$. The values of the Weibull shape parameter used were $\sigma = 1$, .5, and 2. For each value of the shape parameter values of α were selected to make the probability of an observation being censored, p_c, equal to .2 and .5. In each of the six cases (3 values of σ x 2 values of p_c) we generated 1000 samples of size 50. In each case the 8 statistics (W_1 and W_2 with the expected information

Table 4: Empirical Sizes (in %) of the W_1 and W_2 Tests Using the Expected Information Computed Under Different Censoring Models

	Nominal Level	Conditional		Homogeneous Exponential		Kaplan-Meier		Totally Conditional	
		W_1	W_2	W_1	W_2	W_1	W_2	W_1	W_2
$\sigma=1$	10%(.95)	9.4	9.0	9.9	8.8	9.8	8.4	6.0*	5.7*
$p_c=.2$	5%(.69)	4.2	4.1	4.1	4.3	4.3	4.4	2.7*	2.3*
	1%(.31)	1.0	.6	1.0	.7	.9	.8	.3*	.3*
$\sigma=1$	10%	10.5	9.4	9.8	8.9	9.1	8.2	2.0*	2.4*
$p_c=.5$	5%	5.1	5.0	4.7	4.7	4.1	4.8	.3*	1.1*
	1%	1.0	1.5	.7	1.2	.9	1.1	0*	.2*
$\sigma=2$	10%	11.3	10.7	9.3	7.7*	11.3	9.6	6.2*	5.8*
$p_c=.2$	5%	6.5*	5.1	4.8	3.5*	6.1	4.7	3.1*	1.7*
	1%	1.3	.8	.5	.5	1.0	.8	.2*	.2*
$\sigma=2$	10%	10.8	11.0	4.5*	4.3*	10.6	9.0	.4*	2.4*
$p_c=.5$	5%	6.0	6.1	1.9*	2.2*	5.3	4.6	.1*	1.1*
	1%	1.2	1.2	.2*	.7	.9	.8	0*	.1*
$\sigma=.5$	10%	7.0*	7.9*	8.2	9.7	7.1*	7.6*	6.5*	6.1*
$p_c=.2$	5%	3.9	3.0*	4.6	4.1	3.7	3.3*	3.2*	1.9*
	1%	.5	.5	.5	.8	.4	.6	.2*	.3*
$\sigma=.5$	10%	9.9	7.7*	15.3*	18.1*	9.6	8.0*	5.3*	4.2*
$p_c=.5$	5%	4.6	3.9	9.0*	11.4*	4.4	4.0	2.1*	1.8*
	1%	.6	.8	2.5*	3.2*	.8	1.2	.2*	.2*

*indicates the empirical size is more than 2 standard errors from the nominal level.

computed under 4 different censoring models) were computed from the same samples.

In all cases the Kaplan-Meier and conditional models performed well. When the censoring is not exponential ($\sigma = .5$ or 2) and $p_c = .5$, use of the homogeneous exponential censoring model can be quite bad. However, when $p_c = .2$ there seems to be little effect. The totally conditional model is very conservative. The reason for this is that it assumes uncensored observations would never be censored. Consequently, the expected number of censored observations under this model is substantially less than the actual number of censored observations in the sample. Since the variance of the scores decreases as the degree of censoring increases, this means the totally conditional model consistently overestimates the variance.

Of all the methods considered in this section for estimating $I_{\theta\theta|\gamma}$ for the Neyman smooth test for exponentiality, only using the expected information computed under the conditional and Kaplan-Meier censoring models has given reliable results in all the cases considered here. For the other methods, the observed information fails to even give a positive definite matrix in a substantial proportion of samples generated under the null hypothesis, the size of the tests using the jackknife and the bootstrap is considerably larger than the nominal level, use of the totally conditional model is overly conservative, and the size of the tests using the homogeneous exponential model can be both smaller and larger than the nominal level, depending on the true censoring. Of the two reliable methods, only the Kaplan-Meier model can always be computed, whether or not all the potential censoring times are known. The simulations reported in Table 4 were only under homogeneous censoring. How the Kaplan-Meier model will perform when the censoring is not homogeneous is not known. It is possible that when all potential censoring times are known, the conditional model is preferable, since it does not assume the censoring is homogeneous and it can be slightly easier to compute.

VI.3. A Note on Random Number Generation

Exponential or Weibull deviates were computed by generating a stream of uniform pseudo-random numbers and using the obvious probability integral transforms. To generate randomly censored data, two separate streams of uniforms were used, one for the censoring times and one for the failure times. The uniform pseudo-random number generators used were the congruential generators given in the first two columns of Table I of Downham and Roberts (1967). They perform a number of tests for uniformity and randomness and find that both generators are satisfactory.

For the nonparametric bootstrap, we need to simulate sampling from the distribution that places a mass of $1/n$ on each of the values $y_1, \ldots, y_n$. To do this we generated a stream of uniform deviates u, used the transformation $i = \mathrm{INT}(u \cdot n+1)$, where $\mathrm{INT}(x)$ is the integer part of x, and then took the value of the corresponding observation to be y_i.

VI.4. Expected Information Calculations for the Neyman Smooth Test

In this section details are given for computing the expected information for the Neyman smooth test under the censoring models examined in Section VI.2, with complete details given for the exponential and Weibull regression models.

The Kaplan-Meier estimate of the censoring distribution is

$$\hat{\bar{C}}(y) = \prod_{i:y_i \le y} [(n-r_i)/(n-r_i+1)]^{1-z_i}$$

where r_i is the rank of (y_i, z_i) in the lexicographic ordering of the sequence $(y_1, z_1), \ldots, (y_n, z_n)$. If this estimate does not become 0 at some point, we do not set it equal to 0 beyond the largest observation, as is sometimes done, but leave the unassigned mass at $+\infty$. Let $y_{c_1}, \ldots, y_{c_{n_c}}$ denote the censoring times, where $n_c = \sum_{i=1}^{n} (1-z_i)$ is the number of censored observations. Also, let $\Delta\hat{C}(y_{c_h})$ denote the jump in $\hat{C}(y)$ at y_{c_h} and $\hat{\bar{C}}(\infty)$ the value of $\hat{\bar{C}}(y)$ beyond the largest censoring time.

By (4.2)-(4.4), under the Kaplan-Meier censoring model, the components of the expected information are of the form $\sum_{i=1}^{n} \int g_i(y;\gamma)\, \hat{\bar{C}}(y)\, d F_i(y;\gamma)$, which can be rewritten as

$$\sum_{i=1}^{n} \left\{ \int_0^{\infty} g_i(y;\gamma) d F_i(y;\gamma)\, \hat{\bar{C}}(\infty) + \sum_{h=1}^{n_c} \Delta \hat{C}(y_{c_h}) \int_0^{y_{c_h}} g_i(y;\gamma) d F_i(y;\gamma) \right\} .$$

For the Neyman smooth test, using this result together with (4.2)-(4.4) and (6.2), we have

$$n\, I^{(n)}_{\theta_\ell \theta_k} = \sum_{i=1}^{n} k\ell[(k+1)(\ell+1)]^{-1} \left\{ \int_0^\infty \bar{F}^{k+\ell}(y;\gamma)\, d\, F_i(y;\gamma) \cdot \hat{\bar{C}}(\infty) \right.$$

$$\left. + \sum_{h=1}^{n_c} \Delta\, \hat{C}(y_{c_h}) \int_0^{y_{c_h}} \bar{F}^{\,k+\ell}(y;\gamma)\, d\, F_i(y;\gamma) \right\}$$

$$= k\ell[(k+1)(\ell+1)(k+\ell+1)]^{-1} \sum_{i=1}^{n} \left\{ 1 - \sum_{h=1}^{n_c} \Delta\, \hat{C}(y_{c_h})\, \bar{F}_i^{k+\ell+1}(y_{c_h};\gamma) \right\} ,$$

$$n\, I^{(n)}_{\theta_\ell \gamma_j} = \ell(\ell+1)^{-1} \sum_{i=1}^{n} \left\{ \hat{\bar{C}}(\infty)\;\; P_{ij\ell}(\infty) + \sum_{h=1}^{n_c} \Delta\, \hat{C}(y_{c_h})\, P_{ij\ell}(y_{c_h}) \right\}, \qquad (6.3)$$

and

$$n\, I^{(n)}_{\gamma_j \gamma_k} = \sum_{i=1}^{n} \left\{ \hat{\bar{C}}(\infty)\, Q_{ijk}(\infty) + \sum_{h=1}^{n_c} \Delta\, \hat{C}(y_{c_h})\, Q_{ijk}(y_{c_h}) \right\} , \qquad (6.4)$$

where

$$P_{ij\ell}(w) = \int_0^w q_{ij}(y;\gamma)\;\; \bar{F}_i^{\ell}(y;\gamma)\, d\, F_i(y;\gamma) \qquad (6.5)$$

and

$$Q_{ijk}(w) = \int_0^w q_{ij}(y;\gamma)\;\; q_{ik}(y;\gamma)\, d\, F_i(y;\gamma)\;. \qquad (6.6)$$

For the exponential model with $\bar{F}_i(y;\gamma) = \exp\{-y \cdot \exp(x_i'\beta)\}$, $q_{ij}(y;\gamma) = \partial \log[f_i(y;\gamma)/\bar{F}_i(y;\gamma)] / \partial\beta_j = x_{ij}$. Thus

$$P_{ij\ell}(y) = x_{ij}\;\; [1 - \exp\{-y(\ell+1)\exp(x_i'\beta)\}] / (\ell+1)$$

and

$$Q_{ijk}(y) = x_{ij}x_{ik}\;\; [1 - \exp\{-y \cdot \exp(x_i'\beta)\}]\;.$$

From these results, (6.3), (6.4) and the fact that $\hat{\bar{C}}(\infty) + \sum_{h=1}^{n_c} \Delta \hat{C}(y_{c_h}) = 1$, we have

$$n\, I^{(n)}_{\theta_\ell \gamma_j} = \ell(\ell+1)^{-2} \sum_{i=1}^{n} x_{ij} \, [1 - \sum_{h=1}^{n_c} \Delta \hat{C}(y_{c_h}) \exp\{ -y_{c_h} (\ell+1)\exp(x_i'\beta)\}]$$

and

$$n\, I^{(n)}_{\gamma_j \gamma_k} = \sum_{i=1}^{n} x_{ij} x_{ik} \, [1 - \sum_{h=1}^{n_c} \Delta \hat{C}(y_{c_h}) \exp\{ -y_{c_h} \exp(x_i'\beta)\}] ,$$

for the components of the expected information, computed under the Kaplan-Meier censoring model, for the Neyman smooth test for the exponential distribution.

Formulas for $P_{ij\ell}$ and Q_{ijk} are also easily found for the Weibull model. We have $\bar{F}_i(y;\gamma) = \exp\{ -a_i(y)\}$, where $a_i(y) = [\, y \cdot \exp(x_i'\, \beta)]^{\sigma}$, $\gamma_i = \beta_i$, $i = 1, \ldots, p$, and $\gamma_{p+1} = \sigma$. Then $f_i / \bar{F}_i = \sigma \exp(x_i'\, \beta\sigma)\, y^{\sigma-1}$, so

$$q_{ij}(y;\gamma) \begin{cases} \sigma x_{ij} & j = 1, \ldots, p \\ \sigma^{-1} (1 + \log[a_i(y)]) & j = p+1 \, . \end{cases} \tag{6.7}$$

Thus

$$P_{ij\ell}(w) = \sigma x_{ij} \, [1 - \exp\{ -(\ell+1) a_i(w)\}] / (\ell+1) , \qquad j=1,\ldots,p ,$$

and

$$Q_{ijk}(w) = \sigma^2 x_{ij} x_{ik} \, [1 - \exp\{ -a_i(w)\}] , \qquad j, k = 1, \ldots, p .$$

From (6.5) and (6.7)

$$P_{i,p+1,\ell}^{(w)} = \sigma^{-1} \int_0^w [1+\log a_i(y)] \exp\{-(\ell+1)a_i(y)\} [d\,a_i(y)/dy]\,dy$$

$$= [\sigma(\ell+1)]^{-1}\{ [1-\log(\ell+1)][1-\exp\{-(\ell+1)a_i(w)\}]$$

$$+ \Gamma_1[(\ell+1)a_i(w)]\} ,$$

where $\Gamma_h(x) = \int_0^x (\log u)^h e^{-u}\,du$. Similarly, from (6.6) and (6.7),

$$Q_{ij,p+1}^{(w)} = Q_{i,p+1,j}^{(w)} = x_{ij}\,[1-\exp\{-a_i(w)\}] + x_{ij}\,\Gamma_1[a_i(w)] ,$$

$$j = 1, \ldots, p .$$

Finally, from (6.6) and (6.7) we have

$$Q_{i,p+1,p+1}^{(w)} = \sigma^{-2} \int_0^w [1+\log a_i(y)]^2 \exp\{-a_i(y)\}[d\,a_i(y)/dy]\,dy$$

$$= \sigma^{-2}\{1-\exp\{-a_i(w)\} + 2\,\Gamma_1[a_i(w)] + \Gamma_2[a_i(w)]\} .$$

Using these formulas for $P_{ij\ell}$ and Q_{ijk} in (6.3) and (6.4) gives the components of the expected information, computed under the Kaplan-Meier censoring model, for the Neyman smooth test for the Weibull distribution.

The functions Γ_1 and Γ_2 are derivatives of the incomplete gamma function $\Gamma(w,\alpha) = \int_0^w t^{\alpha-1} e^{-t}\,dt$. In particular, $\Gamma_1(w) = \partial\Gamma(w,1)/\partial\alpha$, and $\Gamma_2(w) = \partial^2\,\Gamma(w,1)/\partial\,\alpha^2$. An algorithm for evaluating derivatives of the incomplete gamma function has been given by

Lindstrom (1981). The values for $\Gamma_1(\infty)$ and $\Gamma_2(\infty)$ can be found from tables, by noting that $\Gamma_1(\infty) = \Psi(1)$, where $\Psi(\alpha)$ is the di-gamma function, and $\Gamma_2(\infty) = \Psi'(1) + [\Psi(1)]^2$, where $\Psi'(\alpha)$ is the tri-gamma function.

For the conditional model, $C_i(y) = I(v_i \leq y)$ where $v_1, \ldots, v_n$ are the potential censoring times. Then from (4.2)-(4.4),

$$n\, I^{(n)}_{\theta_\ell \theta_k} = k\ell[(k+1)(\ell+1)(k+\ell+1)]^{-1} \sum_{i=1}^{n} \{1 - \bar{F}_i^{\,k+\ell+1}(v_i;\gamma)\} ,$$

$$n\, I^{(n)}_{\theta_\ell \gamma_j} = \ell(\ell+1)^{-1} \sum_{i=1}^{n} P_{ij\ell}(v_i) ,$$

and

$$n\, I^{(n)}_{\gamma_j \gamma_k} = \sum_{i=1}^{n} Q_{ijk}(v_i) ,$$

where $P_{ij\ell}$ and Q_{ijk} are as defined in (6.5) and (6.6). The formulas for $P_{ij\ell}$ and Q_{ijk} given above in the exponential and Weibull cases are still valid here.

The totally conditional model is the same as the conditional model, except the potential censoring times for uncensored observations are taken to be $+\infty$.

For the homogeneous exponential censoring model with $\bar{C}(v) = \exp\{-\lambda v\}$, the calculations are reasonably simple only for the exponential distribution. For this case, from (4.2)-(4.4) and the expressions $s_{i\ell}(y;\gamma) = \ell \exp\{-\ell\, y \cdot \exp(x_i'\beta)\}/(\ell+1)$ and $q_{ij}(y;\gamma) = x_{ij}$, we have

$$n\ I^{(n)}_{\theta_\ell \theta_k} = k\ell[(k+1)(\ell+1)]^{-1} \sum_{i=1}^{n} \exp(x_i'\beta)/[(k+\ell+1)\exp(x_i'\beta) + \lambda\ ]\ ,$$

$$n\ I^{(n)}_{\theta_\ell \beta_j} = \ell(\ell+1)^{-1} \sum_{i=1}^{n} x_{ij}\ \exp(x_i'\beta)/[(\ell+1)\ \exp(x_i'\beta) + \lambda\ ]\ ,$$

and

$$n\ I^{(n)}_{\beta_j \beta_k} = \sum_{i=1}^{n} x_{ij}x_{ik} \exp(x_i'\beta)\ /\ [\exp(x_i'\beta) + \lambda\ ]\ .$$

VII. Examples

In this section we apply the Neyman smooth test to two data sets taken from the literature. We will be testing goodness-of-fit in the exponential model where $\bar{F}_i(t;\beta) = \exp\{-t \cdot \exp(x_i'\beta)\}$, and the Weibull model where $\bar{F}_i(t;\gamma) = \exp\{-[t \cdot \exp(x_i'\beta)]^{\sigma}\}$. The scores for the Neyman smooth test are given by (6.1). In this section we only use the expected information computed under the Kaplan-Meier censoring model. Details of the expected information calculations needed here are given in Section VI.4.

In addition to the tests, we also use graphical methods to examine lack of fit. For this purpose, in the exponential model, it is convenient to take for the generalized residuals the random variables $E_i = T_i \exp(x_i'\beta)$, which are unit exponential with distribution function $F_E(w) = 1 - e^{-w}$ when the null hypothesis is true. Then

$$- \log [1 - F_E(w)] = w$$

and

$$\log\{-\log [1 - F_E(w)]\} = \log(w) .$$

Thus if F_E is estimated with the MEDF, $\hat{F}_n^*$, computed from the observed residuals $\hat{e}_i = y_i \exp(x_i'\hat{\beta})$, then a plot of $-\log[1-\hat{F}_n^*(w)]$ against w or $\log\{-\log[1-\hat{F}_n^*(w)]\}$ against $\log(w)$ should lie roughly on a straight line through the origin with slope 1, when the null hypothesis is true. Because of the censoring, there will in general be little information about lack of fit in the right tail, so we prefer the second of these plots, since it spreads out the left side of the distribution. Note that for a Weibull departure from the exponential model, $P[E_i > w] = \exp\{-w^{\sigma}\}$, which gives $\log(-\log P[E_i > w]) = \sigma \log(w)$, so the plot of $\log\{-\log[1-\hat{F}_n^*(w)]\}$ against $\log(w)$ should lie roughly on a straight line through the origin with slope σ.

For making plots with the Weibull model, we take the residuals to be $W_i = [T_i \exp(x_i' \beta)]^{\sigma}$. If we estimate the distribution of the W_i with the MEDF $\hat{F}_n^*$ computed from the $\hat{w}_i = [y_i \exp(x_i' \hat{\beta})]^{\hat{\sigma}}$, then again a plot of $\log\{-\log[1-\hat{F}_n^*(w)]\}$ against $\log(w)$ should lie roughly on a straight line through the origin with slope 1 when the null hypothesis is true.

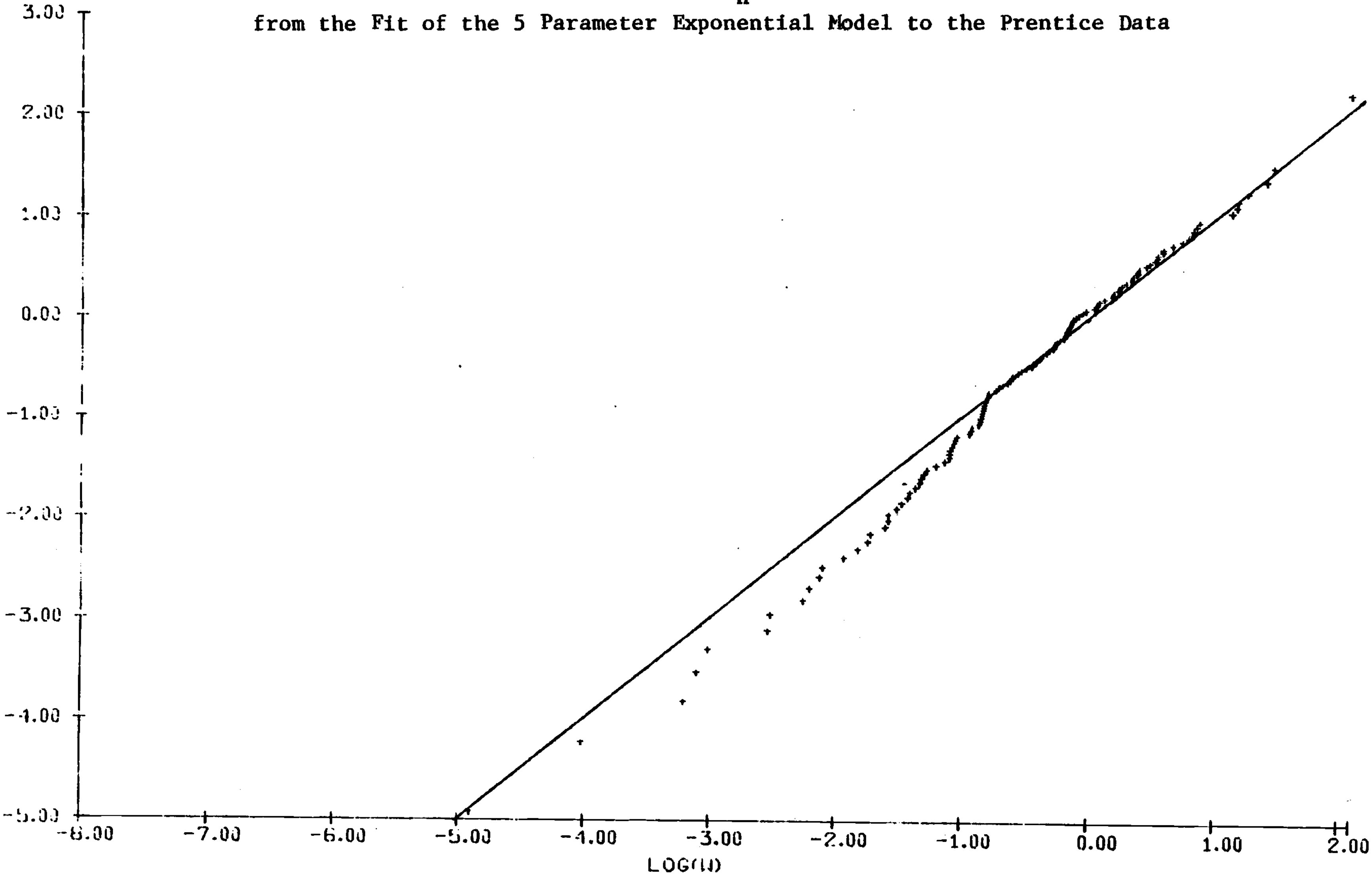
Aitkin and Clayton (1980) also discuss making plots in this setting. In addition to the plots discussed above, they suggest a "variance stabilized" plot of $\sin^{-1}\sqrt{1-\hat{F}_n^*(w)}$ against $\sin^{-1}[\exp(-w/2)]$.

The first data set, taken from Prentice (1973), consists of survival times of 137 lung cancer patients of which 9 are censored. The data have also been discussed by Farewell and Prentice (1977), Kalbfleisch and Prentice (1980), and Aitkin and Clayton (1980). In the exponential analysis of Prentice and the Weibull and proportional hazards analyses of Kalbfleisch and Prentice, only performance status and tumor type (4 categories) were identified as important regression variables. Kalbfleisch and Prentice find no evidence against the exponential model relative to the Weibull model.

We fit an exponential model with the five parameters (constant term, performance status, and three for tumor type). Figure 1 is a plot of $\log\{-\log[1-\hat{F}_n^*(w)]\}$ against $\log(w)$ at those observed residuals $\hat{e}_i$ which correspond to uncensored observations. The points all lie quite close to the line through the origin with slope 1, although there appears to be a slight wave in the plot. The values of the Neyman smooth tests are $W_1 = .19$ and $W_2 = 7.94$. The W_2 test has a p-value $< .025$, so there is substantial evidence here for lack of fit. We also fit the Weibull model, with the same covariables. The plot from the exponential fit does not suggest that the Weibull shape parameter is different from 1, and the likelihood ratio test for the exponential model versus the Weibull model has a value of .91 with an approximate χ_1^2 distribution, so there is very little evidence against the exponential model relative to the Weibull model. The Neyman smooth tests for the Weibull model are 2.08 for the 1 degree of freedom test and 7.25 for the 2 degree of freedom test, so there is also significant evidence against the Weibull model.

This agrees with the results of Farewell and Prentice (1977). They fit a generalized gamma model and found significant evidence against both the Weibull and lognormal models. They also found evidence that the distribution is different for patients who had received prior therapy than for those with no prior therapy. Fitting the exponential model (with the same covariables) to just those patients that have received prior therapy gives $W_1 = .476$ and $W_2 = .480$. For just those patients with no prior therapy, the

Figure 1: Plot of $\log(-\log[1-\hat{F}_n^*(w)])$ Against $\log(w)$
from the Fit of the 5 Parameter Exponential Model to the Prentice Data



results are $W_1 = .76$ and $W_2 = 7.88$. Thus the lack of fit does seem to be concentrated in the no prior therapy group.

To investigate the effect of adding nonsignificant covariables to the model (using the full data set) we fit the exponential model with additional covariables representing months from diagnosis, age in years, receipt of prior therapy, and treatment (standard or test), giving 9 parameters total. The likelihood ratio test for the 9 parameter model versus the 5 parameter model is 1.6 on 4 degrees of freedom, so the extra covariables are not at all significant. The values for the Neyman smooth test for this model are $W_1 = .33$ and $W_2 = 7.40$. To investigate the effect of dropping significant factors, we fit the 2 parameter exponential model with only performance status. The likelihood ratio test for the 2 parameter model versus the 5 parameter model is 18.3 on 3 degrees of freedom, so the omitted tumor type variables are highly significant. The Neyman smooth tests for this model are $W_1 = 2.15$ and $W_2 = 14.48$, so omitting a significant factor has had a substantial effect on the test.

With only 9 of the 137 observations censored, the censoring has had little effect on the test. To see what would happen with more substantial censoring, we generated independent observations from an exponential distribution with scale parameter 120, to use as censoring times for the 128 uncensored observations. This gave a total of 69 censored observations. For the altered data, some of the additional factors were of marginal significance, so we used the 9 parameter model for testing goodness-of-fit. The values of the Neyman smooth test were $W_1 = .68$ and $W_2 = 11.87$. As can be seen in

Figure 2: Plot of $\log(-\log[1-\hat{F}_n^*(w)])$ Against $\log(w)$ from the Fit of the 9 Parameter Exponential Model to the Prentice Data with Additional Censoring

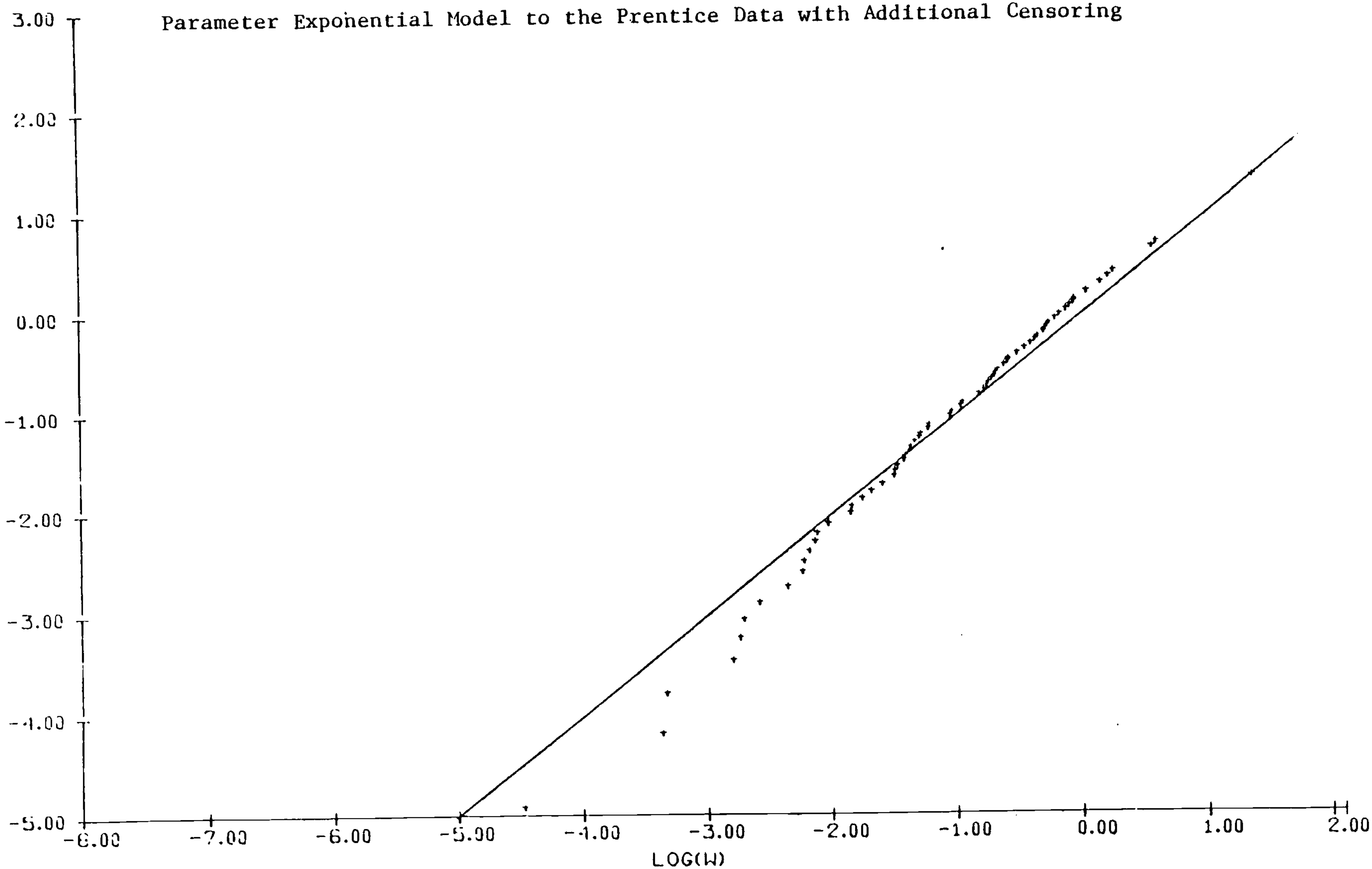


Figure 2, the model does not appear to fit any better with this additional censoring. The increase in the value of W_2 with the extra censoring is surprising, since the censoring mechanism employed was both independent and noninformative.

The second example is taken from Glasser (1967). The data are days survived after surgery for 131 lung cancer patients, with 66 of the observations censored. Two of the times are 0. These were replaced by the value .1 to facilitate the analysis. The patients are in two groups: group one patients have "low" vital capacity/ predicted vital capacity ratios and group two patients have "high" ratios. The age of the patients is also given. In the fit of the exponential model, both age and group are significant. The plot of $\log\{-\log[1-\hat{F}_n^*(w)]\}$ against $\log(w)$ for this model, given in Figure 3, seems to indicate substantial lack of fit. The values of the Neyman smooth test for this model were $W_1 = 4.91$ and $W_2 = 5.02$. The p-value for W_1 is approximately .03, so the Neyman smooth test does give significant evidence of lack of fit.

Examination of Figure 3 indicates that a Weibull model may fit better. This is confirmed in Figure 4 where the plot of $\log\{-\log[1-\hat{F}_n^*(w)]\}$ versus $\log(w)$ from the fit of the Weibull model with the same covariables is given. The values of the Neyman smooth test for the Weibull model were .78 for the 1 degree of freedom test and 3.19 for the 2 degree of freedom test.

The disturbing fact about this example is that the likelihood ratio test for the exponential model versus the Weibull model is 12.9 on 1 degree of freedom. In the uncensored, identically

Figure 3: Plot of $\log(-\log[1-\hat{F}_n^*(w)])$ Against $\log(w)$ from the Fit of the Exponential Model to the Glasser Data

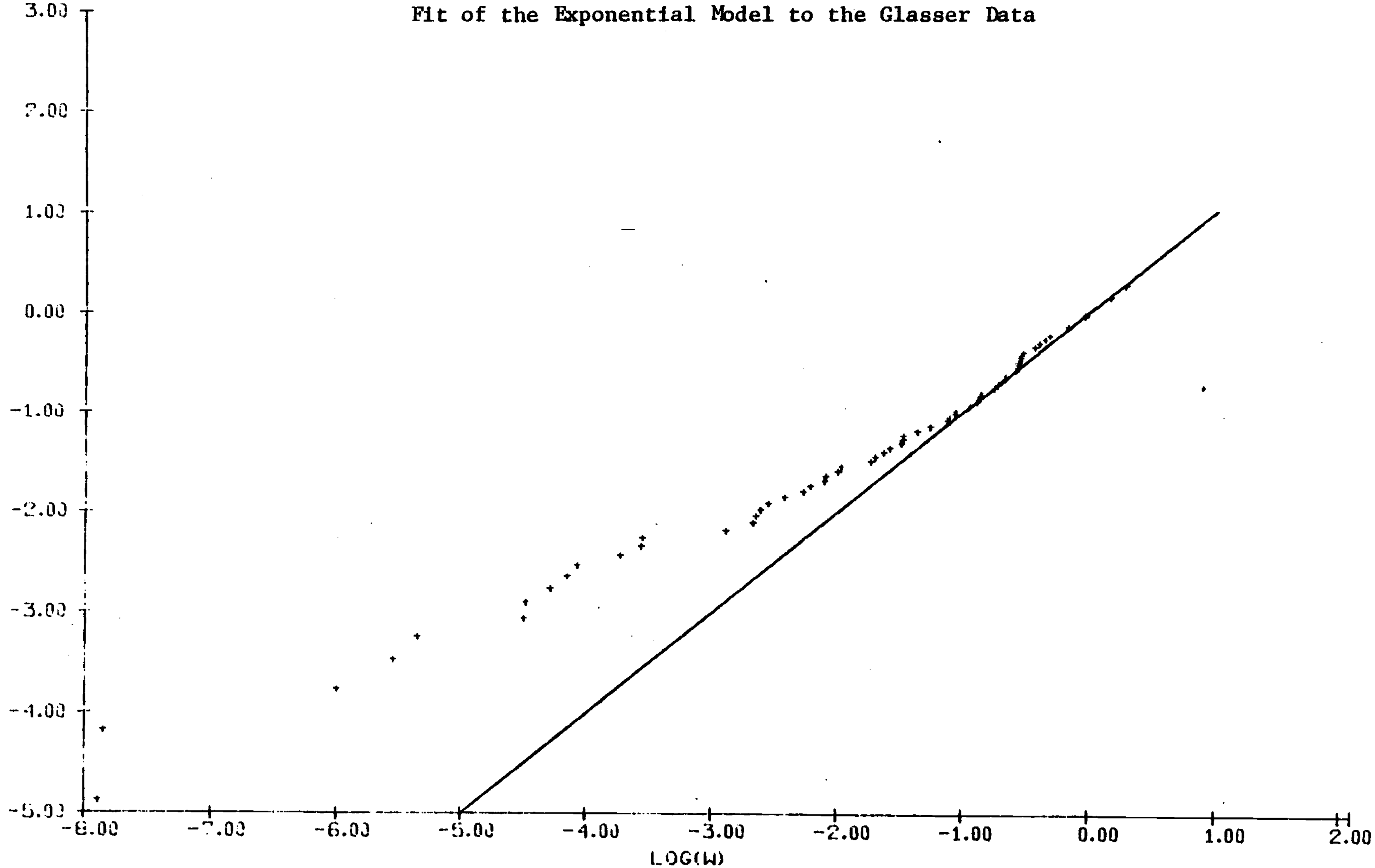
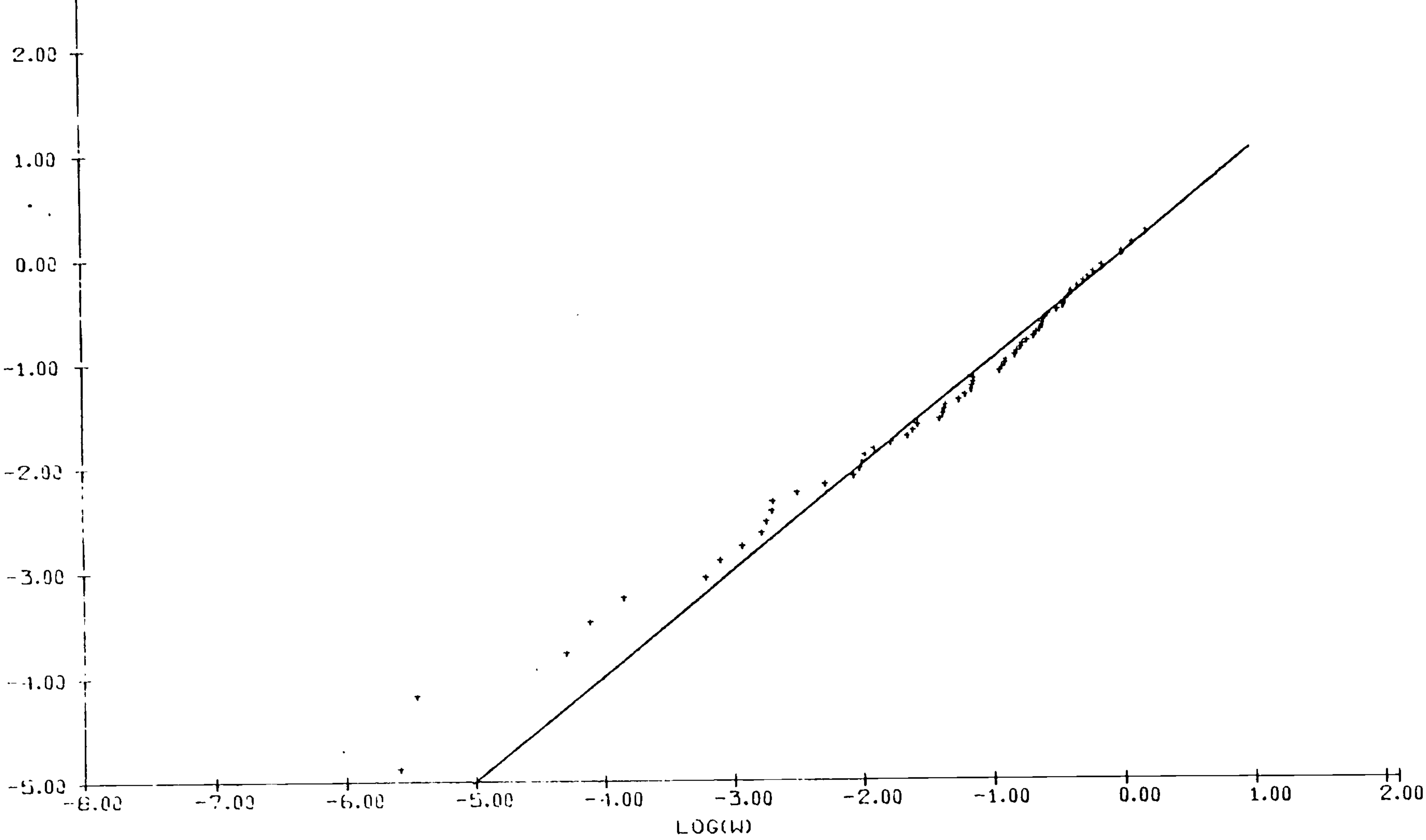


Figure 4: Plot of $\log(-\log[1-\hat{F}_n^*(w)])$ Against $\log(w)$ from the Fit of the Weibull Model to the Glasser Data.



distributed case Kopecky and Pierce (1979) give the local asymptotic relative efficiency of W_1 relative to the efficient score test for the Weibull departure from the exponential as .876, so it is surprising that here there should be such a large discrepancy between the Neyman smooth test and the likelihood ratio test. Also, based on Figure 2 and Figure 3, the exponential model does not appear to fit better here than in the version of the previous example with the extra censoring, yet W_2 there is more significant than either W_1 or W_2 here, and in both examples the proportion of censored observations is roughly .5 . It could be that the effect of the censoring on the power of the Neyman smooth test depends in a complex way on the type of departure and the type of censoring.

VIII. Summary and Conclusions

In this paper our concern has been with testing the underlying distribution in parametric regression models. We have been particularly interested in the complications arising from the presence of censoring in the data. In Section I we presented an alternative to the Kaplan-Meier estimator for estimating the distribution from censored data. This estimate, $\hat{H}_n^*$, defined in equation (1.4), was shown to be a natural modification of the ordinary empirical distribution function to allow for the presence of censored data. We proposed that test statistics be based on functionals of the form $\sqrt{n} \int_0^1 \psi_\ell(u)\, d\, \hat{y}_n^*(u)$. These functionals were shown to have a natural interpretation as the difference between the expectation of ψ_ℓ under the distribution $\hat{H}_n^*$ and the expectation of ψ_ℓ under the null hypothesis distribution. It was shown in Section III that these functionals are also efficient scores from the parametric alternatives (3.1). This connection with efficient scores was seen not only to give further motivation for the proposed statistics, but also to provide the asymptotic distributions of the statistics.

In Section V this connection with the scores was exploited to give a heuristic derivation of the asymptotic distribution of the stochastic process $\hat{y}_n^*(u)$. Using this result it was shown that even for location-scale regression models, the asymptotic distribution depends on the covariables, the true values of the parameters, and the censoring distributions. This makes most other possibilities for goodness-of-fit tests unattractive, since they involve statistics whose asymptotic distributions are difficult to find, and these distributions are essentially different for every problem.

The main difficulty with the class of tests proposed here is how to estimate the asymptotic variance for the quadratic score statistic Q_m. A general discussion of the possibilities was given in Section IV. In Section VI we gave detailed consideration to these possibilities for the Neyman smooth test for exponentiality. The simulation results there suggest that the only one of the methods considered that can always be computed and is reliable under homogeneous random censorship is to use the expected information computed under the Kaplan-Meier censoring model. How this performs when the censoring is not homogeneous is not known. In the special case where potential censoring times are known for all observations, it may be better to use the expected information computed under the conditional model, since this approach does not assume the censoring is homogeneous, and the results in Table 4 suggest the two methods are equally effective when the censoring is homogeneous.

In Section VII we have applied the Neyman smooth test to two data sets taken from the literature. The results indicate that the Neyman smooth test using the Kaplan-Meier censoring model can be a useful diagnostic tool. However, the second example raises some questions about the power of the test with heavy censoring. Further work investigating how different types of censoring effect the power of the test against various departures seems to be indicated.

Bibliography

Aitkin, M. and Clayton, D. (1980). The Fitting of Exponential, Weibull and Extreme Value Distributions to Complex Censored Survival Data Using GLIM. Applied Statistics 29:156-163.

Bargal, A. I. (1981). Efficiency Comparisons of Goodness-of-Fit Tests for Weibull and Gamma Distributions with Singly Censored Data. Ph.D. thesis, Oregon State University.

Barton, D. E. (1956). Neyman's ψ_k^2 Test of Goodness of Fit When the Null Hypothesis is Composite. Skandinaviske Aktuarietidskrift 39:216-246.

Billingsley, P. (1968). Convergence of Probability Measures. New York: John Wiley and Sons.

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.

Burke, M.D., Csörgö, S., and Horváth, L. (1981). Strong Approximations of Some Biometric Estimates Under Random Censorship. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 56:87-112.

Chen, Chen-Hsin. (1981) Correlation-Type Goodness-of-Fit Tests for Randomly Censored Data. Technical Report No. 73, Division of Biostatistics, Stanford University.

Cox, D. R., and Hinkley, D. V. (1974). Theoretical Statistics. London: Chapman and Hall.

Cox, D. R., and Snell, E. J. (1968). A General Definition of Residuals. Journal of the Royal Statistical Society, Series B, 30:248-275.

Cox, D. R., and Snell, E. J. (1971). On Test Statistics Calculated from Residuals. Biometrika 58:589-594.

Crowley, J. and Hu, M. (1977). Covariance Analysis of Heart Transplant Survival Data. Journal of the American Statistical Association 72:27-36.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society, Series B, 39:1-38.

Downham, D. Y. and Roberts, F. D. K. (1967). Multiplicative Congruential Pseudo-Random Number Generators. Computer Journal 10:74-77.

Durbin, J. (1973). Weak Convergence of the Sample Distribution Function When Parameters are Estimated. Annals of Statistics 1:279-290.

Durbin, J. and Knott, M. (1972). Components of Cramér-von Mises Statistics I. Journal of the Royal Statistical Society, Series B,34:290-307.

Efron, B. (1967). The Two Sample Problem with Censored Data. Proceedings of the Fifth Berkeley Symposium IV:831-853.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. Annals of Statistics 7:1-26.

Efron, B. (1981). Censored Data and the Bootstrap. Journal of the American Statistical Association 76:312-319.

Farewell, V. T. and Prentice, R. L. (1977). A Study of Distributional Shape in Life Testing. Technometrics 19:69-76.

Fleming, T. R. and Harrington, D. P. (1981). A Class of Hypothesis Tests for One and Two Sample Censored Survival Data. Communications in Statistics - Theory and Methods A10:763-794.

Fleming, T. R., O'Fallon, J. R., O'Brien, P.C. and Harrington, D.P. (1980). Modified Kolmogorov-Smirnov Test Procedures with Application to Arbitrarily Right-Censored Data. Biometrics 36:607-625.

Gillespie, M. J. and Fisher, L. (1974). Confidence Bands for the Kaplan-Meier Survival Curve Estimate. Annals of Statistics 7:920-924.

Glasser, M. (1967). Exponential Survival with Covariance. Journal of the American Statistical Association 62:561-568.

Habib, G. (1981). A Chi-Square Goodness-of-Fit Test for Censored Data. Ph.D. thesis, Oregon State University.

Hall, W. J. and Wellner, J. A. (1980). Confidence Bands for a Survival Curve from Censored Data. Biometrika 67:133-143.

Hinkley, D. V. (1977). Jackknifing in Unbalanced Situations. Technometrics 19:285-292.

Hollander, M. and Proschan, F. (1979). Testing to Determine the Underlying Distribution Using Randomly Censored Data. Biometrics 35:393-401.

Hyde, J. (1977). Testing Survival Under Right Censoring and Left Truncation. Biometrika 64:225-230.

Kalbfleisch, J. D. and Prentice, R. L. (1980). The Statistical Analysis of Failure Time Data. New York: John Wiley and Sons.

Kaplan, E. L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association 53:457-481.

Kopecky, K. J. and Pierce, D. A. (1979). Efficiency of Smooth Goodness-of-Fit Tests. Journal of the American Statistical Association 74:393-397.

Koziol, J. A. (1980). Goodness-of-Fit Tests for Randomly Censored Data. Biometrika 67:693-696.

Koziol, J. A. and Green, S. B. (1976). A Cramér-von Mises Statistic for Randomly Censored Data. Biometrika 63:465-474.

Lindstrom, F. T. (1981). An Algorithm for the Evaluation of the Incomplete Gamma Function and the First Two Partial Derivatives with Respect to the Parameter. Communications in Statistics - Simulation and Computation B10:465-478.

Loynes, R. M. (1980). The Empirical Distribution Function of Residuals from Generalized Regression. Annals of Statistics 8:285-298.

Mihalko, D. P. and Moore, D. S. (1980). Chi-Square Tests of Fit for Type II Censored Data. Annals of Statistics 8:625-644.

Miller, R. G. (1974). The Jackknife - a Review. Biometrika 61:1-15.

Nair, V. N. (1981). Plots and Tests for Goodness of Fit with Randomly Censored Data. Biometrika 68:99-103.

Neyman, J. (1937). "Smooth Test" for Goodness of Fit. Skandinaviske Aktuarietidskrift 20:150-199.

Pettit, A. N. (1976). Cramér-von Mises Statistics for Testing Normality with Censored Samples. Biometrika 63:475-481.

Pettit, A. N. (1977). Tests for the Exponential Distribution with Censored Data Using Cramér-von Mises Statistics. Biometrika 64:629-632.

Pierce, D. A. and Kopecky, K. J. (1979). Testing Goodness of Fit for the Distribution of Errors in Regression Models. Biometrika 66:1-5.

Prentice, R. L. (1973). Exponential Survivals with Censoring and Explanatory Variables. Biometrika 60:279-288.

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.

Smith, R. M. and Bain, L. J. (1976). Correlation Type Goodness-of-Fit Statistics with Censored Sampling. Communications in Statistics - Theory and Methods A5:119-132.

Thomas, D. R. and Pierce, D. A. (1979). Neyman's Smooth Goodness-of-Fit Test When the Hypothesis is Composite. Journal of the American Statistical Association 74:441-445.

Turnbull, B. W. and Weiss, L. (1978). A Likelihood Ratio Statistic for Testing Goodness of Fit with Randomly Censored Data. Biometrics 34:367-375.

APPENDIX

Appendix

In this appendix we establish the claim made in Section I that functionals of $\hat{y}_n$ have "no local power" against omitting covariables from the model in the location-scale case. The following notation is used for the parameters: $\nu = (\alpha,\gamma)$, $\nu_n = (\alpha_n,\gamma_o)$ and $\nu_o = (0,\gamma_o)$, where $\gamma = (\beta,\sigma)$. For $1 \le i \le n$, we assume T_{in} are independent with continuous distribution functions $F_{in}(t;\nu) = G([t - x_i'\beta - w_i'\alpha] / \sigma)$ and densities $f_{in}(t;\nu)$, where x_i and w_i are vectors of known constants. We assume the X portion of the regression model contains a constant term. The null hypothesis is $\alpha = 0$. We are interested in the asymptotic distribution of the process $\hat{y}_n(u)$, defined in Section I, under the sequence of alternatives $H_n(\lambda) : \alpha = \alpha_n = \lambda/\sqrt{n}$. The true value of the nuisance parameter γ is denoted by γ_o. We use the result of Loynes (1980). If his conditions A1, A2, A4, A5, A7 and A9 are satisfied, then under the sequence of alternatives, $\hat{y}_n(u)$ converges weakly to a Gaussian process $y(u)$ with

$$E[y(u)] = \{[\psi_\gamma(u)]'I_{\gamma\gamma}^{-1} I_{\gamma\alpha} - [\psi_\alpha(u)]'\}\lambda \ , \quad 0 \le u \le 1 \ ,$$

and

$$\mathrm{Cov}[y(u_1) \ , \ y(u_2)] = \min(u_1,u_2) - u_1u_2 - [\psi_\gamma(u_1)]' \, I_{\gamma\gamma}^{-1} \, \psi_\gamma(u_2) \ , \qquad 0 \le u_1 \ , \ u_2 \le 1 \ ,$$

where

$$I_{\gamma\gamma} = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} E\left\{[\partial \log f_{in}(T_{in};\nu_n)/\partial\gamma][\partial \log f_{in}(T_{in};\nu_n)/\partial\gamma]' \Big|_{\nu=\nu_n}\right\},$$

$$I_{\gamma\alpha} = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} E\left\{[\partial \log f_{in}(T_{in};\nu)/\partial\gamma][\partial \log f_{in}(T_{in};\nu)/\partial\alpha]' \Big|_{\nu=\nu_o}\right\},$$

$$\psi_{\gamma}(u) = -n^{-1} \sum_{i=1}^{n} [\partial F_{in}(F_{in}^{-1}(u;\nu_n);\nu)/\partial\gamma] \Big|_{\nu=\nu_n},$$

and

$$\psi_{\alpha}(u) = -n^{-1} \sum_{i=1}^{n} [\partial F_{in}(F_{in}^{-1}(u;\nu_n);\nu)/\partial\alpha] \Big|_{\nu=\nu_n}.$$

We show the distribution of $y(u)$ does not depend on λ, so that the distribution is the same under the sequence of alternatives as under the null hypothesis. First we show $[\psi_{\gamma}(u)]' I_{\gamma\gamma}^{-1} I_{\gamma\alpha} = [\psi_{\alpha}(u)]'$ so that $E[y(u)] \equiv 0$.

As discussed by Pierce and Kopecky (1979), the location-scale regression model can always be reparametrized so that in the information matrix $I_{\nu\nu}$, the components $I_{\beta\sigma}$ and $I_{\alpha\sigma}$ are zero. For such a reparametrization (still using the same symbols for the new parameters), let

$$i_{\mu\mu} = E\ [\ g'(E_i) / g(E_i)\ ]$$

and

$$i_{\sigma\sigma} = E\ [\ g'(E_i) / g(E_i)\ ],$$

where $E_i \sim G$ and $g(t) = dG(t)/dt$. Note that $i_{\mu\mu}$ and $i_{\sigma\sigma}$ are intrinsic constants of the distribution G and do not depend on ν. Then $I^{(n)}_{\beta_j\beta_k} = \sigma^{-2}\, n^{-1} \sum_{i=1}^{n} x_{ij}x_{ik} i_{\mu\mu}$ and $I^{(n)}_{\sigma\sigma} = \sigma^{-2}\, n^{-1} \sum_{i=1}^{n} i_{\sigma\sigma}$, so

$$I^{(n)}_{\gamma\gamma} = \sigma^{-2} \begin{pmatrix} n^{-1} X_n' X_n\, i_{\mu\mu} & 0 \\ 0 & i_{\sigma\sigma} \end{pmatrix}, \qquad \text{(A.1)}$$

where $X_n = [x_1, \ldots, x_n]'$. Similarly, with $W_n = [w_1, \ldots, w_n]'$, we have

$$I^{(n)}_{\gamma\alpha} = \sigma^{-2} \begin{bmatrix} n^{-1} X_n' W_n\, i_{\mu\mu} \\ 0 \end{bmatrix} .$$

Now $F^{-1}_{in}(u;\nu_n) = \sigma\, G^{-1}(u) + x_i'\beta + w_i'\alpha_n$, $\partial F_{in}(t;\nu)/\partial\beta_j = -\sigma^{-1} x_{ij}\, g(e_i)$ and $\partial F_{in}(t;\nu)/\partial\sigma = -\sigma^{-1} e_i\, g(e_i)$, where $e_i = [t - x_i'\beta - w_i'\alpha]/\sigma$. Thus

$$\psi^{(n)}_{\gamma}(u) = \sigma^{-1} \begin{pmatrix} n^{-1} X_n' 1_n\, d(u) \\ h(u)\, d(u) \end{pmatrix}, \qquad \text{(A.2)}$$

where $d(u) = g[h(u)]$, $h(u) = G^{-1}(u)$, and 1_n is an n-dimensional vector of ones. Similarly,

$$\psi^{(n)}_{\alpha}(u) = W_n' 1_n\, d(u) / (n\sigma) .$$

Thus

$$[\psi_\gamma(u)]' I_{\gamma\gamma}^{-1} I_{\gamma\alpha} = \lim_{n\to\infty} \{[\psi_\gamma^{(n)}(u)]'[I_{\gamma\gamma}^{(n)}]^{-1} I_{\gamma\alpha}^{(n)}\}$$

$$= \lim_{n\to\infty} [d(u)\ 1_n'\ X_n(X_n' X_n)^{-1} X_n' W_n/(n\sigma)]$$

$$= \lim_{n\to\infty} \quad d(u)\ 1_n' W_n/(n\sigma)$$

$$= [\psi_\alpha(u)]'$$

The next to last equality is due to the fact that $X_n(X_n' X_n)^{-1} X_n' 1_n = 1_n$, since $X_n (X_n' X_n)^{-1} X_n'$ is the orthogonal projection on the span of the columns of X_n, and 1_n is in this span because the regression model is assumed to contain a constant term.

Thus $E[y(u)] \equiv 0$. Clearly from (A.1) and (A.2), the term $[\psi_\gamma(u_1)]' I_{\gamma\gamma}^{-1} \psi_\gamma(u_2)$ in the covariance function of $y(u)$ does not depend on λ either. Thus the claim is established.