Yandell

P. R. Krishnaiah and P. K. Sen, eds., *Handbook of Statistics, Vol. 4*

26

Tests for Exponentiality

Kjell A. Doksum and Brian S. Yandell

1. Introduction

The exponential hypothesis is important because of its implications concerning the random mechanism operating in the experiment being considered. In reliability, the exponential assumption may apply when one is dealing with failure times of items or equipment without any moving parts, such as for instance transistors, fuses, air monitors, car fenders, etc. In these examples, failure is not brought about by wear, but by a random shock, and the exponential assumption corresponds to assuming that this shock follows a Poisson process distribution. Thus testing the exponential assumption about the failure time distribution is equivalent to testing the Poisson assumption about the process producing the shock that causes failure.

Tests for exponentiality are subject to the usual dilemma concerning goodness of fit tests, namely, only when the hypothesis is rejected do we have a significant result. Thus, if the significance level of a test of exponentiality is $\alpha = 0.05$ and the true underlying model is Weibull and not exponential, the probability of falsely accepting H_0 can be nearly $1 - \alpha = 0.95$.

On the other hand, when a test rejects exponentiality it justifies the use of other more complicated models and the probabilistic and statistical methods that go along with these models. Such models and methods can be found in the books by Barlow and Proschan (1975) and Kalbfleisch and Prentice (1980).

In this paper we present some of the tests available for testing exponentiality. It is not a complete treatment of the topic and reflects the author's interests and biases.

In Section 2, we introduce some of the common parametric and nonparametric alternatives to exponentiality. The next section discusses tests designed for parametric models and it is shown that one of these tests is appropriate in a nonparametric setting. Spacings tests are discussed in Section 4, and their isotonic properties for increasing failure rate alternatives are

This research was supported in part by National Science Foundation Grant MCS-8102349. Brian Yandell is now at the departments of Statistics and Horticulture, University of Wisconsin, Madison.

developed. In Section 5, tests based on the total time on test transform are considered, while in Section 6 the nonparametric optimality of the total time on test statistic is developed.

Some of the common distance type statistics are discussed in Section 7 and graphical methods based on *Q–Q* plots and the total time on test transforms are given in Section 8. Section 9 gives some tests designed to detect 'New Better than Used' alternatives.

The rest of the paper concerns testing for exponentiality in the presence of right censoring. Censoring arises in many practical problems when individuals under study cannot be observed until failure. Section 10 presents several common types of censoring and details the notation used for this part. Estimates and their properties are reviewed in Section 11. Very few small-sample results exist for censored data problems (Chen, Hollander and Langberg, 1980). We therefore briefly review in Section 12 the weak convergence of the survival curve and of the cumulative failure rate, or hazard function, to a Gaussian process. This is used in later sections to examine asymptotic properties of several generalizations of tests without censoring. These include maximal deviation (Kolmogorov–Smirnov) tests which can be inverted to yield simultaneous confidence bands (Section 13); tests based on spacings and the total time on test (14); and others including average deviation, or Cramér–von Mises tests, and linear rank tests (15); Monte Carlo simulation results are summarized in Section 16. The question of exponentiality is explored using several tests on data from a prostate cancer study in Section 17.

2. Some alternatives to exponentiality

2a. Parametric models

It is often useful to study properties of nonparametric methods at certain important parametric models. Two such alternatives to the exponential model are the gamma and Weibull models whose probability densities are respectively

$$\begin{aligned} f_G(t;\theta,\lambda) &= \lambda(\lambda t)^{\theta-1}\,e^{-\lambda t}/\Gamma(\theta), \quad t>0\,, \\ f_W(t;\theta,\lambda) &= \lambda\theta(\lambda t)^{\theta-1}\,e^{-(\lambda t)^\theta}, \quad t>0\,, \end{aligned} \tag{2.1}$$

where θ, $\lambda > 0$. Their properties are discussed in Barlow and Proschan (1975) and Kalbfleisch and Prentice (1980). When $\theta = 1$ both reduce to the exponential density so the exponential hypothesis can be written H_θ: $\theta = 1$.

Properties of alternatives to exponentiality are most conveniently expressed through the *failure* (or *hazard*) *rate* function defined by

$$\lambda(t) = f(t)/[1 - F(t)], \quad t > 0\,,$$

where F is the failure distribution defined by $F(t) = P(T \leq t)$ and $f(t)$ is its density. The failure times of equipment or components with moving parts are modelled to have an increasing failure rate distribution since wear would increase the rate of failure.

In the case of gamma and Weibull distributions, we find that the failure rate is monotone increasing if $\theta > 1$ and monotone decreasing if $\theta < 1$. In fact, for the Weibull density, we have

$$\lambda_{\mathrm{W}}(t) = \theta\lambda^{\theta}t^{\theta-1}, \quad t > 0 .$$

Two other interesting, but less familiar parametric models are

$$\begin{aligned} f_{\mathrm{L}}(t;\, \theta, \lambda) &= \lambda(1 + \theta\lambda t)\exp\{-(\lambda t + \tfrac{1}{2}\theta\lambda^2 t^2)\}, \\ f_{\mathrm{M}}(t;\, \theta, \lambda) &= \lambda[1 + \theta K(\lambda t))]\exp\{-[\lambda t + \theta(\lambda t - K(\lambda t))]\} \end{aligned} \tag{2.2}$$

where $\theta \geq 0$, $t > 0$, and $K(x) = 1 - \exp(-x)$, $x > 0$. We refer to these as the linear failure rate density and Makeham (type) density, respectively. They were introduced by Bickel and Doksum (1969). Their failure rates are

$$\lambda_{\mathrm{L}}(t) = \lambda(1 + \theta\lambda t), \qquad \lambda_{\mathrm{M}}(t) = \lambda[1 + \theta(1 - \mathrm{e}^{-\lambda t})] .$$

These densities reduce to the exponential density when $\theta = 0$, the failure rates are increasing when $\theta > 0$.

2b. Nonparametric models

It is usually hard to determine exactly which parametric family of densities is appropriate in a given experiment. Thus it is useful to turn to nonparametric classes of distributions that arise naturally from physical considerations of aging and wear. Three such natural classes of nonparametric models are listed below.

(1) The class of all IFR (Increasing Failure Rate) distributions. This is the class of distribution functions F that have failure rate $\lambda(t)$ nondecreasing for $t > 0$.

(2) The class of IFRA (IFR Average) distributions which is the class of F where the failure rate average

$$a(t) = t^{-1}\int_0^t \lambda(x)\,\mathrm{d}x = -t^{-1}\log[1 - F(t)] \tag{2.3}$$

is nondecreasing. This class has nice closure properties: It is the smallest class of F's which includes the exponential distribution and is closed under the formation of coherent systems (Birnbaum, Esary and Marshall, 1966) and it is closed under convolution (Block and Savits, 1976).

(3) The class of NBU (New Better than Used) distributions F is the class with

$$S(s+t) \leq S(s)S(t), \quad s \geq 0,\ t \geq 0\,, \tag{2.4}$$

where $S(t) = 1 - F(t)$ is the survival function. Note that (2.4) is equivalent to stating that the conditional survival probability $S(s+t)/S(s)$ of a unit of age s is less than the corresponding survival probability of a new unit.

The three above classes satisfy

$$\text{IFR} \subset \text{IFRA} \subset \text{NBU}\,.$$

Thus the gamma and Weibull distributions with $\theta > 1$ are examples of F's for all three classes as are the F_L and F_M distributions when $\theta > 0$.

For further results on these nonparametric classes, see Barlow and Proschan (1975), and Hollander and Proschan (1984, this volume).

3. Parametric tests

In this section we consider tests that are asymptotically (approximately, for large sample size) optimal for parametric alternatives in the sense that in the class of all level α tests (assuming scale λ unknown) they maximize the asymptotic power. We will find that one of these tests is consistent for the nonparametric class of all IFRA alternatives.

Let $T_1, \ldots, T_n$ denote n survival or failure times assumed to be independent and to follow a continuous distribution F satisfying $F(0) = 0$. The exponential hypothesis H_0 is that $F(t) = K_\lambda(t)$, some λ, where

$$K_\lambda(t) = 1 - e^{-\lambda t}, \quad t > 0,\ \lambda > 0\,.$$

Suppose we have a parametric alternative with density $f(t;\, \theta, \lambda)$ in mind, where θ is a real shape parameter, λ is a real scale parameter, and $\theta = \theta_0$ corresponds to the exponential hypothesis.

With this setup, it is natural to apply the likelihood ratio test which is based on the likelihood ratio statistic

$$R(\boldsymbol{t}) = \frac{\sup_{\theta,\lambda} L(\boldsymbol{t};\, \theta, \lambda)}{\sup_\lambda L(\boldsymbol{t};\, \theta_0, \lambda)}$$

where $\boldsymbol{t} = (t_1, \ldots, t_n)$ is the observed sample vector, $L(\boldsymbol{t};\, \theta, \lambda) = \Pi_{i=1}^n f(t_n;\, \theta, \lambda)$ is the likelihood function, and the sup is over $\lambda > 0$ and $\theta \in \Theta$, where Θ is the parameter set for θ. In the examples of Section 2, $\Theta = [0, \infty]$.

Note that since the maximum likelihood estimate of λ in the exponential model is $\hat{\lambda} = 1/\bar{t}$, then

$$\sup_\lambda L(\boldsymbol{t};\, \theta_0, \lambda) = L\left(\boldsymbol{t};\, \theta_0, \frac{1}{\bar{t}}\right) = \frac{1}{\bar{t}}\, e^{-n}\,.$$

For smooth models, as in Section 2(a), the value of $R(t)$ can be computed on a computer. The test rule based on $R(t)$ is to reject exponentiality when $R(t) \geq k_\alpha$, where k_α is the $(1-\alpha)$-th quartile of a χ^2 distribution with one degree of freedom (e.g., Bickel and Doksum, 1977, p. 229).

Another test suitable for a parametric alternative $f(t; \theta, \lambda)$ is Neyman's (1959) asymptotically most powerful $C(\alpha)$ test. This test is asymptotically most powerful in the class of all similar tests, that is, in the class of all tests that have level α no matter what the value of the unknown parameter λ is.

Let

$$h(t) = \frac{\partial}{\partial \theta} \log f(t; \theta, 1)\bigg|_{\theta=\theta_0^+},$$

then it can be easily shown that in our setup the $C(\alpha)$ test reduces to a test which rejects exponentiality for large values of the test statistic

$$T(h) = (1/\sqrt{n}) \sum_{i=1}^{n} h(t_i/\bar{t})/\tau(h) \tag{3.1}$$

where

$$\bar{t} = \frac{1}{n} \sum_{i=1}^{n} t_i$$

and

$$\tau^2(h) = \int_0^\infty h^2(t)\, e^{-t}\, dt - \left[\int_0^\infty th(t)\, e^{-t}\, dt\right]^2. \tag{3.2}$$

The test rule is to reject H_0 when $T(h) \geq c_\alpha$, where c_α is the upper α critical value from a standard normal distribution, i.e. $c_{0.05} = 1.645$. For the four parametric models f_G, f_W, f_L and f_M of the previous section, we find, after some simplification,

$$T_G = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\log(t_i/\bar{t}) + E]/\sqrt{\tfrac{1}{6}\pi^2 - 1},$$

$$T_W = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{1 + [1 - (t_i/\bar{t})] \log(t_i/\bar{t})\}/\sqrt{\tfrac{1}{6}\pi^2},$$

$$T_L = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [1 - \tfrac{1}{2}(t_i/\bar{t})^2], \qquad T_M = \frac{1}{\sqrt{n}} \sum_{n=1}^{n} [2K(t_n/\bar{t}) - 1]/\sqrt{\tfrac{1}{12}}$$

respectively, where E = Euler's constant = 0.5772 and K is the standard exponential distribution function $1 - e^{-x}$.

Next, we consider the question of whether any of these four test statistics will have desirable properties not only for the parametric alternative they were derived for but also for nonparametric classes of failure distributions. We find that

THEOREM 3.1. *The test that rejects H_0 when $T_L \geq c_\alpha$ is consistent for any alternative F in the class of IFRA distributions.*

PROOF. Rewrite T_L as

$$T_L = \tfrac{1}{2}\sqrt{n}\left[1 - \frac{\hat{\sigma}^2}{\bar{t}^2}\right],$$

where $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(t_i - \bar{t})^2$. Under H_0, T_L converges (in law) to a standard normal random variable. For IFRA alternatives, $\sigma^2 = \mathrm{Var}(T)$ exists, thus $(\hat{\sigma}^2/\bar{t}^2) \to \sigma^2/\mu^2$ a.s. as $n \to \infty$, where $\mu = E(T)$. If F is an IFRA distribution different from $K_\lambda(t)$, then from Barlow and Proschan (1975, p. 118), $(\sigma/\mu) < 1$; thus $T_L \to \infty$ (a.s.). Thus the power of the test converges to one as $n \to \infty$.

Note that this test is equivalent to rejecting H_0 for large values of the sample coefficient of variation $\bar{t}/\hat{\sigma}$ and that it can be carried out on any calculator that computes $\bar{t}$ and $\hat{\sigma}^2$.

EXAMPLE 3.1. In Table 3.1 we give 107 failure times for right rear breaks on D9G-66A Caterpillar tractors. These numbers are reproduced from Barlow and Campo (1975). We find $\bar{t} = 2024.26$ and $\hat{\sigma} = 1404.35$, thus $T_L = 2.68$ and the level $\alpha = 0.01$ test based on T_L rejects the hypothesis. The p-value is $p_L = 0.0037$. By comparison we find $T_M = 4.20$, so the test based on this statistic rejects H_0 with negligible p-value.

Shorack (1972) derived the uniformly most powerful invariant test for gamma alternatives. It is equivalent to the $C(\alpha)$ test based on T_G. Spiegelhalter (1983)

Table 3.1
Failure data for right rear brake on D9G-66A caterpillar tractor

56	806	1253	1927	2325	3185
83	834	1313	1957	2337	3191
104	838	1329	2005	2351	3439
116	862	1347	2010	2437	3617
244	897	1454	2016	2454	3685
305	904	1464	2022	2546	3756
429	981	1490	2037	2565	3826
452	1007	1491	2065	2584	3995
453	1008	1532	2096	2624	4007
503	1049	1549	2139	2675	4159
552	1069	1568	2150	2701	4300
614	1107	1574	2156	2755	4487
661	1125	1586	2160	2877	5074
673	1141	1599	2190	2879	5579
683	1153	1608	2210	2922	5623
685	1154	1723	2220	2986	6869
753	1193	1769	2248	3092	7739
763	1201	1795	2285	3160	

derived the locally most powerful test for Weibull alternatives and obtained the $C(\alpha)$ test based on T_W.

4. Tests based on spacings

Let $T_1, \ldots, T_n$ denote n survival or failure times assumed to be independent and to follow a continuous distribution F satisfying $F(0) = 0$. The exponential hypothesis is that

$$F(x) = 1 - e^{-\lambda x}, \quad x > 0, \ \lambda \geq 0 . \tag{4.1}$$

We look for a simple transformation of $T_1, \ldots, T_n$ that will yield new variables $D_1, \ldots, D_n$ with a distribution which is sensitive to IFR deviations from the exponential assumption. Such a transformation is defined by

$$D_i = (n + 1 - i)(T_{(i)} - (T_{(i-1)}), \quad i = 1, \ldots, n, \tag{4.2}$$

where $T_{(0)} = 0$ and $T_{(1)} < \cdots < T_{(n)}$ are the ordered T's. Using the Jacobian result on transformations of random variables, (e.g., Bickel and Doksum, 1977, p. 46), we find that under the exponential hypothesis, $D_1, \ldots, D_n$ are independent and each has the exponential distribution (4.1).

The D's are called the *normalized sample spacings*, or just *spacings* for short. They are useful since for the important class of IFR alternatives, there will be a stochastic downward trend in the spacings and tests that are good for trend will be good for IFR alternatives. To make this claim precise, we define a distribution F to be *more IFR* than G, written $F <_c G$, if $G^{-1}F$ is convex, where $G^{-1}F$ is defined by $P_F(T^{-1}F(T) \leq t) = G(t)$, $t \geq 0$ (Van Zwet, 1964; Bickel and Doksum, 1969). With this definition, 'F is IFR' is equivalent to '$F <_c K$', where $K(x)$ denotes the standard exponential distribution $1 - e^{-x}$. Moreover, for the gamma and Weibull families $F_{G,\theta}$ and $F_{W,\theta}$ of Section 2; $F_{G,\theta_2} <_c F_{G,\theta_1}$ and $F_{W,\theta_2} <_c F_{W,\theta_1}$ are both equivalent to $\theta_1 < \theta_2$.

We say that there is a *stronger downward trend* in $D_1, \ldots, D_n$ than $D'_1, \ldots, D'_n$ if D'_i/D_i is nondecreasing in i.

Now we can make precise the notion that the more increasing the failure rate, the stronger the stochastic downward trend in the spacings.

LEMMA 4.1. *Suppose F is more IFR than G. Let $T_1, \ldots, T_n$ be a sample from F with corresponding spacings $D_1, \ldots, D_n$. Then there is a sample $T'_1, \ldots, T'_n$ with distribution G and spacings $D'_1, \ldots, D'_n$ such that there is a stronger downward trend in $D_1, \ldots, D_n$ than in $D'_1, \ldots, D'_n$.*

PROOF. Let $T_{(1)} < \cdots < T_{(n)}$ be the ordered failure times and let $T'_{(i)} = G^{-1}F(T_{(i)})$, $i = 1, \ldots, n$. Then $T'_{(1)}, \ldots, T'_{(n)}$ are distributed as order statistics from G. Next let $D'_i = (n - i + 1)(T'_{(i)} - T'_{(i-1)})$. Since the function $G^{-1}F$ is convex, its slope is increasing and thus

$$(T'_{(i)} - T'_{(i-1)})/(T_{(i)} - T_{(i-1)}) \leq (T'_{(j)} - T'_{(j-1)})/(T_{(j)} - T_{(j-1)})$$

for $i < j$. It follows that $(D'_i/D_i) \leq (D'_j/D_j)$, $i < j$.

As an application of this Lemma, we note that there is a stronger downward trend in spacings from an IFR population than in spacings from an exponential population. Since spacings from an exponential population form a sample from an exponential distribution, there is no trend in these spacings.

Figure 4.1 shows a downward trend in the spacings for the tractor data of Example 3.1. The spacings D_i are plotted against $i/(n+1)$.

We consider two types of test statistics appropriate for testing no trend vs. downward trend. The first is the class of *linear rank statistics* of the form

$$\frac{1}{n}\sum_{i=1}^{n} c\Big(\frac{i}{n+1}\Big)J\Big(\frac{R_i}{n+1}\Big)$$

where R_i is the rank of D_i, and $c(1/(n+1)), \ldots, c(n/(n+1))$, $J(1/(n+1)), \ldots, J(n/(n+1))$ are constants to be chosen subject to the condition that $-c(i/(n+1))$ and $J(i/(n+1))$ are nondecreasing in i. Proschan and Pyke (1967) proposed $J(i/(n+1)) = i/(n+1)$, while Bickel and Doksum (1969) showed that it is both better and asymptotically optimal for all alternatives $f(t;\theta,\lambda)$ to choose $J(i/(n+1)) = -\log(1 - i/(n+1))$. Thus we will from now on consider

$$W = \frac{1}{n}\sum_{i=1}^{n} c\Big(\frac{i}{n+1}\Big)\Big[-\log\Big(1-\frac{R_i}{n+1}\Big)\Big].$$

The choice of c depends on the alternative, and for the parametric alternatives f_G, f_W, f_L, f_M of Section 2 the respective asymptotically optimal choices of c are (Bickel and Doksum, 1969)

$$c_G(u) = (1-u)^{-1}\int_{-\log(1-u)}^{\infty} x^{-1}e^{-x}\,dx, \qquad c_W(u) = -\log[-\log(1-u)],$$
$$c_L(u) = \log(1-u), \qquad c_M(u) = -u. \tag{4.3}$$

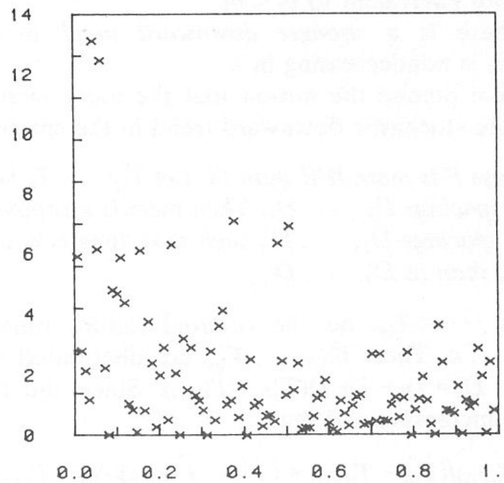

Fig. 4.1. Plot of spacings D_i vs. $i/(n+1)$ for the tractor data of Table 3.1.

The second class of statistics is the class of (*standardized*) *linear spacings statistics* which are of the form

$$S = \sum_{n=1}^{n} c\Big(\frac{i}{n+1}\Big)D_i \Big/ \sum_{i=1}^{n} D_i$$

where $c(i/(n+1))$ again is nonincreasing in i. This class was considered by Barlow and Prochan (1966). For the four parametric alternatives of Section 2, the optimal c to use in S is precisely as in (4.3) above (Bickel and Doksum, 1969; Bickel, 1969). We denote these asymptotically optimal spacings statistics by S_G, S_W, S_L and S_M respectively.

Next we turn to nonparametric properties of these two classes of statistics. We say that a statistic $T = T(D_1, \ldots, D_n)$ is *trend monotonic* if $T(D_1, \ldots, D_n) \geq T(D'_1, \ldots, D'_n)$ when there is a stronger downward trend in $D_1, \ldots, D_n$ than in $D'_1, \ldots, D'_n$.

From Lehmann (1966) and Bickel and Doksum (1969) we can conclude:

THEOREM 4.1. *If* $-c(i/(n+1))$ *and* $J(i/(n+1))$ *are nondecreasing in i, then the linear rank and spacings statistics W and S are trend monotonic.*

Recall that a similar test is one where the probability of rejecting H_0 when H_0 is true is the same for all values of the scale parameters λ. This probability is the significance level α. Tests that reject H_0 when $T \geq k$, where k is a critical constant and T is trend monotonic, are similar. This is because the downward trend in $\lambda D_1, \ldots, \lambda D_n$ is the same as that of $D_1, \ldots, D_n$, thus $T(D_1, \ldots, D_n) = T(\lambda D_1, \ldots, \lambda D_n)$.

From Lemma 4.1 and Theorem 4.1 we get the following important result.

COROLLARY 4.1. *Let* $\beta(T, F)$ *denote the power of the test that rejects* H_0 *when* $T \geq k$, *where T is trend monotonic. Then the test is unbiased and has isotonic power with respect to the IFR ordering, i.e. if F is in the IFR class, then the power* $\beta(T, F)$ *is greater than the significance level* $\alpha = \beta(T, K)$, *and if F is more IFR than G, then* $\beta(T, F) \geq \beta(T, G)$.

At this point, we have two classes of tests that are good for the nonparametric IFR class in the sense of being unbiased and having isotonic power. In each of the two classes of tests, we can obtain the asymptotically optimal test for a parametric alternative $f(t; \theta, \lambda)$, by choosing

$$c_h(u) = \frac{1}{1-u}\int_{-\log(1-u)}^{\infty} h'(t)\,\mathrm{e}^{-t}\,\mathrm{d}t \tag{4.4}$$

where

$$h(t) = \frac{\partial}{\partial\theta}\log f(t; \theta, 1)\Big|_{\theta=\theta_0}$$

as in Section 3. The resulting statistics W_h and S_h can be shown to be asymptotically equivalent (in the sense of having the same asymptotic power) to the $C(\alpha)$ test based on $T(h)$ given in Section 3. Thus the rank and spacings tests with c given by (4.4) are asymptotically most powerful for $f(t;\theta,\lambda)$ in the sense of maximizing the asymptotic power. See also Bickel (1969). Formula (4.4) was used to compute the examples given (in (4.3)).

Let

$$\bar{c} = n^{-1}\sum_{i=1}^{n} c(i/(n+1)) \quad \text{and} \quad s_c^2 = n^{-1}\sum_{i=1}^{n}(c_i-\bar{c})^2,\ c_i = c\left(\frac{i}{n+1}\right),$$

then the distribution of both $\sqrt{n}(W-\bar{c})/s_c$ and $\sqrt{n}(S-\bar{c})/s_c$ converge to a standard normal distribution under H_0. Thus approximate level α tests based on W and S reject H_0 when these quantities exceed the upper level α critical value c_α of a standard normal distribution.

Note that using integral approximations to sums, $\bar{c}$ and s_c^2 can be approximated by $\mu(c)$ and $\sigma^2(c)$ where

$$\mu(c) = \int_0^1 c(u)\,du \quad \text{and} \quad \sigma^2(c) = \int_0^1 c^2(u)\,du - \mu^2(c).$$

For the four examples c_G, c_W, c_L and c_E of (4.3), we find

$$\mu(c_G) = 1,\quad \sigma^2(c_G) = \tfrac{1}{6}\pi^2 - 1,\qquad \mu(c_W) = 1 - E,\quad \sigma^2(c_W) = \tfrac{1}{6}\pi^2,$$
$$\mu(c_L) = -1,\quad \sigma^2(c_L) = 1,\qquad \mu(c_E) = -\tfrac{1}{2},\quad \sigma^2(c_E) = \tfrac{1}{12},$$

where $E = 0.5772$.

In the case of $c_M(u) = -u$, we have $\bar{c}_M = -\frac{1}{2}$ and $s_M^2 = \frac{1}{12}(n-1)/(n+1)$.

Example 4.1. For the tractor data of Example 3.1, we find $S_L = -0.689$, $\bar{c}_L = -0.979$, $s_L = 0.931$ and $\sqrt{n}(S_L - \bar{c}_L)/s_L = 3.22$ which should be compared with the 'asymptotic equivalent' value $T_L = 2.68$ of Example 3.1. Similarly, $S_M = -0.370$ and $\sqrt{n}(S_M - \bar{c}_M)/s_M = 4.69$ as compared with $T_M = 4.20$ is Example 3.1. Clearly, S_L and S_M both reject exponentiality. Note that, in this example, the spacings tests appear to do better than the $C(\alpha)$ tests.

Finally, we remark that in terms of finite sample size Monte Carlo power, the spacing tests were shown in Bickel and Doksum (1969) to do better than the rank tests.

5. Test based on the total time on test transform

In this section, we introduce another transformation and other test statistics whose distributions are sensitive to IFR models. Suppose we put n independent items on test at the same time. Let $T_{(1)} < \cdots < T_{(n)}$ denote their ordered failure times. At time $T_{(i)}$, the total time the n items have spent on test is

$$TT_i = nT_{(1)} + (n-1)(T_{(2)} - T_{(1)}) + \cdots + (n+1-i)(T_{(i)} - T_{(i-1)})$$
$$= \sum_{i=1}^{i} (n+1-j)(T_{(j)} - T_{(j-1)}) = \sum_{j=1}^{i} D_j$$

where $T_{(0)} = 0$. Note that $TT_n = \sum_{j=1}^{n} D_j = \sum_{i=1}^{n} T_{(i)}$.

The transformation considered in this section is the one that transforms the survival or failure times $T_1, \ldots, T_n$ into $W_1, \ldots, W_{n-1}$, where

$$W_i = \frac{\sum_{j=1}^{i} D_j}{\sum_{j=1}^{n} D_j}.$$

We call $W_1, \ldots, W_{n-1}$ *total time on test transforms*, or *total time transforms* for short. Under H_0, $W_1, \ldots, W_{n-1}$ are distributed as the order statistics in a sample of size $n-1$ from a distribution uniform on (0, 1) (Epstein, 1960).

This transformation is useful since W_i tends to be larger for an IFR distribution than it is for an exponential distribution, more precisely:

THEOREM 5.1. *Suppose F is more IFR than G. Let $T_1, \ldots, T_n$ be a sample from F with corresponding total time transforms $W_1, \ldots, W_{n-1}$. Then there is a sample $T'_1, \ldots, T'_n$ with distribution G and total time transforms $W'_1, \ldots, W'_n$ with*

$$W_i \geq W'_i, \quad i = 1, \ldots, n-1.$$

The proof can be found in Barlow and Proschan (1966), Barlow and Doksum (1972), and Barlow, Bartholomew, Bremner and Brunk (1972).

The result suggests using tests based on statistics that are *monótonic* in the W's in the sense that they are coordinate-wise increasing, i.e.

$$T(W_1, \ldots, W_{n-1}) \geq T(W'_1, \ldots, W'_{n-1})$$

whenever $W'_n \geq W'_i$, $i = 1, \ldots, n-1$.

For such tests we find

THEOREM 5.2. *Let $\beta(T, F)$ denote the power of the test that rejects H_0 when $T \geq k$, where T is monotonic. Then the test is monotonic and has isotonic power with respect to the IFR ordering, i.e. if F is in the IFR class, then the power $\beta(T, F)$ is greater than the significance level α, and if F is more IFR than G, then $\beta(T, F) \geq \beta(T, G)$.*

One important monotonic statistic is the total time on the test statistic which is defined by

$$V = \sum_{i=1}^{n-1} W_i .$$

Since V is distributed as the sum of uniform variables under H_0, its distribution is very close to normal. The exact distribution is tabled in Barlow et al. (1972) for $n \leq 12$. For $n > 9$,

$$\sqrt{12(n-1)}\left[\frac{V}{n-1} - \tfrac{1}{2}\right]$$

has practically a standard normal distribution.

A little algebra shows that

$$V = (n+1)\left(\frac{n}{n+1} + S_M\right)$$

where $S_M = -\sum_{i=1}^{n} iD_i / \sum_{i=1}^{n} D_i$ as in Section 4. Thus V is equivalent to S_M, asymptotically equivalent to T_M, and asymptotically most powerful for the Makeham alternative $f_M(t; \theta, \lambda)$.

Barlow and Doksum (1972) investigated a more general class of monotonic statistics, namely

$$V_J = \sum_{i=1}^{n-1} J(W_i)$$

where J is some nondecreasing function on (0, 1). They found that for a given parametric alternative $f(t; \theta, \lambda)$, the test based on V_J will be asymptotically most powerful if $J(u)$ is chosen to equal $-c(u)$ where $c(u)$ is the function given in (4.1) and (4.2). Thus for the linear failure rate alternative $f_L(t, \theta, \lambda)$, $-\sum_{i=1}^{n-1} \log(1 - W_i)$ is asymptotically optimal, while for the Weibull alternative $f_W(t; \theta, \lambda)$, $\sum_{i=1}^{n-1} \log[-\log(1 - W_i)]$ is asymptotically optimal.

Other tests based on the spacings D_i or total time transforms W_i, have been considered by Störmer (1962), Seshadri, Csörgő and Stephens (1969), Csörgő, Seshadri and Yalovsky (1975), Koul (1978), Azzam (1978), Parzen (1979) and Csörgő and Révész (1981b), among others. An excellent source for results on spacings is the paper by Pyke (1965).

6. Nonparametric optimality

In Sections 3, 4 and 5, we have seen that different IFR parametric alternatives lead to different asymptotically optimal tests. Thus we have no basis on which to choose one test as being better than the others.

In this section, we outline the development of a theory that leads to one test, namely the one based on the total time on test statistc V, as being asymptotically optimal. These results are from Barlow and Doksum (1972).

We define the *total time on test transform* $\mathcal{H}_F^{-1}$ of the distribution function F as

$$\mathcal{H}_F^{-1}(u) = \int_0^{F^{-1}(u)} [1 - F(v)]\, dv, \quad 0 < u < 1,$$

and the standardized total time on test transform as

$$H_F^{-1}(u) = \mathcal{H}_F^{-1}(u)/\mathcal{H}_F^{-1}(1), \quad 0 < u < 1.$$

Note that $H_F^{-1}(1) = E_F(T_i)$ = mean of T_i. The reason for the inverse notation is that H_F^{-1} can be regarded as the inverse of a distribution in (0, 1). We let H or H_F denote this distribution. Note that W_i of the previous section can be regarded as $H_{F_n}^{-1}(i/n)$ where F_n is the empirical distribution function of $T_1, \dots, T_n$.

It is easy to check that when F is exponential, $H(u) = u$, $0 \leq u \leq 1$; while F is IFR iff $H(t)$ is convex and $H(t) \leq t$ on [0, 1]. Thus the problem of testing for exponentiality can be formulated in terms of H as testing

H_0: H is uniform on [0, 1]

vs.

H_1: H is convex, $H(t) \leq t$ and H is not uniform on [0, 1].

The optimality criteria we are going to consider is the minimax criteria, i.e. we want to find the test that maximizes (asymptotically) the minimum power over a nonparametric class Ω. The term minimax is used since in decision theory terminology, risk = 1 − power.

We cannot take Ω to be the whole IFR class since then the minimum power would always be α. The total time on test transform H_F^{-1} gives us a convenient way of separating alternatives from H_0. We let $\Omega(\Delta)$, $0 < \Delta < 1$, be the class of all distributions F where H is convex, $H(t) \leq t$, and

$$\sup[t - H(t)] \geq \Delta.$$

If Δ is fixed, the minimum power over $\Omega(\Delta)$ will tend to one, thus we must allow $\Delta = \Delta_n$ to depend on n, in fact the interesting cases have

$$\Delta_n = O(n^{-1/2}).$$

Let $\beta(\varphi_T, F)$ denote the power of the level α Total Time on Test test which rejects H_0 when $V \geq k_\alpha$, then

Lemma 6.1. *Assume that* $\lim_{n\to\infty}(\sqrt{n}\Delta_n)$ *exists and equals* c *where* c *is some number in* $[0, \infty]$, *then*

$$\lim_{n\to\infty} [\inf_{F\in\Omega(\Delta_n)} \beta(\varphi_T, F)] \leq \Phi(-k_\alpha + \sqrt{3}c).$$

Now suppose that $\beta(\varphi_J, F)$ denotes the power of the test which rejects H_0 when $V_J = \Sigma_{i=1}^{n-1} J(W_i)$ is greater than the appropriate critical constant. We want to choose J to maximize the limiting minimum power. This is achieved by choosing $J(w) = w$; thus the Total Time on Test test φ_T is optimal in the sense of being asymptotically minimax. The result follows from the fact that if $\lim_{n\to\infty}(\sqrt{n}\Delta_n) = c$, $c \in [0, \infty]$, then

$$\lim_{n\to\infty} [\inf_{F\in\Omega(\Delta_n)} \beta(\varphi_J, F)] \leq \Phi(-k_\alpha + \sqrt{3}c) .$$

The proof can be obtained (under appropriate conditions) from Barlow and Doksum (1972) and Koul and Staudte (1976).

7. Distance statistics

If there is no natural alternative class of distributions (such as the IFR class), one can use statistics based on the distance between the exponential distribution $K_\lambda(t)$ and the empirical distribution $F_n(t)$ defined as $F_n(t) = n^{-1}[\#T_i \leq t]$. If $\lambda = \lambda_0$ is specified, the Kolmogorov statistic is given by

$$D_n(\lambda_0) = \max_t |F_n(t) - K_{\lambda_0}(t)| .$$

For tables, see Owen (1962).

In the more realistic case with λ unknown, we replace λ in K_κ by $\hat{\lambda} = 1/\bar{t}$ and use

$$D_n^* = \max_t |F_n(t) - K_{\hat{\lambda}}(t)|$$

where $K_{\hat{\lambda}}(t) = 1 - \exp(-\hat{\lambda}t)$.

The distribution of D_n^* has been studied by Lilliefors (1969), Stephens (1974) and Durbin (1975), among others. A very good approximation to the level α critical values k_α of D_n^* for $\alpha = 0.01$, 0.05 and 0.10 are given by

$$k_\alpha = \frac{0.2}{n} + \frac{d_\alpha}{\sqrt{n} + 0.26 + (0.5/\sqrt{n})}$$

where $d_\alpha = 1.308$, 1.094, 0.990 for $\alpha = 0.01$, 0.05, 0.10, respectively.

An alternative approach to estimating λ in $D_n(\lambda)$ is to first make a transformation of $T_1, \ldots, T_n$ to obtain new variables whose distribution does not depend on λ. Thus we could use the distance between the empirical distribution of $W_1, \ldots, W_{n-1}$ and the uniform distribution on (0, 1). The distribution of the resulting statistic is the same as that of the one-sample Kolmogorov statistic. Tables can be found in Owen (1962, p. 423).

For other distance statistics and their properties, see Seshadri, Csörgő and Stephens (1969), Durbin (1973, 1975), Csörgő, Seshadri and Yalovsky (1975), Sarkadi and Tusnady (1977), and Csörgő and Révész (1981(a)).

8. Graphical method in the uncensored case

8a. The Q–Q (Quantile–Quantile) Plot

The exponential quantile function evaluated at the population distribution function is

$$Q_F(t) = K^{-1}[F(t)]$$

where $K^{-1}(u) = -\log(1-u)$ is the inverse of the exponential distribution. If the exponential hypothesis is satisfied and in fact $F(t) = 1 - e^{-\lambda t} = K(\lambda t)$, then we find

$$Q_K(t) = \lambda t .$$

Thus a graphical method for checking exponentiality is to plot

$$Q_{F_n}(t) = K^{-1}[F_n(t)] = -\log[1 - F_n(t)]$$

and check if this plot falls close to a straight line through the origin. Since we cannot use the log of zero, we use the modification

$$\hat{Q}(t) = K^{-1}\left[F_n(t) - \frac{1}{2n}\right]$$

and plot $\hat{Q}(t)$ for $t = t_{(i)}$, $i = 1, \ldots, n$, where $\{t_{(i)}\}$ are the order statistic of the sample. Since the $t_{(i)}$ are sample quantiles, the resulting plot of $(t_{(i)}, K^{-1}[i - \frac{1}{2}/n])$ is called a Q–Q plot.

The reliability of $\hat{Q}(t)$ can be judged by giving the simultaneous level α confidence band

$$[K^{-1}(\hat{F}_n(t) - k_\alpha),\ K^{-1}(\hat{F}_n(t) + k_\alpha)] ,$$

where k_α is the level α critical value for the D_n^* test of Section 7. We reject exponentiality if the line $t/\bar{t}$ does not fall entirely within the band. This graphical test is equivalent to the D_n^* test of the previous section.

Note that, using Section 4, a convex shape for $\hat{Q}(t)$ indicates an IFR alternative.

8b. The total time on test plot

Barlow and Campo (1975) demonstrated that

$$H_n^{-1}(u) = H_{F_n}^{-1}(u), \quad 0 < u < 1,$$

where H_F^{-1} is the standardized total time on test transform of F defined in Section 6, gives a useful plot for checking exponentiality. Under exponentiality, H_n^{-1} should fall close to the identity function on (0, 1), while for IFR alternatives, we would expect $H_n^{-1}(t) \geq t$ and $H_n^{-1}(t)$ concave (see Section 6). Figure 8.1 shows this plot for the tractor data of Example 3.1. An IFR distribution is strongly indicated for this data.

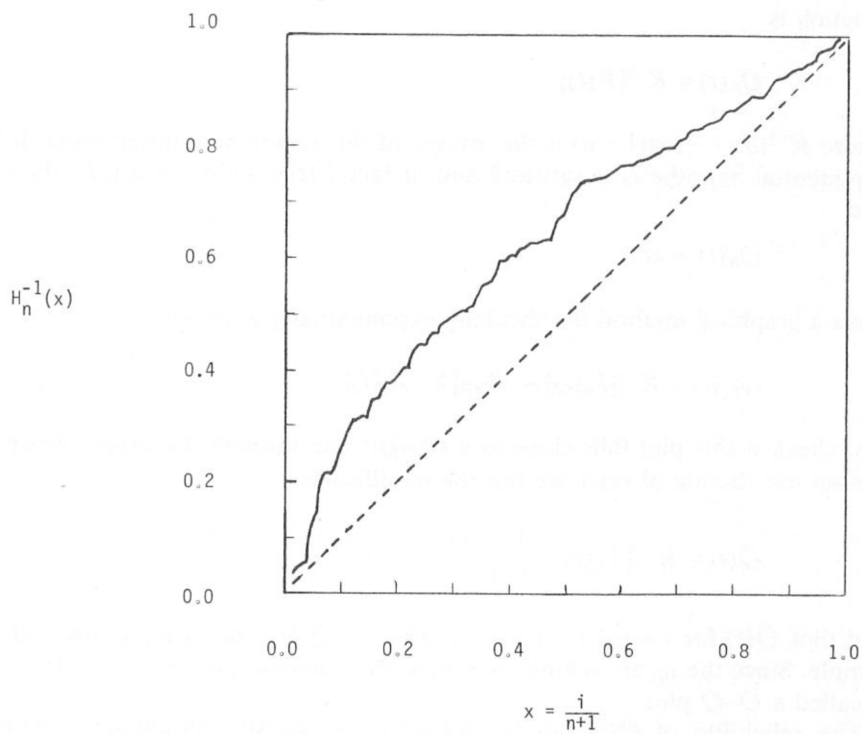

Fig. 8.1. Total time on test plot for the tractor data of Table 3.1.

The reliability of $H_n^{-1}(u)$ can be judged by using the asymptotic simultaneous level α confidence band

$$\left[H_n^{-1}(u) - \frac{b_\alpha}{\sqrt{n}},\ H_n^{-1}(u) + \frac{b_\alpha}{\sqrt{n}}\right], \quad 0 < u < 1,$$

where b_α is the critical value of the maximum of the Brownian Bridge on [0, 1]. Thus b_α is given in Owen (1962, p. 439).

9. NBU alternatives

Tests designed to detect NBU alternatives are motivated by measures of the deviation of F from exponentiality towards NBU alternatives. One such measure, considered by Hollander and Proschan (1972), is

$$\gamma(F) = \int_0^\infty \int_0^\infty [S(t)S(v) - S(t+v)]\, \mathrm{d}F(t)\, \mathrm{d}F(v)\,.$$

When F is exponential, $\gamma(F) = 0$, while it is positive when F is NBU. Thus an intuitive rule is to reject exponentiality for large values of $\gamma(F_n)$. Hollander and Proschan (1972) give the appropriate critical values and prove consistency of this test rule.

Koul (1977) considered

$$\alpha(F) = \inf_{t,v \geq 0} \{S(t+v) - S(t)S(v)\}$$

as a measure of NBUness. $\alpha(F)$ is 0 when F is exponential and negative when F is NBU. Koul (1977) gave critical values of the test based on $\alpha(F_n)$ for selected values of α and n.

Deshpande (1983) measures NBUness through

$$\xi(F) = \int_0^\infty [S^2(t) - S(2t)]\, \mathrm{d}F(t)$$

and considers the corresponding test statistic $\xi(F_n)$. He develops the asymptotic distribution and gives the Pitman asymptotic relative efficiencies 0.931, 1.006 and 0.946 of $\xi(F_n)$ to $\gamma(F_n)$ for linear failure rate, Weibull and Makeham alternatives, respectively.

For further results on measures of NBU alternatives, see Koul (1978) and Hollander and Proschan (this volume, Chapter 27).

10. Types of censoring

Censoring may arise in a variety of ways, leading to several possible assumptions about the form of censoring. Here we consider primarily right censoring. For individual i, $i = 1, \ldots, n$, the observed length of life, or time on test, is $Y_i = \min(T_i, C_i)$, in which T_i is the failure time with survival curve $S(t) = P(T_i \geq t)$, and C_i is the censoring time with censoring curve $G(t) = P(C_i \geq t)$. T_i and C_i are assumed to be independent.

'Type I' censoring concerns experiments in which observation is terminated at a predetermined time $C_i = C$, $i = 1, \ldots, n$. Thus a random number of failures are observed. For 'type II' censoring, observation continues until $r \leq n$ failures occur, with r fixed. Type II censoring may arise when one wants at least r failure times, for reasons of power, but cannot afford to wait until all individuals fail.

In many clinical trials, the beginning and end of the observation period is fixed, but individuals may enter the study at any time. This is an example of 'fixed' or 'progressive type I' censoring, in which the C_i, $i = 1, \ldots, n$, are fixed but not necessarily equal.

'Random censorship' refers to experiments in which the censoring times are randomly distributed. This may occur when censoring is due to competing risks, such as loss to follow-up or accidental death. However, T_i and C_i may be dependent, as is the case when individuals are removed from study based on mid-term diagnosis. The lack of independence brings problems of identifiability and interpretation (Horvath, 1980; see Prentice et al. (1978) for review).

Several other possible assumptions deserve mention. Hyde (1977) and Mihalko and Moore (1980) considered left truncation with right censoring. Left truncation may correspond to birth or to entering the risk stage of a disease (Chiang, 1979). Mantel (1967), Aalen (1978), Gill (1980) and others generalize this to arbitrary censoring.

Various authors (Koziol and Green, 1976; Hollander and Proschan, 1979; Koziol 1980; Chen, Hollander and Langberg, 1982) assumed a 'proportional hazards' model for censoring. That is, $G = S^{\beta}$ with β the 'censoring parameter'.

All these types of censoring are special cases of the multiplicative intensity model (Aalen, 1975, 1976, 1978; Gill, 1980). For our purposes, let $N(t)$, $t \geq 0$, be the number of failures in $[0, t]$ and $R(t)$ be the number at risk of failure at time $t \geq 0$. If we are only concerned with right censorship, then $R(t) = \#(Y_i \geq t)$. More generally $R(t)$ must be predictable, that is left-continuous with right-hand limits and depending only on the history of the process $\{N(u), R(u); 0 \leq u \leq t\}$. We assume that for each $t > 0$, the jump $\mathrm{d}N(t)$ is a zero–one random variable with expectation $R(t)\,\mathrm{d}H(t)$, in which $H(t)$ is the cumulative rate, or hazard function. Aalen (1975, 1978) and Gill (1980) and later authors use the fact that

$$N(t) - \int_0^t R(u)\,\mathrm{d}H(u), \quad t \geq 0,$$

is a square-integrable martingale to derive asymptotic properties of the estimators and tests presented below. Note that one does not need to assume continuity of the survival S or censoring G curve.

The remainder of this paper concerns right censorship unless otherwise noted.

11. Estimates in the censored case

The tests presented in later sections embody estimates of the survival curve, the censoring curve, and/or the hazard function. The survival curve is usually estimated by the Kaplan–Meier product limit estimator

$$S_n(t)\begin{cases} = \prod_{\{i|Y_i\leq t\}}\left(1-\frac{1}{R(Y_i)}\right)I(T_i\leq C_i) & \text{if } 0\leq t< Y_{(n)},\\ =0 & \text{if } t>Y_{(n)},\end{cases}$$

with the Efron (1967) convention that the last event is considered a failure. The censoring curve may be estimated in a similar fashion, with the relation

$$G_n(t)S_n(t)=1-R(t^+)/n\,.$$

S_n and G_n are biased but consistent and self-consistent (Efron, 1967). If S is continuous and G is left-continuous, then S_n is asymptotically normal (Breslow and Crowley, 1974). If S and G are both continuous then S_n is strongly uniformly consistent on any finite interval in the support of both S and G (Földes and Rejtö, 1981).

The hazard function is estimated by the Nelson (1969) estimator

$$H_n(t)=\sum_{\{i|Y_i\geq t\}}\left(\frac{I(t_i\leq C_i)}{R(Y_i)}\right)=\int_0^t R^{-1}\,\mathrm{d}N\,.$$

H_n is biased, consistent and asymptotically normal (Aalen, 1978) under the same conditions as those for S_n. It is also strongly uniformly consistent (Yandell, 1983).

Some tests rely on a survival curve estimator based on H_n, namely

$$\hat{S}(t)=\exp(-H_n(t))>t\geq 0\,.$$

$\hat{G}(t)$ is defined in an analogous manner. The properties of these estimates are presented in Fleming and Harrington (1979).

The asymptotic variance V of $\sqrt{n}(H_n-H)$ and of $\sqrt{n}(S_n-S)/S$ has the form (Breslow and Crowley, 1974; Gill, 1980)

$$V(t)=\int_0^t S^{-1}G^{-1}\,\mathrm{d}H\,.$$

It can be estimated consistently by

$$V_n(t)=n\int_0^t R^{-1}(R-1)^{-1}\,\mathrm{d}N=n\sum_{\{i|Y_i\leq t\}}\left(\frac{I(T_i\leq C_i)}{R(Y_i)(R(Y_i)-1)}\right).$$

12. Weak convergence

Several asymptotic tests for censored survival data are based on the weak convergence of the survival curve S_n or the hazard function to a Gaussian process. Throughout this section we assume that S is continuous, G is left-continuous, and censoring and survival (failure) act independently.

Breslow and Crowley (1974) first proved the weak convergence for S_n and H_n with G continuous. Meier (1975) handled the case of fixed censorship for S_n. Aalen (1978) and Gill (1980, 1983) considered the case of G left-continuous. The results can be stated in terms of Brownian motion B on $[0, \infty)$ or a Brownian bridge B^0 on $[0, 1]$ with a time change (Efron, 1967; Gillespie and Fisher, 1979; Hall and Wellner, 1980). Let $\circ$ denote composition.

THEOREM 12.1. *Let* $Z_n = \sqrt{n}(H_n - H)$ *or* $Z_n = \sqrt{n}(S_n - S)/S$. *Then*

$$Z_n \Rightarrow B \circ V, \qquad Z_n/(1+V) \Rightarrow B^0 \circ (V/(1+V))$$

in $D[0, T]$ *for* $T < T_{SG} = \inf\{t;\, S(t)G(t) > 0\}$.

Gill (1981) extended this result to the whole line:

THEOREM 12.2. *Let* $Z_n = \sqrt{n}(S_n - S)/S$. *Then*

$$Z_n/(1+V) \Rightarrow B^0 \circ (V/(1+V))$$

in $D[0, T_{SG}]$. *In addition,*

$$Z_n/(1+V_n) \Rightarrow B^0 \circ (V/(1+V))$$

in $D[0, T_{SG}]$ *provided that*

$$\int_0^{T_{SG}} S^2 \, dV = \int_0^{T_{SG}} SG^{-1} \, dH < \infty .$$

Nair (1980, 1981) and Gill (1983) introduced weight functions which allow weak convergence to weighted versions of B and B^0.

THEOREM 12.3 (Nair 1980). *Let* $Z_n = \sqrt{n}(H_n - H)$ *or* $Z_n = \sqrt{n}(S_n - S)/S$. *Let* q *be continuous and nonnegative on* $[0, 1]$, *and* $T_n \xrightarrow{P} T < T_{GS}$. *Then*

$$Z_n V_n^{-1/2}(T_n) q \circ (V_n/V_n(T_n)) \Rightarrow (Bq) \circ (V/V(T)), \tag{12.1}$$

$$(Z_n/(1+V_n)) q \circ (V_n/(1+V_n)) \Rightarrow (B^0 q) \circ (V/(1+V)) \tag{12.2}$$

on $D[0, T]$.

Gill (1983) proved a similar result on the whole line for a restricted class of weight functions.

THEOREM 12.4 (Gill 1983). *Let* $Z_n = \sqrt{n}(S_n - S)/S$. *Let q be continuous on* $[0, 1]$, *symmetric at* $\frac{1}{2}$, *nondecreasing on* $(0, \frac{1}{2})$,

$$\int_0^1 q^{-2}(t)\,\mathrm{d}t < \infty$$

and $(1-t)q^{-1}(t)$ *nonincreasing near* $t = 1$. *Then*

$$(Z_n/(1+V))q\circ(V/(1+V)) \Rightarrow (B^0 q)\circ(V/(1+V)).$$

These results will be used with various weight functions in later sections.

Csörgő and Horváth (1982a, 1982b) showed that Z_n (for the survival curve or hazard function) can be strongly approximated on $[0, T]$ by a Brownian bridge process. They required continuity of G, but mention in Remark 3.3 (Csörgő and Horvath, 1982) that continuity and independence of competing risks may not be needed (see Horvath, 1980). Their results yield the same test statistics as those available from the weak convergence results. In addition they provide the rate of convergence, and Chung and Strassen type laws of the iterated logarithm (Csörgő and Horvath, 1983).

13. Maximal deviation tests

One class of goodness-of-fit tests relies on the Kolmogorov–Smirnov metric of the maximal deviation of the empirical from the theoretical distribution. Here we exhibit results for a completely specified null distribution (S or H). For the exponential family, $S(x) = \mathrm{e}^{-\lambda x}$ or $H(x) = \lambda x$, one may view these as conservative in the sense that if no choice of λ yields a curve close enough to the empirical curve, then the hypothesis of exponentiality is rejected. In other words, if one cannot place a straight line completely within the $1-\alpha$ confidence bands for $H(t)$, $a_n \leq t \leq T_n$, then the exponential hypothesis is rejected at level α.

The basic result (Aalen, 1976; Gillespie and Fisher, 1979; Hall and Wellner, 1980; Nair, 1980, 1981; Gill, 1980) is for $Z_n = \sqrt{n}(H_n - H)$ or $Z_n = \sqrt{n}(S_n - S)/S$,

$$\sup_{a_n \leq t \leq T_n} \left| \frac{Z_n(t)}{\sqrt{V_n(T_n)}} q\left(\frac{V_n(t)}{V_n(T_n)}\right)\right| \Rightarrow \sup_{a \leq x \leq b} |q(x)B(x)|,$$

$$\sup_{a_n \leq x \leq T_n} \left| \frac{Z_n(t)}{1+V_n(t)} q\left(\frac{V_n(t)}{1+V_n(t)}\right)\right| \Rightarrow \sup_{a \leq x \leq b} |q(x)B^0(x)|.$$

The cited authors restrict attention to the finite intervals $[a_n, T_n]$ with $T_n \xrightarrow{\mathrm{p}} T < T_{SG}$ and $a_n \geq 0$. The limiting distribution then depends upon S and G, with

$$a = V(a_n)/V(T_n) \quad \text{and} \quad b = 1$$

for the convergence to Brownian motion, and

$$a = \frac{V(a_n)}{1 + V(a_n)} \quad \text{and} \quad b = \frac{V(T_n)}{1 + V(T_n)}$$

for the convergence to B^0. In the latter case, if q satisfies the conditions of Theorem 12.4, then one can extend the sup to the whole line for $Z_n = \sqrt{n}(S_n - S)/S$.

The maximal deviation statistics can be inverted to yield simultaneous confidence bands. More precisely, the $1 - \alpha$ confidence band for $H(t)$ is

$$H_n(t) \pm K_{q,\alpha} n^{-1/2}(1 + V_n(t)) q^{-1}\left(\frac{V_n(t)}{1 + V_n(t)}\right)$$

with $K_{q,\alpha}$ the $1 - \alpha$ point of $\sup|qB^0|$.

Consider several choices of q for this band. If $q(u) = 1/(1 - u)$, one gets bands proportional to 1, with asymptotic distribution that of $\sup|B|$. This has distribution (Feller, 1971)

$$\Pr\{\sup_{0 \le u \le 1} |B(u)| < x\} = \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{(-1)^n}{2n + 1} \exp\left(-\frac{(2n + 1)^2 \pi^2}{8x^2}\right) \approx 4\phi(x) - 3$$

in which $\phi(x)$ is the standard normal distribution. With the choice $q(u) = u^{-1/2}(1 - u)^{-1/2}$, the bands are proportional to $[V_n(t)]^{1/2}$, with asymptotic distribution

$$2\Pr\left\{\sup_{a<u<1} \left|\frac{B(u)}{\sqrt{u}}\right| \le x\right\} = 2\Pr\left\{\sup_{a/(1+a)<u<1/2} \frac{B^0(u)}{\sqrt{u(1 - u)}} \le x\right\}$$

tabled by Borovkov and Sycheva (1968). The choice $q(u) = 1$ yields bands proportional to $(1 + V_n(t))$ with asymptotic distribution equivalent to the Kolmogorov–Smirnov distribution tabled by Pearson and Hartley (1976, Table 54) (see Hall and Wellner (1980) for the case $T < T_{SG}$).

Useful approximations to the distributions of $\sup_{a \le x \le b}|q(x)B(x)|$ and $\sup_{a \le x \le b}|q(x)B^0(x)|$ can be found in the papers by Jennen and Lerche (1981) and Jennen (1981).

The above tests are consistent but biased against continuous alternatives. They are distribution-free asymptotically, up to the choice of interval end points. The choice of $q(\cdot)$ is open, with the obvious remark that different choices emphasize different intervals of the survival or hazard function. The Borovkov–Sycheva (1968) type choice is appealing as the bands are then proportionally wider than pointwise confidence intervals. These bands also have equal variance at every point.

Fleming and Harrington (Fleming, O'Fallon, O'Brien and Harrington, 1980; Fleming and Harrington, 1981) introduced a class of Kolmogorov–Smirnov type tests which differ in an important manner from those considered above. They point out (Fleming and Harrington, 1981) that the asymptotic distribution of tests based on (12.1) with $q = 1$ depend on the maximum of a Gaussian process with variance function which depends on the censoring curve,

$$V(t)/V(T) = \int_0^t G^{-1}S^{-1}\,\mathrm{d}H \Big/ \int_0^T G^{-1}S^{-1}\,\mathrm{d}H\,.$$

They claim that such a test 'has the undesirable property that its probability of rejection of [the null hypothesis] based upon information up to time t systematically tends to zero when censorship of data after time t is increased'. The tests based on (12.1) are asymptotically distribution-free (Nair, 1980), but the power against alternatives will certainly depend on the choice of T and the degree of censoring.

Fleming and Harrington (1981) propose instead to examine

$$Z_{n,a}(t) = \int_0^t \tfrac{1}{2}[S^a(u) + \hat{S}^a(u)]\hat{G}^{-1/2}(u)\,\mathrm{d}(H^*(u) - H_u(u))$$

with $a \geq 0$. This converges weakly to a zero mean independent increment Gaussian process with variance

$$V_a(t) = \int_0^t S^{(2a-1)}\,\mathrm{d}H$$

which does not depend on the censoring G. Their statistic is

$$K_{n,a} = \sup_{0\leq t\leq T} (V_{n,a}(T))^{-1/2}|Z_{n,a}(t)|$$

with $V_{n,a}(t) = \int_0^t \hat{S}^{(2a-1)}\,\mathrm{d}H_n$ and $T < T_{SG}$. K_+^a and K^a converge in distribution to $\sup B$ and $\sup|B|$, respectively. The parameter $a > 0$ acts as the weighting factor; the early part of the distribution S is more emphasized if $0 < a < 1$ while the tail is more heavily weighted if $a > 1$. This can be seen by noting that

$$-\mathrm{d}(S^a) = -S^{a-1}\,\mathrm{d}S = S^a\,\mathrm{d}H\,.$$

See Fleming and Harrington (1981). Note that T may be replaced by $T_n \xrightarrow{\mathrm{p}} T$ and the weights and transformations discussed in Section 12 may be used here, with the obvious modifications.

One-sided maximal deviation tests and simultaneous confidence bands arise in an analogous manner. See the above references for details.

14. Spacings and total time on test

Barlow and Proschan (1969) first derived the distribution of the total time on test plots under the exponential hypothesis for censored data. Barlow and Campo (1975) considered several types of censoring, showing the form of the total time on test and indicating how censoring may affect the stochastic ordering of scaled total time on test plots. Others (Lurie, Hartley and Stroud, 1974; Mehrotra, 1982) considered weighted spacings tests under type II censoring. Aalen and Hoem (1978) considered the multiplicative intensity model of Aalen (1978), generalizing earlier results to arbitrary censorship. The Aalen–Hoem approach will be considered here.

We construct a random time change on the counting process of failures to derive a stationary Poisson process under the null hypothesis of exponentiality. The total time on test transform, based on this random time change, has the same distribution as that in the noncensored case. Define

$$\psi(t) = \int_0^t R(u)\,\mathrm{d}u$$

in which $R(u)$ is defined as in Section 10. If $t_0 = 0$ and $t_1 < t_2 < \cdots < t_k$ are the k distinct failure times, assuming no tied failures, then

$$D_i = \int_{t_{i-1}}^{t_i} R(u)\,\mathrm{d}u = \psi(t_i) - \psi(t_{i-1})$$

is the i-th spacing. Aalen and Hoem (1978) show that if the survival curve is $S(t) = \exp(-H(t))$ then

$$N^*(t) = N(\psi^{-1}(t))$$

is a Poisson process with parameter $h(\cdot) = H'(\cdot)$ (their results are more general). If $S(\cdot)$ is exponential then N^* is a stationary process. Hence

$$N^*(\psi(t_i)) = i, \quad i = 1, \ldots, k,$$

and $(D_1, \ldots, D_k)$ has the same distribution as a random sample from $S(\cdot)$. For exponential $S(x) = \mathrm{e}^{-\lambda x}$,

$$\begin{aligned}\Pr\{D_1 > x\} &= \Pr\{t_1 > \psi^{-1}(x)\} = \Pr\{N(\psi^{-1}(x)) = 0\} \\ &= \Pr\{N^*(x) = 0\} = \mathrm{e}^{-\lambda x}.\end{aligned}$$

Thus many results for noncensored data apply to $(D_1, \ldots, D_k)$. The scaled total time on test transform is

$$\psi(t_i)/\psi(t_k), \quad i = 0, 1, \ldots, k\,.$$

This is plotted for some censored data on prostate cancer in Section 17. The tests based on spacings presented in the first part of this paper generalize in a natural way. In particular the cumulative total time on test statistic of Barlow et al. (1972) becomes, for fixed k,

$$V_k = \sum_{i=1}^{k-1} \psi(t_i)/\psi(t_k)\,.$$

15. Other tests

Several other goodness of fit tests have been proposed in the literature for censored data. These include tests based on contingency tables (Mihalko and Moore, 1980), average deviations (Koziol and Green, 1976; Csörgő and Horvath, 1981; Nair, 1980, 1981), generalized ranks (Breslow, 1975; Hyde, 1977; Hollander and Proschan, 1979; Gill, 1980; Anderson et al., 1981; Harrington and Fleming, 1981), and kernel density or failure rate estimators (Yandell, 1983; see Bickel and Rosenblatt, 1973). We briefly present general forms of the average deviation and generalized linear rank tests.

The average deviation, or Cramér–von Mises, tests are based on weighted average deviations, from the null distribution. Let $K_n(x) = V_n(x)/(1 + V_n(x))$ and

$$Z_n(x) = \sqrt{n}\, q(K_n(x))(1 - K_n(x))(S_n(x) - S(x))/S(x)\,.$$

Then the statistics are of the form

$$W_n^i = \int_0^{T_n} |Z_n|^i \,\mathrm{d}K_n, \quad i = 1, 2\,,$$

for specified weight function q. Similar statistics obtain for the hazard function and for the transform based on equation (12.1). Asymptotic distribution of W_n^2 for $q = 1$ corresponds to that of the classical Cramér–von Mises test

$$\psi_n^2 = -n \int (S_n - S)^2 \,\mathrm{d}S_n$$

in the case of no censoring, and is tabled in Pearson and Hartley (1976, Table 54). Koziol and Green (1976) show that ψ_n^2 converges to a distribution which depends on the censoring parameter β of the proportional hazards model (Koziol and Green, 1976). Clearly, the choice of weights $q(\cdot)$ will force emphasis on different aspects of the distribution S.

Generalized linear rank tests take the form

$$\int_0^{T_n} K(s)(\mathrm{d}H_n(s) - \mathrm{d}H^*(s))$$

in which $H^*(t) = \int_0^t I[R(s) > 0]\,\mathrm{d}H(s)$ is the estimable portion of $H(\cdot)$. $K(t)$ is some function of the history of the survival process, $\{(N(u), R(u)),\ u \in [0, t]\}$. If $K(t) = R(t)$, this becomes, with $T_n = Y_{(n)}$,

$$N(T_n) - \int_0^{T_n} T\,\mathrm{d}H = N(T_n) + \sum_{i=1}^{n} \log(S(Y_i))$$

which is equivalent to Breslow's (1975)

$$\left(N(T_n) - \int_0^{T_n} R\,\mathrm{d}H\right)^2 \Big/ \int_0^{T_n} R\,\mathrm{d}H$$

which converges to chi square with one degree of freedom. Hyde's (1977) statistic is a modification of this to allow left truncation. The asymptotic theory for general $K(\cdot)$ is presented in Anderson et al. (1981) and Gill (1980). Finally we mention that Burke (1982) has constructed a test for the hypothesis that both T and C have exponential distributions.

16. Simulation results

A few Monte Carlo results concerning goodness of fit tests with censored data are available. Koziol (1980) compared the censored Kolmogorov–Smirnov test $\hat{D}_n$ of Hall and Wellner (1980), the Cramér–von Mises statistic

$$\hat{W}_n^2 = n \int_0^T (S_n - S)^2/(S^2(1 + V_n)^2)\,\mathrm{d}(V_n/(1 + V_n))$$

and a 'traditional analogue' of the Cramér–von Mises statistic

$$\psi_n^2 = -n \int_0^T (S_n - S)^2\,\mathrm{d}S_n\,.$$

He considered scale ($S(t) = \mathrm{e}^{-\lambda t}$) and Weibull ($S(t) = \exp(-t^\gamma)$) alternatives to the unit exponential in the Koziol–Green (1976) proportional hazards model, with 1000 trials and sample sizes 20 and 50. At level 0.05, ψ_n^2 had the right size, with $\hat{W}_n^2$ a close second. The size of $\hat{D}_n$ was between 0.066 and 0.107, depending on the degree of censoring ($\beta = 0.5, 1$). ψ_n^2 and W_n^2 had better power against Weibull alternatives, but D_n and W_n^2 had more power than ψ_n^2 against scale alternatives. One may be surprised that D_n performed as well as it did, since the alternatives represent small changes along the whole distribution rather than marked change at any one point. The power of D_n against Weibull alternatives dropped from 0.904 to 0.576 as the censoring parameter increased from 0.5 to 1. This suggests looking at the statistics of Fleming and Harrington (1981).

Unfortunately (for our situation), the simulations of Fleming et al. (1980) (Fleming and Harrington, 1981; Harrington and Fleming, 1981) were only done for 2-sample situations. Further, these simulations concern statistics which differ from those considered here. They examine variations on Kolmogorov–Smirnov tests and several linear rank tests.

Hollander and Proschan (1979) compare the Cramér–von Mises statistic ψ_n^2 with two linear rank statistics.

17. Data analysis

Data were obtained from Hollander and Proschan (1979) on 211 patients with stage IV prostate cancer who were treated with estrogen in a Veterans Administration Cooperative Urological Research Group (1967) study. The observations span the years 1967 through March, 1977. Ninety patients died of prostate cancer, 105 died of other diseases and 16 were alive in March, 1977. The live patients and deaths from other causes were counted as censored.

Koziol and Green (1976) failed to reject the hypothesis of exponentiality with parameter $\lambda = 1/100$. Using the ψ_n^2 Cramér–von Mises statistic with the data truncated at an earlier date, Hollander and Proschan (1979) could not reproduce the earlier value of ψ_n^2, but their value and those of the Hyde (1977) and their own test were not significant at $\alpha = 0.10$. The significance probabilities of the tests varied considerably (0.86, 0.49, 0.14, respectively). Csörgő and Horvath (1981) state that Koziol has computed the Cramér–von Mises W_n^2, the Kolmogorov–Smirnov D_n, and the Kuiper statistic with p-values of 0.15, 0.1, and 0.04, respectively. The ordering of p-values reflects the deviation of S_n from S in Figure 1 of Hollander and Proschan (1979). Csörgő and Horvath's (1981) version of the Cramér–von Mises test is somewhat more significant ($p = 0.0405$).

Our graphical tests indicate that the data may not be exponential, or may at least be a borderline situation. Figure 17.1 is the total time on test plot, showing the same criss-cross of the exponential case curve as seen in Figure 1 of Hollander and Proschan (1979). The hazard function plot of Figure 17.2 suggests that the data may be exponential over most of its range, but the rate appears to taper off. Figures 17.3 and 17.4 are both transformations of the survival curve (see Nair, 1981). Confidence bands are 80% based on the Borokov–Sycheva (1968) weights.

The P–P plot in Figure 17.3 shows some discrepancy with the exponential. This is a plot of

$$(u, S_n(S^{-1}(u))), \quad 0 \leq u \leq 1 ,$$

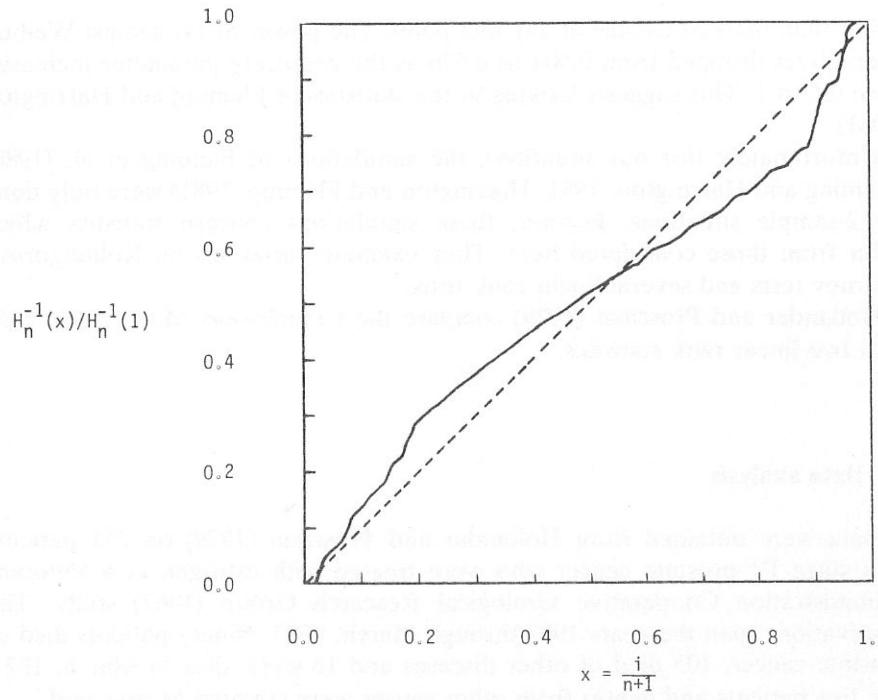

Fig. 17.1. Total time on test plot for the prostate data.

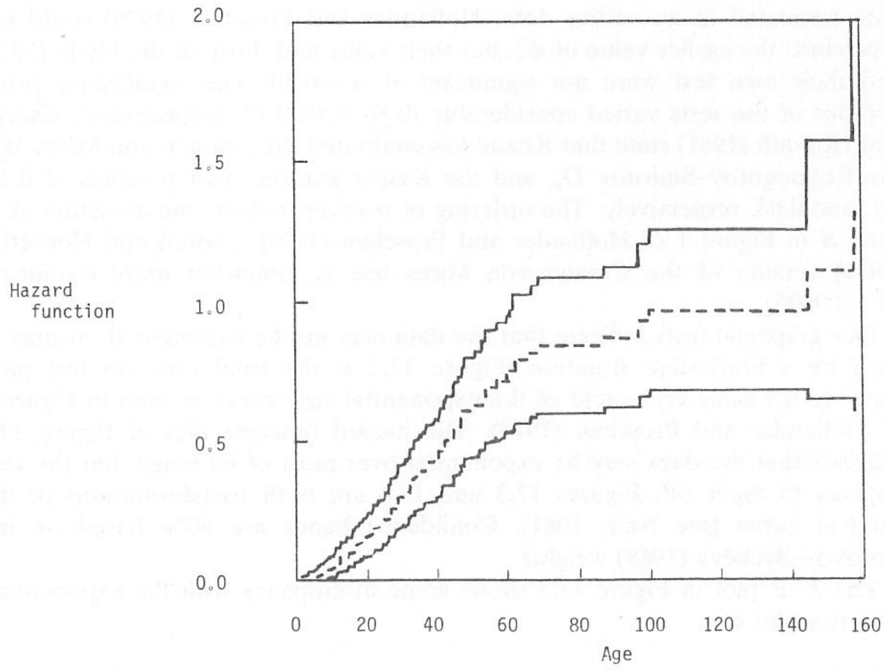

Fig. 17.2. Hazard function plot for the prostate data.

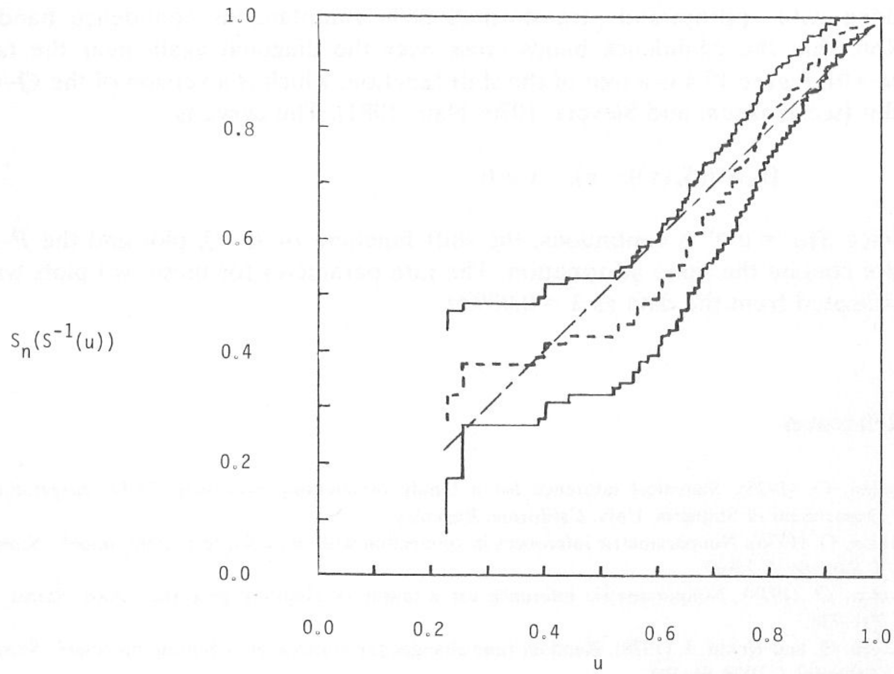

Fig. 17.3. *P–P* plot with 80% simultaneous confidence band for the prostate data. The straight line (diagonal) represents the exponential hypothesis.

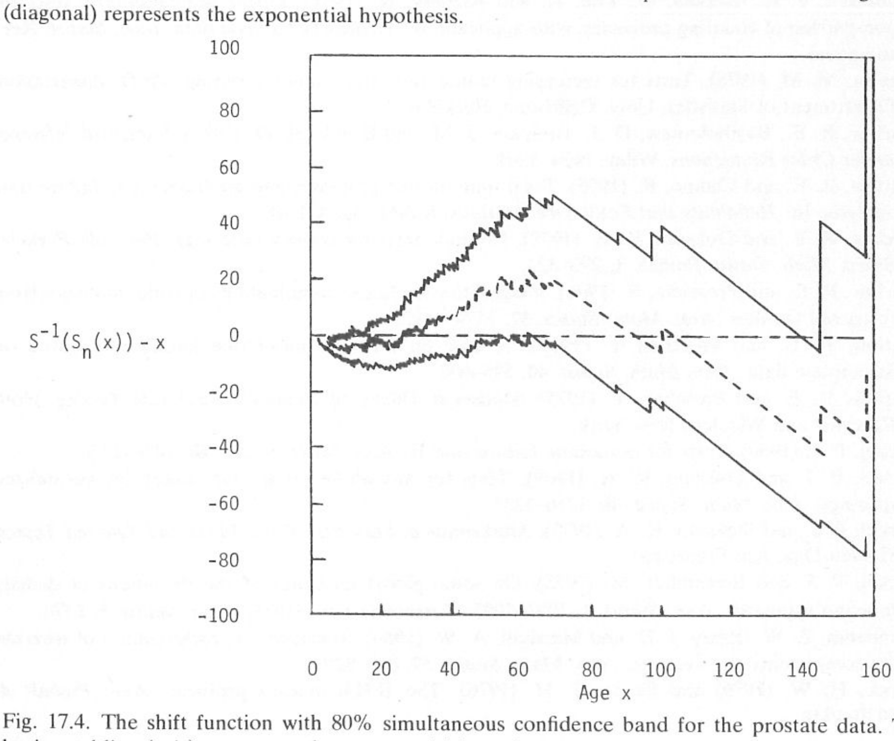

Fig. 17.4. The shift function with 80% simultaneous confidence band for the prostate data. The horizontal line (axis) represents the exponential hypothesis.

along with appropriately transformed 80% simultaneous confidence bands. Note that the confidence bands cross over the diagonal again near the tail ($u \to 0$). Figure 17.4 is a plot of the shift function, which is a version of the Q–Q plot (see Doksum and Sievers, 1976; Nair, 1981). The curve is

$$(x, S^{-1}(S_n(x)) - x), \quad x \geq 0 .$$

Since $S(u) = e^{-u\lambda}$ is continuous, the shift function, or Q–Q, plot and the P–P plot contain the same information. The rate parameter for these two plots was estimated from the data as $\hat{\lambda} = 0.00939$.

References

Aalen, O. (1975). Statistical inference for a family of counting processes. Ph.D. dissertation, Department of Statistics, Univ. California, Berkeley.

Aalen, O. (1976). Nonparametric inferences in connection with multiple decrement models. *Scand. J. Statistic.* **3**, 15–27.

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6**, 701–726.

Aalen, O. and Hoem, J. (1978). Random time changes for multivariate counting processes. *Scand. Actuarial J. 1978*, 81–101.

Andersen, P. K., Borgan, O., Gill, R. and Kieding, N. (1981). Linear nonparametric tests for comparison of counting processes, with applications to censored survival data. *Inter. Statist. Rev.*, to appear.

Azzam, M. M. (1978). Tests for increasing failure rate and convex ordering. Ph.D. dissertation, Department of Statistics, Univ. California, Berkeley.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions.* Wiley, New York.

Barlow, R. E. and Campo, R. (1975). Total time on test processes and applications to failure data analysis. In: *Reliability and Fault Tree Analysis*, SIAM, pp. 451–481.

Barlow, R. E. and Doksum, K. A. (1972). Isotomic tests for convex orderings. *Proc. 6th Berkeley Symp. Math. Statist. Probab.* **I**, 293–323.

Barlow, R. E. and Proschan, F. (1966). Inequalities for linear combinations of order statistics from restricted families. *Ann. Math. Statist.* **37**, 1574–1592.

Barlow, R. E. and Proschan, F. (1969). A note on tests for monotone failure rate based on incomplete data. *Ann. Math. Statist.* **40**, 595–600.

Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing.* Holt, Rinehart and Winston, New York.

Bickel, P. J. (1969). Tests for monotone failure rate II. *Ann. Math. Statist.* **40**, 1250–1260.

Bickel, P. J. and Doksum, K. A. (1969). Tests for monotone failure rate based on normalized spacings. *Ann. Math. Statist.* **40**, 1216–1235.

Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics.* Holden-Day, San Francisco.

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1**, 1071–1095. Correction note (1975), *Ann. Statist.* **3**, 1370.

Birnbaum, Z. W., Esary, J. D. and Marshall, A. W. (1966). Stochastic characterization of wearout for components and systems. *Ann. Math. Statist.* **37**, 816–825.

Block, H. W. (1975) and Savits, T. H. (1976). The IFRA closure problem. *Ann. Probab.* **4**, 1030–1032.

Borovkov, A. A. and Sycheva, N. M. (1968). On asymptotically optimal non-parametric criteria. *Theory Probab. Appl.* **13**, 359–393.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* **2**, 437–453.

Burke, M. D. (1982). Tests for exponentiality based on randomly censored data. *Colloquia Math. Soc. J. Bolyai* **32**, 89–101.

Chen, Y. Y., Hollander, M. and Langberg, N. A. (1982). Small-sample results for the Kaplan–Meier estimator. *J. Amer. Statist. Assoc.* **77**, 141–144.

Csörgő, M. and Révész, P. (1981a). *Strong Approximations in Probability and Statistics.* Academic Press, New York.

Csörgő, M. and Révész, P. (1981b). Quantile processes and sums of weighted spacings for composite goodness-of-fit. In: M. Csörgő, D. A. Dawson, J. N. K. Rao and A. K. Md. E. Saleh, eds., *Statistics and Related Topics.* North-Holland, Amsterdam, pp. 69–87.

Csörgő, M., Seshadri, V. and Yalovsky, M. (1975). Applications of Characterizations in the Area of Goodness-of-Fit. In: C. P. Patil, K. Kotz and J. K. Ord, eds., *Statistical Distributions in Scientific Work, Vol. 2.* Reidel, Boston, pp. 79–90.

Csörgő, S. and Horváth, L. (1981). On the Koziol–Green model for random censorship. *Biometrika* **68**, 391–401.

Csörgő, S. and Horváth, L. (1982a). On cumulative hazard processes under random censorship. *Scand. J. Statist.* **9**, 13–21.

Csörgő, S. and Horváth, L. (1982b). On random censorship from the right. *Acta Sci. Math.* (Szeged), **44**, 23–34.

Csörgő, S. and Horváth, L. (1983). The rate of strong uniform consistency for the product-limit estimator. *Z. Warsch. Verw. Geb.* **62**, 411–462.

Doksum, K. A. and Sievers, G. L. (1976). Plotting with confidence: Graphical comparisons of two populations. *Biometrika* **63**, 421–434.

Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function.* Society for Industrial and Applied Mathematics.

Durbin, J. (1975). Kolmogrov–Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests of spacings. *Biometrika* **62**, 5–22.

Efron, B. (1967). The two-sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. Probabl.* **4**, 831–853.

Epstein, B. (1960). Tests for the validity of the assumption that the underlying distribution of life is exponential. *Technometrics* **2**, 83–101.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications.* Wiley, New York.

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.

Fleming, T. R. and Harrington, D. P. (1979). Nonparametric estimation of the survival distribution in censored data. Technical Report Series No. 8, Mayo Clinic.

Fleming, T. R. and Harrington, D. P. (1981). A class of hypothesis tests for one and two sample censored survival data. *Commun. Statist. A* **10**, 763–794.

Földes, A. and Rejtö, L. (1981). Strong uniform consistency for nonparametric survival curve estimators from randomly censored data. *Ann. Statist.* **9**, 122–129.

Gill, R. D. (1980). Censoring and stochastic integrals. Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam.

Gill, R. D. (1983). Large sample behaviour of the product-limit on the whole line. *Ann. Statist.* **11**, 59–67.

Gillespie, M. J. and Fisher, L. (1979). Confidence bands for Kaplan–Meier survival curve estimate. *Ann. Statist.* **7**, 920–924.

Hall, W. J. and Wellner, J. A. (1980). Confidence bands for a survival curve from censored data. *Biometrika* **67**, 133–143.

Harrington, D. P. and Fleming, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika* **69**, 553–566.

Hollander, M. and Proschan, F. (1972). Testing whether new is better than used. *Ann. Math. Stat.* **43**, 1136–1146.

Hollander, M. and Proschan, F. (1979). Testing to determine the underlying distribution using randomly censored data. *Biometrics* **35**, 393–401.

Hollander, M. and Proschan, F. (1984). Nonparametric concepts and methods in reliability. In: P. R. Krishnaiah and P. K. Sen, eds., *Handbook of Statistics, Vol. 4, Nonparametric Methods,* this volume.

Horváth, L. (1980). Dropping continuity and independence assumptions in random censorship models. *Studia Sci. Math. Hung.* **15**, 381–389.

Hyde, J. (1977). Testing survival under right censoring and left truncation. *Biometrika* **64**, 225–230.

Jennen, C. (1981). Asymptotishe Bestimmung von Kenngrössen sequentieller verfahren. Doctorial Dissertation. University of Heidelberg.

Jennen, C. and Lerche, H. R. (1981). First exit densities of Brownian motion through one-sided moving boundaries. *Z. Wahrsch. Verw. Gebiete* **55**, 133–148.

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

Koul, H. L. (1977). A test for new is better than used. *Comm. Statist. A* **6**, 563–573.

Koul, H. L. (1978a). A class of tests for testing "new is better than used". *Canad. J. Statist.* **6**, 249–271.

Koul, H. L. (1978b). Testing for new is better than used in expectation. *Comm. Statist. A*, **7**, 685–701.

Koul, H. L. and Staudte, Jr., R. G. (1976). Power bounds for a Smirnov statistic in testing the hypothesis of symmetry. *Ann. Statist.* **4**, 924–935.

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 Cramer–von Mises statistic for randomly censored data. *Biometrika* 63, 465–474.

Lilliefors, H. W. (1969). On the Kolmogorov–Smirnov test for the exponential distribution with mean unknown. *J. Amer. Statist. Assoc.* **64**, 387–389.

Lurie, D., Hartley, H. O. and Stroud, M. R. (1974). A goodness of fit test for censored data. *Commun. Statist.* **3**, 745–753.

Mantel, N. (1967). Ranking procedures for arbitrarily restricted observations. *Biometrics* **65**, 311–317.

Mehrotra, K. G. (1982). On goodness of fit tests based on spacings for type II censored samples. *Commun. Statist.* **11**, 869–878.

Meier, P. (1975). Estimation of a distribution from incomplete observations. In: J. Gani, ed., *Perspectives in Probab. and Statistic.*: *Papers in Honour of M. S. Bartlett.* Academic Press, New York, pp. 67–82.

Mihalko, D. P. and Moore, D. S. (1980). Chi square tests of fit for type II censored data. *Ann. Statist.* **8**, 625–644.

Nair, V. N. (1980). Goodness of fit test for multiply right censored data. Tech. Report, 1 Sept. 1980, Bell Tele. Lab., Holmdel, NJ, 22 pp.

Nair, V. N. (1981). Plots and tests for goodness of fit with randomly censored data. *Biometrika* **68**, 99–103.

Nelson, W. (1969). Hazard plotting for incomplete failure data. *J. Qual. Tech.* **1**, 27–52.

Neyman, J. (1959). Optimal asymptotic tests of composite statistical hypothesis. *Probab. and Statist.* The Harold Cramer Volume, Almquist and Wiksells, Uppsala, Sweden, pp. 213–234.

Owen, D. B. (1962). *Handbook of Statistical Tables.* Addison-Wesley, Reading, MA.

Parzen, E. (1979). Nonparametric statistical data modeling. *J. Amer. Statist. Assoc.* **74**, 105–131.

Pearson, E. S. and Hartley, H. O. (1975). *Biometrika Tables for Statisticians, Vol. 2.* Griffin, London.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Jr., Fournoy, N., Farewell, V. T. and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34**, 541–554.

Proschan, F. and Pyke, R. (1967). Tests for monotone failure rate. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **3**, 293–312.

Pyke, R. (1965). Spacings. *J. Roy. Statist. Soc. Ser. B* **27**, 395–436.
Sarkadi, K. and Tusnady, G. (1977). Testing for normality and the exponential distribution. *Proc. Fifth Conf. Probab. Theory*. Brasov, Romania, 99–118.
Seshadri, V., Csörgő, M. and Stephens, M. A. (1969). Tests for the exponential distribution using Kolmogorov-type statistics. *J. R. Statist. Soc. B* **31**, 499–509.
Shorack, G. R. (1972). The best test of exponentiality against gamma alternatives. *J. Amer. Statist. Assoc.* **67**, 213–214.
Spiegelhalter, D. J. (1983). Diagnostic tests of distributional shape. *Biometrika* **70**, 401–410.
Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *J. Amer. Statist. Assoc.* **69**, 730–737.
Störmer, H. (1962). On a test of the exponential distribution. *Metrika* **5**, 128–137.
Veterans Administration Cooperative Urological Research Group. (1967). Treatment and survival of patients with cancer of the prostate. *Surgery, Gynecology, Obstetrics* **124**, 1011–1017.
Yandell, B. S. (1983). Nonparametric inference for rates with censored survival data. *Ann. Statist.* **11**, 1119–1135.
van Zwet, W. (1964). *Convex Transformations of Random Variables*. Math. Centrum, Amsterdam.