The Annals of Statistics
1999, Vol. 27, No. 1, 338–360

TESTING UNIFORMITY VERSUS A MONOTONE DENSITY[1]

BY MICHAEL WOODROOFE AND JIAYANG SUN

University of Michigan and Case Western Reserve University

The paper is concerned with testing uniformity versus a monotone density. This problem arises in two important contexts, after transformations, testing whether a sample is a simple random sample or a biased sample, and testing whether the intensity function of a nonhomogeneous Poisson process is constant against monotone alternatives. A penalized likelihood ratio test (P-test) and a Dip likelihood test (D-test) are developed. The D-test is analogous to Hartigan and Hartigan's (1985) Dip test for bump hunting problems. While nonparametric, both the P- and D-tests are quite efficient in comparison to the most powerful (MP) tests for some simple alternatives and also the Laplace test developed for nonhomogeneous Poisson process. The P- and D-tests have higher power than the above MP tests under different sets of monotone alternatives and so have greater applicability. Moderate sample size performance and applications of our tests are illustrated via simulations and examination of an air-conditioning equipment data set.

1. Introduction. Let f denote a density on the unit interval $(0, 1]$, taken to be continuous from the left, and suppose that a sample $X_1, \ldots, X_n \sim_{\text{ind}} f$ is available. The problem considered here is that of testing the null hypothesis H_0: $f = 1$, that f is the standard uniform density, against the alternative H_1, that f is *nonincreasing* on $(0, 1]$. Two tests of this hypothesis are derived and studied through asymptotic analysis and simulation experiments. The problem arises in two important contexts, after transformations. After a probability integral transformation, the problem of testing whether a sample from a population is a simple random sample against alternatives that involve a monotone selection effect is of the form considered. Also, the problem of testing whether the intensity function of a possibly nonhomogeneous Poisson process is constant against monotone alternatives can be transformed into the form considered here.

The two potential applications are described in more detail in Section 2. A penalized likelihood ratio test (P-test) and a Dip likelihood test (D-test) of the hypotheses are derived in Section 3. The D-test is similar to Hartigan and Hartigan's (1985) Dip test for bump hunting problems, and reduces to a one-sided Kolmogorov–Smirnov test with an optional smoothing parameter in the present context. The P-test starts with a penalized nonparametric maximum likelihood estimator (PNPMLE), $\hat{f}$ say, of f (assumed to be nonincreasing) and

Received July 1997; revised August 1998.
[1]Supported in part by NSF Grants DMS-95-04515, DMS-96-26108 and DMS-96-26347.
AMS 1991 *subject classifications*. Primary 62G10, 62A10; secondary 62G05, 62E25, 62F30, 60F05, 62P20, 62P99.
Key words and phrases. Air-conditioning equipments, penalized maximum likelihood estimates, selection bias, nonhomogeneous Poisson process.

then compares the penalized likelihood at $f = \hat{f}$ to the penalized likelihood at $f = 1$. The properties of these two tests are examined in Section 4 through a heuristic derivation of the asymptotic distributions and simulations of the null distribution. Examples and power studies of the two tests and some most powerful (MP) tests are presented in Section 5. The P- and D-tests are competitive with the MP-tests under the alternatives for which the MP-tests are optimal, and do better than the MP-tests under different alternatives. An application to the air-conditioning equipment data from Moeller (1976) is also provided. The mathematical development is informal in Sections 1–5. Proofs are completed in an Appendix. Some familiarity with Robertson, Wright and Dykstra (1988) [RWD hereafter] is assumed in the mathematical development.

The role of the penalty terms deserves comment. The reader may be familiar with Grenander's (1956) nonparametric maximum likelihood estimator (NPMLE) of a nonincreasing density function. If not, it is briefly described in the Appendix and admirably described by RWD, Section 7.2. Unfortunately, Grenander's estimator is inconsistent at the endpoints and so is not appropriate for some applications, including the present one. The inconsistency is sometimes called the "spiking problem," because the estimator is too big at the left endpoint. It is also too small at the right one. This problem was addressed by Woodroofe and Sun (1993) [hereafter WS], who were especially interested in $f(0+) = \lim_{x \downarrow 0} f(x)$. They showed that a suitable PNPMLE had the same form as Grenander's estimator, but with a deformed x-axis. They further showed that the PNPMLE is consistent at the left endpoint and obtained an asymptotic distribution for rescaled estimation error. Here both endpoints are important, and it is necessary to treat the penalty terms differently from WS. See Section 3 and the Appendix for the details. The inclusion of the penalty term in the likelihood is essential for the P-test. Without it, the contributions from the endpoints dominate, and the unpenalized log-likelihood ratio test statistic is unbounded for large n under the null hypothesis [cf. Groeneboom and Pyke (1983)]. Also the power of the P-test was higher than that of the unpenalized likelihood ratio test in the cases studied in Section 5. The penalty term plays a much smaller role in the D-test, though it does not appear to harm the power and may improve it in some cases.

Testing homogeneity of parameters versus monotonicity of parameters, a related testing problem, has been studied by Cohen, Perlman and Sackrowitz (1990), Wang (1994) and others. See also RWD, Chapter 5. There appears to be little work on nonparametric versions of the problem. In effect, Boswell (1966) proposes an unpenalized likelihood ratio test in the context of testing the rate of a Poisson process. Similarly, Barlow (1968) considers an unpenalized likelihood ratio test of a constant failure rate versus a nondecreasing, or nondecreasing on the average failure rate. The focus here on penalized likelihood and large sample properties is quite different from those of the early authors, however. Cohen and Sackrowitz (1993) compare several methods for testing the rate of a nonhomogeneous Poisson process and provide a good entry into the literature on this question. The application to biased sampling models is motivated by Sun and Woodroofe (1997).

2. Examples. More detail on the two applications is provided in this section.

EXAMPLE 1. *Nonhomogeneous Poisson processes*. Let N_t, $0 \le t < \infty$, be a possibly nonhomogeneous Poisson process (NHPP) with intensity function λ. Thus, N_t, $0 \le t < \infty$, has independent increments and $N_t - N_s$ has a Poisson distribution with mean $\Lambda(t) - \Lambda(s)$ for $0 \le s < t < \infty$, where $\Lambda(t) = \int_0^t \lambda(u)\,du$. In quality control models, for example, N_t might represent the number of defective items produced by time t. In such models, a constant λ represents a stable process, an increasing λ represents system deterioration, and there is interest in testing the two hypotheses. See Ascher and Feingold (1984) for a survey on applications of NHPP's and Misra (1983) for applications in software reliability analysis. Suppose that the process is observed over a time interval $0 \le t \le T$, and let $Y_1 < Y_2 < \cdots < Y_n$ denote the times at which events occur. Then the conditional joint density of $Y_1, \dots, Y_n$ given $N_T = n$, is

$$n!\frac{\lambda(y_1) \times \cdots \times \lambda(y_n)}{\Lambda(T)^n}, \qquad 0 \le y_1 \le \cdots \le y_n \le T.$$

Thus, the conditional joint distribution of $Y_1/T, \dots, Y_n/T$ is the same as the joint distribution of the order statistics of a sample from the density $f(x) = T\lambda(Tx)/\Lambda(T)$, $0 \le x \le 1$, and the problem becomes testing $f = 1$ against the alternative that f is nondecreasing. This is equivalent to the problem considered, as it is easy to adapt our tests for nonincreasing alternatives to ones for nondecreasing alternatives. See also Remark 1 in the next section.

Of course, one might ask how our tests compare with those using the information that the data are from a NHPP. Bain, Engelhardt and Wright (1985) compared, via Monte Carlo simulation, the power functions of six different tests for a constant intensity against the alternative of an increasing intensity function in a NHPP. They suggested using the L-, Z- and W-tests, described below. Cohen and Sackrowitz (1993) have the same conclusion and explain further why these three tests are recommended. They concluded, "... there is essentially no reasonable test based on ranks [of interfailure times] that can have an acceptable power function." Our P- and D-tests are based on a penalized maximum likelihood estimator of a monotone density. If the penalty parameters are zero, the P-test is the same as the W-test, and the D-test is the one-sided Kolmogorov–Smirnov test. It is well known and shown by WS that the unpenalized estimator suffers a spiking problem. So, our P-test with nonzero penalty should work better than the W-test. Indeed, Boswell (1966), the author of the W test, applies a bound (which is often unknown) on the intensity function to control the spiking problem. Our tests are nonparametric since the alternative hypothesis H_1 only specifies that the density be nonincreasing. In the context of a NHPP, the P- and D-tests are based on the order statistics (not ranks) of failure times (not interfailure times), so their competitive performance shown in Section 5 (in comparison to L and Z tests) does not contradict Cohen and Sackrowitz's (1993) conclusion. In Section 5, we show that the P- and D-tests are efficient; that is, they do not lose much power in comparison to the L- and Z-tests under the two specific alternatives for which L and Z are uniformly most powerful unbiased tests, and the P- and D-tests (especially P) do better than L- and Z-tests under other alternatives, for example, some step functions.

EXAMPLE 2 *Testing a sample for bias*. Consider a population with a known, continuous distribution function F. If a simple random sample is taken from the population, then the result is of the form $Y_1, \dots, Y_n \sim_{\text{ind}} F$. On the other hand, if subjects with large values of y are selected with lower probability, then $Y_1, \dots, Y_n \sim_{\text{ind}} F^{\#}$, where

$$dF^{\#}(y) = \frac{w(y)\,dF(y)}{\kappa}, \qquad y \in \mathbb{R},$$

w is a nonnegative, nonincreasing function, and κ denotes a normalizing constant. If F is the standard uniform distribution, then the problem is of the form considered in Section 1. If F is not uniform, a probability integral transformation may be used to transform it to the uniform. A difficulty may arise if F is unknown or partially unknown. We leave this case to future studies. It is encouraging as shown in Section 5 that our P- and D-tests are comparable to the most powerful tests for simple alternatives that require the selection function $w(x)$ to be length biased and power biased.

3. The test statistics.

A Penalized MLE. Let $X_1, \dots, X_n$ be independent and identically distributed with common density f and suppose that f is known to be *nonincreasing* and left continuous on $(0, 1]$. Denote the order statistics by $0 < x_1 < \cdots < x_n < 1$, and let $x_0 = 0$, $x_{n+1} = 1$ and

$$l_{\alpha,\beta}(f) = \sum_{i=1}^{n} \log[f(x_i)] - n\alpha f(0+) + n\beta \log[f(x_n)],$$

where α, $\beta > 0$. Then $l_{\alpha,\beta}$ is called the *penalized log-likelihood function*. The penalty terms $n\alpha f(0+)$ and $n\beta \log[f(x_n)]$ are included to suppress the spiking problem, as in WS. The important difference between the penalty here and that in WS is that both endpoints (rather than one) are penalized here and the two points are penalized in different ways. Of course, there is some subjectivity in the choice of the penalty terms. The choice made here leads to a mathematically tractable problem and performed well in extensive simulations (not reported here) when compared to competitors, like $n\beta f(x_n)$ or $n\beta \log[f(1)]$. See also Remark 2 below.

The next step is to find the penalized maximum likelihood estimator. For this the penalized log-likelihood function must be maximized subject to the constraint that f be a nonincreasing, left-continuous density. It is easily seen that the penalized log-likelihood function is maximized when f is a step function, say $f(x) = f_k$ for $x_{k-1} < x \le x_k$ and $k = 1, \dots, n$ and $f(x) = 0$ for $x > x_n$. The maximizing values of f_k, $k = 1, \dots, n$, may be found in WS, who analyzed the special case $\beta = 0$. Let $c_i = 1/n$ for $i = 1, \dots, n-1$, $c_n = 1/n + \beta$, $w_1 = \alpha + \gamma x_1$, $w_i = \gamma(x_i - x_{i-1})$, $i = 2, \dots, n$, and

$$f_k(\gamma) = \min_{i \le k} \max_{j \ge k} \frac{c_i + \cdots + c_j}{w_i + \cdots + w_j} \tag{1}$$

for $k = 1, \dots, n$ and $\gamma > 0$. The following may be proved as in WS. See the Appendix.

PROPOSITION 1. *If $x_n > \alpha/(1+\beta)$, then the equation $\gamma = 1 + \beta - \alpha f_1(\gamma)$ has a unique positive solution $\hat{\gamma}$, and $l_{\alpha,\beta}(f)$ is maximized by $\hat{f}(x) = \hat{f}_k$, $x_{k-1} < x \le x_k$, $k = 1, \dots, n+1$, where $\hat{f}_k = f_k(\hat{\gamma})$, $k = 1, \dots, n$ and $\hat{f}_{n+1} = 0$.*

The proof is outlined in the Appendix.

In Sections 4 and 5, $\alpha = \alpha_n \to 0$ as $n \to \infty$, so that $x_n > \alpha/(1+\beta)$ with probability approaching 1 (rapidly). It is convenient to let $\hat{\gamma} = 1/n$ (or other small positive constant) and define $\hat{f}$ as in the proposition when $x_n \le \alpha/(1+\beta)$.

The P-test. The first test statistic to be studied is a *penalized likelihood ratio test* (P-test),

$$\Lambda = l_{\alpha,\beta}(\hat{f}) - l_{\alpha,\beta}(1) = \sum_{k=1}^{n} \log[\hat{f}(x_k)] - n\alpha[\hat{f}(x_1) - 1] + n\beta \log[\hat{f}(x_n)]. \tag{2}$$

This Λ has an alternative expression (3) below which is easier for analysis. Write $\hat{W}_0 = 0 = t_0$, $\hat{W}_k = \hat{w}_1 + \cdots + \hat{w}_k = \alpha + \hat{\gamma} x_k$, and $t_k = \hat{W}_k/\hat{W}_n = (\alpha + \hat{\gamma} x_k)/(\alpha + \hat{\gamma} x_n)$ for $k = 1, \dots, n$. Next let G: $[0,1] \to \mathbb{R}$ be a right-continuous step function for which

$$G(t_k) = \begin{cases} k/[n(\alpha + \hat{\gamma} x_n)], & \text{if } k \le n-1, \\ (1+\beta)/(\alpha + \hat{\gamma} x_n), & \text{if } k = n. \end{cases}$$

Let $\tilde{G}$ be the least concave majorant of G and let $\tilde{g}$ be the left-hand derivative of $\tilde{G}$. Then

$$\hat{f}(x_k) = \hat{f}_k = \min_{0 \le i < k} \max_{k \le j \le n} \frac{G(t_j) - G(t_i)}{t_j - t_i} = \tilde{g}(t_k)$$

for $k = 1, \dots, n$, using Theorem 1.4.4 from RWD, page 23. So, the log-likelihood ratio statistic is

$$\begin{aligned} \Lambda &= n(\alpha + \hat{\gamma} x_n) \int_0^1 \log[\tilde{g}(x)]\, dG(x) - n\alpha[\tilde{g}(t_1) - 1], \\ &= n(\alpha + \hat{\gamma} x_n) \int_0^1 \log[\tilde{g}(x)]\, d\tilde{G}(x) - n\alpha[\tilde{g}(t_1) - 1], \end{aligned} \tag{3}$$

where the final equality follows since $\tilde{G}$ and G are equal at the points where $\tilde{g}$ jumps. This can be derived from RWD, Theorems 1.3.3 and 1.3.5, page 17.

It is useful to relate G to the empirical distribution function

$$F^*(x) = \frac{\#\{k \le n\colon x_k \le x\}}{n}.$$

Let $U\colon [0,1] \to [0,1]$ be a continuous piecewise linear function for which $U(t_k) = x_k$ for all $k = 0, \dots, n$. Then

$$G(t) = \frac{1}{\alpha + \hat{\gamma} x_n}\left[F^* \circ U(t) + \beta\delta_1(t)\right]$$

for all $0 \le t \le 1$, where $\circ$ denotes composition, $F \circ U(t) = F[U(t)]$ and δ_b denotes the point mass at b; that is, $\delta_b(t) = 0$ for $t < b$ and $\delta_b(t) = 1$ for $t \ge b$.

The D-test. The second test statistic is an analogue of Hartigan and Hartigan's (1985) Dip test for testing unimodality versus nonunimodality, a bump hunting problem. The idea of the Dip test is to measure the distance of the sample distribution of a sample to the class of all unimodal distributions with support on $(0, 1]$, $\mathscr{U}$ say, by

$$\text{Dip} = \inf_{G \in \mathscr{U}} \sup_{t \in (0,1)} \left|F^*(t) - G(t)\right|,$$

where F^* is the empirical distribution function of the data. Our *Dip likelihood test* (D-test) statistic is

$$D := \sqrt{n} \sup_{t \in (0,1)} \left|\hat{F}(t) - t\right|, \tag{4}$$

where $\hat{F}$ is the cumulative distribution function corresponding to the penalized maximum likelihood estimator $\hat{f}$ (as in Proposition 1). It is easy to see that $\sup_{t \in (0,1)} |\hat{F}(t) - t| = \sup_{t \in (0,1)}[\hat{F}(t) - t]$ when the alternative consists of nonincreasing densities; if the penalty parameters are zero, then $\sup_{t \in (0,1)}[\hat{F}(t) - t] = \sup_{t \in (0,1)}[F^*(t) - t]$. So, the D-test is the one-sided Kolmogorov–Smirnov test when $\alpha = \beta = 0$.

4. Limiting distributions and critical values.

Preliminaries. For the asymptotics, it is convenient to subscript the estimators by n. Thus, write $\hat{\gamma}_n$ for $\hat{\gamma}$, $\hat{f}_{n,k}$ for $\hat{f}_k$, $\hat{f}_n(x)$ for $\hat{f}(x)$, D_n for D and so forth. Consider local alternatives in which the common density of $X_1, \dots, X_n$ is of the form

$$f_n(x) = 1 + \frac{1}{\sqrt{n}}\varphi(x), \qquad 0 < x \le 1,$$

where φ is a bounded, left-continuous, nonincreasing function on $(0, 1]$ for which $\int_0^1 \varphi(x)\,dx = 0$. Then the distribution function of $X_1, \dots, X_n$ is

$$F_n(x) = x + \frac{1}{\sqrt{n}}\Phi(x), \qquad 0 \le x \le 1, \tag{5}$$

where $\Phi(x) = \int_0^x \varphi(y)\,dy$ for $0 < x \le 1$. Of course, H_0 is a special case with $\varphi = 0$.

For the remainder of the paper, suppose that

$$\alpha_n = \beta_n = \frac{c}{\sqrt{n}} \tag{6}$$

for all $n \ge 1$, where $0 < c < \infty$. Then the following results hold, again as in WS.

PROPOSITION 2. *Suppose* $\hat{f}_n(0+) \to 1$ *in probability as* $n \to \infty$.

COROLLARY. *Then* $\hat{\gamma}_n = 1 + o_p(1/\sqrt{n})$, *in probability as* $n \to \infty$.

PROOFS. For $\alpha_n = c/\sqrt{n}$, $\beta_n = 0$ and $\varphi = 0$, Proposition 2 is a special case of Theorem 2 of WS, and the proof given there extends to $\beta_n = c/\sqrt{n}$ without essential change. The general case in which $\varphi \ne 0$ may then be obtained by showing that the distributions of $X_1, \dots, X_n$ under local alternatives are contiguous to their joint distribution under the null hypothesis ($\varphi = 0$) and using Le Cam's second lemma in [Bickel, Klaassen, Ritov and Wellner (1993), page 500]. Alternatively, Proposition 2 may be proved by retracing the steps in WS. The details have been omitted.

The Corollary follows easily from the Proposition since

$$1 - \hat{\gamma}_n = \alpha_n[\hat{f}_n(x_1) - 1]. \qquad \square$$

Asymptotic distributions. An heuristic derivation of the asymptotic distribution of Λ_n is described next. A proof may be found in the Appendix. Write F_n^*, G_n, and so on for F^*, G, and so on and let

$$\mathbb{F}_n(t) = \sqrt{n}[F_n^*(t) - F_n(t)]$$

for $0 \le t \le 1$ and $n \ge 1$. Then there is a sequence $\mathbb{B}_n$, $n \ge 1$, of Brownian bridges for which

$$\sup_{0 \le t \le 1} \left|\mathbb{F}_n(t) - \mathbb{B}_n(t)\right| = O_p\left[n^{-1/4}\log(n)\right] \tag{7}$$

as $n \to \infty$, by the Skorohod embedding theorem and Levy's modulus of continuity for Brownian motion. See, for example, Breiman [(1992), pages 257 and 293] or Csörgő and Revesz (1981). Then

$$\begin{aligned} G_n(t) - t &= \frac{1}{\alpha_n + \hat{\gamma}_n x_{nn}}\left[F_n^* \circ U_n(t) + \beta_n \delta_1(t)\right] - t \\ &= \frac{1}{\alpha_n + \hat{\gamma}_n x_{nn}} \\ &\quad \times \left[F_n \circ U_n(t) - (\alpha_n + \hat{\gamma}_n x_{nn})t + \frac{1}{\sqrt{n}}\mathbb{F}_n \circ U_n(t) + \frac{c}{\sqrt{n}}\delta_1(t)\right]. \end{aligned} \tag{8}$$

Let

$$H_n(t) := \sqrt{n}[G_n(t) - t], \qquad 0 \le t \le 1,$$

and

$$\mathbb{B}^{\varphi}_{c,n}(t) = \mathbb{B}_n(t) + \Phi(t) - c\bar{\delta}_0(t) + c\delta_1(t), \qquad 0 \le t \le 1$$

for $n \ge 1$, where $\bar{\delta}_0(t) = 0$ for $t \le 0$ and $\bar{\delta}_0(t) = 1$ for $t > 0$. Then the least concave majorants of G_n and H_n are related by $\tilde{H}_n(t) := \sqrt{n}[\tilde{G}_n(t) - t]$ for $0 \le t \le 1$ and $n \ge 1$, since H_n and $\sqrt{n}G_n$ differ by a linear function of t. Now

$$\begin{aligned} U_n(t_{nk}) = x_{nk} &= \alpha_n + \hat{\gamma}_n x_{nk} - \left[\alpha_n + (\hat{\gamma}_n - 1)x_{nk}\right] \\ &= (\alpha_n + \hat{\gamma}_n x_{nn})t_{nk} - \alpha_n - (\hat{\gamma}_n - 1)x_{nk} \\ &\approx (\alpha_n + \hat{\gamma}_n x_{nn})t_{nk} - \alpha_n, \end{aligned} \tag{9}$$

for $1 \le k \le n$. Using (7), (8) and (9), it may be shown that $H_n(t) \approx \mathbb{B}^{\varphi}_{c,n}(t)$ and, therefore, that $\tilde{H}_n(t) \approx \tilde{\mathbb{B}}^{\varphi}_{c,n}(t)$. Denoting derivatives by lower case letters, it then follows that $\sqrt{n}[\tilde{g}_n(t) - 1] = \tilde{h}_n(t) \approx \tilde{b}^{\varphi}_{c,n}(t)$. Combining these observations with a Taylor series expansion of $\tilde{g}(t)\log[\tilde{g}(t)]$ about $\tilde{g}(t) = 1$ suggests that

$$\Lambda_n \Rightarrow \tfrac{1}{2}\int_0^1 \tilde{b}^{\varphi}_c(t)^2\,dt + W, \tag{10}$$

where $\mathbb{B}^{\varphi}_c(t) = \mathbb{B}(t) + \Phi(t) - c\bar{\delta}_0(t) + c\delta_1(t)$, $0 \le t \le 1$, $\mathbb{B}$ denotes a standard Brownian bridge, W denotes a standard exponential random variable which is independent of $\mathbb{B}^{\varphi}_c$, and $\Rightarrow$ denotes convergence in distribution. Of course, the null limiting distribution is a special case with $\varphi = 0$. See the Appendix for a proof of (10).

Similarly, we can show that the limiting distribution of D_n, the D-test statistic in (4), is

$$D_n = \sqrt{n} \sup_{t \in (0,1)} [\hat{F}_n(t) - t] \Rightarrow \sup_{t \in (0,1)} \tilde{\mathbb{B}}^{\varphi}_c(t). \tag{11}$$

The proof for $c > 0$ is essentially the same as that for $c = 0$, and the latter is well known.

REMARK 1. It is easy to see that in the *nondecreasing* case, all the results for $\hat{f}$ above and P- and D- tests below are valid after replacing x_i by $x'_i = 1 - x_i$ for $i = 1, \dots, n$.

REMARK 2. There is an alternative formulation of the problem with a different limiting distribution of Λ_n. If the penalty term $n\beta \log f(x_n)$ is replaced by $n\beta \log f(1)$ in the equation (2), and if $\alpha + \hat{\gamma} x_n$ is replaced by $\alpha + \hat{\gamma}$ in the definitions of t_k and G, then

$$\Lambda_n \Rightarrow \tfrac{1}{2}\int_0^1 \tilde{b}^{\varphi}_c(t)^2\,dt.$$

We have chosen to analyze form (2) because it is slightly more challenging and the convergence appeared to be faster in our simulation studies.

REMARK 3. Under the null hypothesis ($\varphi = 0$), $\tilde{\mathbb{B}}_c^{\varphi}$ and $\mathbb{B}_c^{\varphi}$ are denoted by $\tilde{\mathbb{B}}_c$ and $\mathbb{B}_c$. Using the relationship $\sup_{0\le t\le 1}\tilde{\mathbb{B}}_c(t) = \sup_{0\le t\le 1}\mathbb{B}_c(t)$, and the known boundary crossing probability of a Brownian bridge [cf. Breiman (1992), page 290], it is easily seen that

$$P\Big\{\sup_{0\le t\le 1}\tilde{\mathbb{B}}_c(t) > \lambda\Big\} = \exp(-2(c+\lambda)^2)$$

for all $\lambda > 0$. This provides the limiting null distribution of D_n in (11).

We do not know the analytical from for the asymptotic null distribution of Λ_n. Results of Groeneboom (1983) may be relevant here. He gives an analytic form for the density of $\tilde{b}_0^0(t)$ for fixed t.

REMARK 4. In principle, the power of the P- and D-tests against local alternatives is determined by (10) and (11) (but see Remark 3 above). Both tests are consistent against any fixed alternative. In fact, The D-tests are consistent against any alternative distribution function F that is stochastically smaller than the uniform [that is, $F(x) \ge x$ for $0 \le x \le 1$ with strict inequality for some x], and the P tests are consistent against any continuous distribution function that is stochastically smaller than the uniform. This is well known for the KS-test, easily seen for the D-test with $c > 0$. For the P-tests, consistency may be established by showing that $\hat{f}_n(x) \to \tilde{f}(x)$ w.p.1 for $0 \le x < 1$, where $\tilde{f}$ is the derivative of the least concave majorant of F. (The latter may be established as in Theorem 2 of WS.) It then follows that $\liminf_{n\to\infty}\Lambda_n \ge \int_0^1 \tilde{f}(x)\log[\tilde{f}(x)]\,dx$, which is positive if F is stochastically smaller than the uniform, and this implies consistency.

Critical values. As noted above, the limiting null distributions in (10) do not admit simple analytical descriptions (to the best of the authors' knowledge). The null distribution of Λ_n may be approximated by simulations, however. We have done extensive simulations for computing Monte Carlo estimates of the 95 percentiles of the null distributions of Λ_n (P-test statistic) and D_n (D-test statistic) for $n = 20, 30, \dots, 90, 100, 150, \dots, 450, 500, \infty$ and various values of penalty parameter c. The simulation size is 10,000 throughout. Values for $n = \infty$ were computed from the limiting distribution for D_n and by simulating Brownian motion on a grid of width 0.001 for the P-test. The 95 percentiles of D_n *stabilize at around* $n = 200$, considering their estimated standard deviations. The percentiles of the P-statistic Λ_n stabilize more slowly, especially for small c. This is to be expected, since there is no limiting distribution for the P-test when $c = 0$. The P-test with $c = 0$ is the unpenalized likelihood ratio test or the W test named by Bain, Englehardt and Wright (1985), and appears to have a logarithmic growth in Λ_n when $c = 0$. This is consistent with a limit theorem by Groeneboom and Pyke (1983). Standard deviations of the estimated percentiles in the range from 0.008 to 0.012 were common for the D-tests, and estimates from 0.036 to 0.065 for the P-tests. We have also examined the means, medians, standard deviations and quartiles, and they all admit similar patterns. Those for the D-test stablize at about $n = 200$, and those for the P test stablize more slowly. The null distributions of P- and D-test statistics are skewed to the right. As expected and shown by the power studies in the next section, the D-test is *relatively* insensitive to the value of c. The P-test with $c = 0.2$ seems to produce the best power. We suggest using $c = 0.2$ for both P- and D-tests and provide level 0.05 critical values for $c = 0.2, 0.25$ and 0.3 in Table 1. Critical values for other n and α, needed in the air-conditioning equipment data described in Section 5, were computed similarly.

5. Power studies and comparisons. Power studies of our two tests and other competing tests and an application are provided in this section. Since the alternative is a very large space, no one test can be universally best. The P- and D-tests performed well, however, for a variety of alternatives in the simulations below.

Biased Sampling. Consider comparisons of the P- and D-tests with the most powerful (MP) tests for two simple alternatives. The first alternative is that $f(x) \sim x$, which corresponds to the popular length-biased sampling with

TABLE 1
Level 0.05 *critical values*

	***D*-tests**			***P*-tests**		
n	***c* = 0.2**	**0.25**	**0.3**	***c* = 0.2**	**0.25**	**0.3**
10	0.960	0.918	0.892	3.44	3.30	3.120
20	0.929	0.882	0.841	3.73	3.55	3.41
30	0.943	0.888	0.840	3.85	3.64	3.51
40	0.945	0.893	0.839	4.03	3.84	3.68
50	0.956	0.903	0.849	4.04	3.83	3.65
60	0.952	0.898	0.847	4.19	3.95	3.76
70	0.968	0.912	0.857	4.20	3.97	3.74
80	0.964	0.910	0.857	4.27	4.00	3.78
90	0.984	0.929	0.878	4.25	4.09	3.76
100	0.979	0.926	0.873	4.29	3.982	3.78
150	0.992	0.937	0.886	4.44	4.16	3.98
200	0.998	0.945	0.893	4.45	4.15	3.93
250	0.991	0.940	0.887	4.49	4.20	3.97
300	0.990	0.940	0.888	4.51	4.22	3.96
350	0.995	0.942	0.889	4.53	4.21	3.94
400	0.999	0.946	0.893	4.52	4.20	3.95
450	0.994	0.942	0.891	4.62	4.28	4.01
500	0.999	0.948	0.895	4.57	4.25	3.99
∞	1.024	0.974	0.924	4.87	4.52	4.23

$w(x) = x$, and the second is that $f(x) \sim e^x$, which corresponds to the power-biased sampling with $w(x) = e^x/e$. These are increasing alternatives rather than H_1. See Remark 1 in Section 3. It is easy to see that the most powerful tests are

$$\phi_Z(x) = 1 \quad \text{if } Z = \frac{1}{n}\sum_{i=1}^{n} \log X_i > c_Z \text{ and } \phi_Z(x) = 0 \text{ otherwise,} \tag{12}$$

$$\phi_L(x) = 1 \quad \text{if } L = \frac{1}{n}\sum_{i=1}^{n} X_i > c_L \text{ and } \phi_L(x) = 0 \text{ otherwise,} \tag{13}$$

for $f(x) \sim x$ and $f(x) \sim e^x$, respectively. Here 1 indicates rejection and 0 nonrejection. Both normalized L and Z statistics have normal limiting distributions. The approximate level α critical values for the L- and Z-tests are

$$c_L = \frac{1}{2} + z_\alpha \frac{1}{\sqrt{12n}} \quad \text{and} \quad c_Z = -1 + z_\alpha \frac{1}{\sqrt{n}}, \tag{14}$$

where z_α is the upper α quantile of the standard normal distribution. In fact, the exact critical value of the Z-test can be read off from a Gamma table or a χ^2 table after a rescaling. In our power studies, we use simulated critical values from the same random samples as those for P- and D-tests.

Figures 1 and 2 present power comparisons of the L- and Z-tests, D-tests and P-tests with $c = 0$, 0.2, 0.25, 0.3, 0.4 (total 12 tests) for selected values of n, based on the critical values from Table 1. Clearly, power increases as n increases for all 12 tests. Among P-tests, $c = 0.2$ appears to be best, and among D-tests, there is not much difference as c varies, for most values of n. This relative insensitivity of the D-test to the value of c is expected, since the main effect of the penalty terms is only on the two end points of the density estimate. Note that P-test with $c = 0.2$ is better than the D-tests when $w(x) = x$, while the D-test is better than the P-test, with $c = 0.2$, when $w(x) = e^x/e$, especially for smaller sample size n.

Overall, Figures 1 and 2 show that the P- and D-tests (with $c = 0.2$) are competitive with the Z- and L-tests on the latter's home ground, for length and power alternatives. For other alternatives, the Z- and L-tests are no longer most powerful. One important type of alternative is that a proportion of the population is either missed entirely or included with a much lower probability. Powers of the four tests for such alternatives are presented in Table 2, where

$$w_1(x) = 0 \text{ or } 1 \quad \text{if } 0 \le x < b \text{ or } b \le x \le 1,$$

$$w_2(x) = 0.1 \text{ or } 1 \quad \text{if } 0 \le x < b \text{ or } b \le x \le 1,$$

$$w_3(x) = 0 \text{ or } 1/2 \text{ or } 1 \quad \text{if } 0 \le x < b \text{ or } b \le x < 1/2 \text{ or } 1/2 \le x \le 1.$$

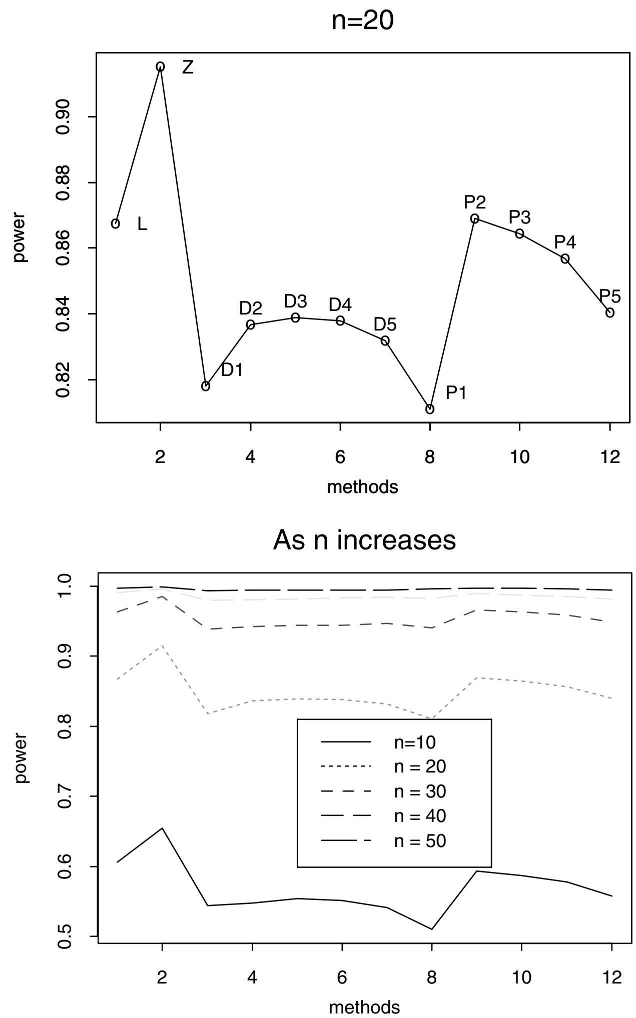

FIG. 1. *Power comparisons, Case* 1: $w(x) = x$. *D*1–*D*5 *(P*1–*P*5*) correspond to D-test (P-test) with* $c = 0$, 0.2, 0.25, 0.3, 0.4, *respectively. The L- and Z-tests are as in* (13) *while Z is MP in this case. Power is* 1 *for all tests for* $n \geq 50$.

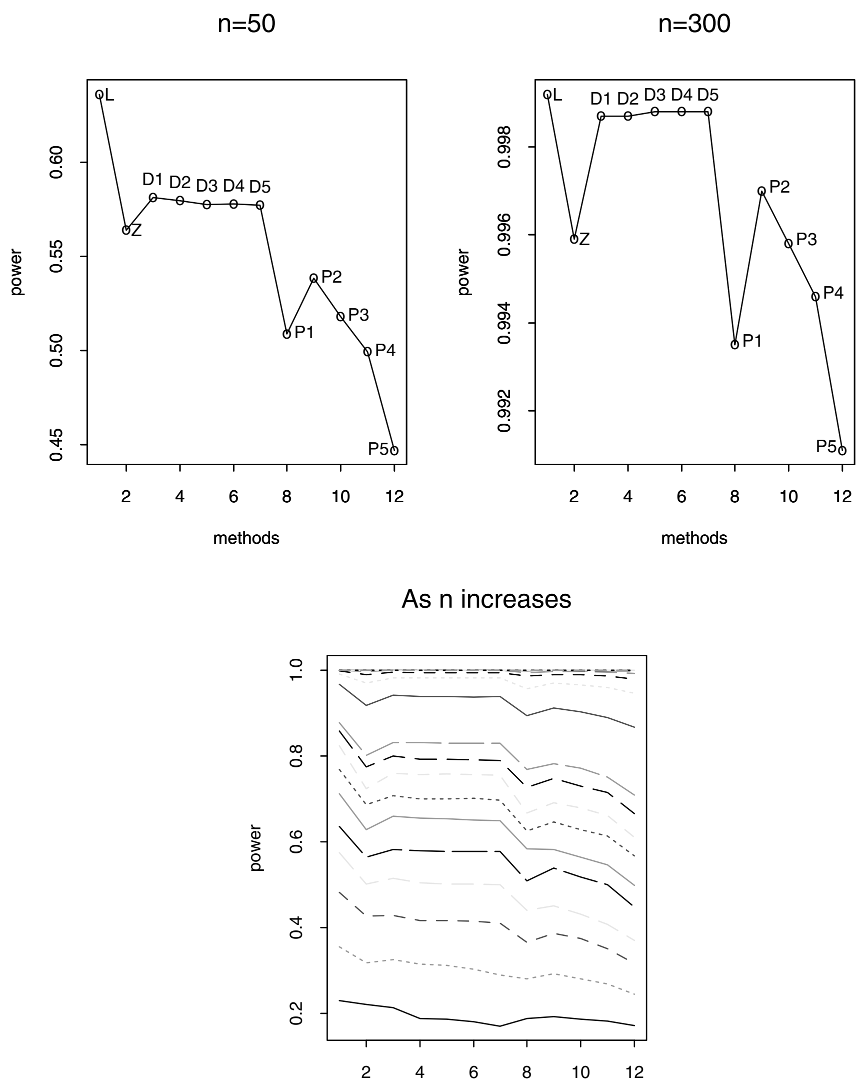

FIG. 2. *Power comparisons, Case* 2: $w(x) = e^x/e$. *The left picture is for* $n = 50$, *the middle is for* $n = 300$, *and the right is for* $n = 20, 30, \ldots, 100, 150, \ldots, 450, 500$. *The L-test is MP in this case.*

TABLE 2
Powers of various tests when $w(x)$ is a step function ($c = 0.2$)

	b = 1/3				b = 0.2				b = 0.1			
n	Z	L	D	P	Z	L	D	P	Z	L	D	P
						$w_1(x)$						
10	0.838	0.612	1.000	1.000	0.368	0.249	0.297	0.383	0.144	0.112	0.114	0.134
20	0.998	0.915	1.000	1.000	0.749	0.448	0.860	1.000	0.261	0.159	0.171	0.315
30	1.000	0.989	1.000	1.000	0.941	0.618	1.000	1.000	0.426	0.219	0.216	0.686
40	1.000	0.999	1.000	1.000	0.988	0.757	1.000	1.000	0.555	0.269	0.287	1.000
50	1.000	1.000	1.000	1.000	0.998	0.824	1.000	1.000	0.660	0.312	0.351	1.000
						$w_2(x)$						
10	0.609	0.464	0.684	0.684	0.302	0.220	0.251	0.313	0.132	0.108	0.104	0.123
20	0.882	0.778	0.914	0.856	0.575	0.362	0.605	0.715	0.233	0.148	0.156	0.273
30	0.968	0.920	0.986	0.965	0.788	0.521	0.718	0.786	0.358	0.187	0.190	0.527
40	0.990	0.977	0.999	0.994	0.892	0.646	0.839	0.878	0.455	0.240	0.238	0.718
50	0.997	0.991	1.000	0.999	0.943	0.716	0.927	0.949	0.546	0.268	0.280	0.737
60	0.999	0.998	1.000	1.000	0.977	0.798	0.979	0.980	0.629	0.306	0.335	0.746
						$w_3(x)$						
10	0.963	0.833	1.000	1.000	0.703	0.558	0.574	0.650	0.436	0.372	0.354	0.369
20	1.000	0.992	1.000	1.000	0.973	0.859	0.963	1.000	0.744	0.613	0.606	0.703
30	1.000	1.000	1.000	1.000	1.000	0.962	1.000	1.000	0.916	0.789	0.764	0.951
40	1.000	1.000	1.000	1.000	1.000	0.993	1.000	1.000	0.974	0.888	0.872	1.000
50	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.992	0.931	0.927	1.000
60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.969	0.967	1.000

Notes: Due to space limit, the results for larger n values are omitted. They all showed that the P-tests appear superior. For example, when the alternative is $w_1(x)$, powers of these tests are 1 (except L-test for $b = 0.2$ and Z-, L,- and D-tests for $b = 0.1$) for $n = 60, \dots, 100$.

The P- and D-tests have higher power than the Z- and L-tests for most of the cases considered, and the P-test has higher power than the D-test except when $b = 1/3$, when all powers are very high.

Nonhomogeneous Poisson process. In this part, we replace X_i by Y_i/T where Y_i and T are failure times and the last observation time described in Section 2. Then the so-called Laplace test and Crow's (1974) test, recommended by Bain, Englehardt and Wright (1985) and Cohen and Sackrowitz (1993), are the L-test and Z-test in (13). They are uniformly most powerful unbiased tests for alternatives that $\lambda(t) = ae^{bt}$ and $\lambda(t) = (\beta/\theta)(t/\theta)^{\beta-1}$, respectively, under the NHPP. The W-test (which happens to be the P-test with $c = 0$) is the third test these two papers suggested. However, as illustrated in Figure 2, the W-test is affected by the spiking problem in the unpenalized maximum likelihood estimator of the density and is dominated by our P-test with $c = 0.2$ (see Figures 1 and 2). Thus we shall mainly compare our P- and D-tests with the L- and Z-tests. Our setup is similar to that of Bain, Englehardt and Wright (1985). The results are presented in

TABLE 3
Estimated powers for testing H_0: $\lambda(t) = constant$ versus various H_1

n	L	Z	D	P	L	Z	D	P
	H_1: $\lambda(t) = be^t$							
	$T = 1$				$T = 2$			
10	0.223	0.214	0.188	0.190	0.545	0.503	0.444	0.457
20	0.346	0.306	0.310	0.279	0.805	0.728	0.731	0.704
40	0.569	0.483	0.503	0.437	0.974	0.933	0.947	0.926
60	0.708	0.619	0.647	0.579	0.997	0.983	0.991	0.986
80	0.823	0.723	0.755	0.688	1.000	0.996	0.998	0.998
100	0.848	0.747	0.838	0.787	1.000	0.998	1.000	0.999
	H_1: $\lambda(t) = bt^{b-1}$							
	$b = 2$				$b = 4$			
10	0.599	0.654	0.542	0.588	0.995	0.992	0.987	0.994
20	0.870	0.917	0.835	0.871	1.000	1.000	1.000	1.000
40	0.992	0.998	0.984	0.992	1.000	1.000	1.000	1.000
	H_1: $\lambda(t)$ is in (15)							
	$T = 4$, $b = 1/3$				$T = 6$, $b = 1/3$			
10	0.191	0.271	0.217	0.272	0.130	0.152	0.128	0.15
20	0.332	0.580	0.479	0.959	0.193	0.319	0.211	0.39
40	0.604	0.936	1.00	1.00	0.319	0.636	0.361	1.0
60	0.764	0.994	1.00	1.00	0.424	0.837	0.642	1.0
	$T = 4$, $b = 1/2$				$T = 6$, $b = 1/2$			
10	0.357	0.539	0.532	0.679	0.193	0.271	0.216	0.273
20	0.649	0.923	1.0	1.0	0.335	0.586	0.483	0.964
40	0.931	1.0	1.0	1.0	0.597	0.933	1.0	1.0
60	0.988	1.0	1.0	1.0	0.756	0.994	1.0	1.0
	$T = 4$, $b = 2/3$				$T = 6$, $b = 2/3$			
10	0.608	0.834	1.0	1.0	0.301	0.440	0.394	0.504
20	0.920	0.999	1.0	1.0	0.540	0.842	1.0	1.0
40	0.999	1.0	1.0	1.0	0.851	0.998	1.0	1.0

Notes: Under the first set of alternatives: $\lambda(t) = be^t$, b cancels out in the sampling density $f(x) = T\lambda(Tx)/\Lambda(T)$; while under the next set of alternatives, T cancels out in the density.

the first two blocks of Table 3, where our figures for L- and Z-tests are consistent with those of Bain, Englehardt and Wright (1985). Our simulation size is 10,000 and the critical values are from 10,000 Monte Carlo simulations rather than normal approximations in (14). The results in these two blocks again show that our P- and D-tests are comparable to the Laplace and Crow tests under the alternative for which the Laplace or Crow's test is optimal.

Next we shall compare our tests with the L- and Z-tests under other monotone alternatives. For simplicity, the following step functions are considered:

$$\lambda(t) = 0 \quad \text{if } 0 \le t \le b, \text{ and } 1 \text{ if } t > b, \tag{15}$$

where b are chosen to be 1/3, 1/2 and 2/3. The results are presented in the last three blocks of Table 3. Clearly, the P-test appears to be best, and the D-test is quite good.

Failure of air-conditioning equipment in 13 Boeing 720 aircraft. Finally, we apply our procedures to the air-conditioning equipment data from Cox and Lewis (1966) and compare our results with Moeller's (1976) analysis. Moeller fitted a Rasch–Weibull Process to the data set. The Rasch–Weibull process is a nonhomogeneous Poisson process with intensity function $\lambda(t) = \lambda\gamma t^{\gamma-1}$. Here $\gamma = 1$ corresponds to a constant intensity function, $\gamma > 1$ indicates that the intensity function is increasing (or the system deteriorates as it ages), and $\gamma < 1$ indicates that the intensity function is decreasing. When an estimate of γ is close to 1, how do we know if it is significantly different from 1? Table 4 lists Moeller's estimates $\hat{\gamma}$ of γ for 12 sets of air-conditioning equipment and our test results at selected significance levels. The original eleventh set of equipment is not considered in this study as there was only one failure time.

It is clear that for the cases that Moeller's $\hat{\gamma}^{-1} > 1$, our tests do not reject H_0. At the significance level 0.05, most tests indicate that those with $\hat{\gamma}^{-1} < 0.65$ are significant. At the significance level 0.3, which is commonly used for quality control problems, most tests conclude that those with $\hat{\gamma}^{-1} < 0.9$ are significant. In summary, our results are consistent with Moeller's, who assumed the Rasch–Weibull process.

Conclusion. We developed two tests for testing uniformity versus a monotone density and *recommend* P-test and D-test with $c = 0.2$. They are quite efficient in comparison to the MP tests developed for special alternatives. The P- and D-tests do not lose much power under the length-biased or power-biased alternatives (for which the Z- or L-test is optimal) and the P-test or D-test (mostly the P-test) is much better than the L- and Z-tests under the step function alternatives. Our tests have greater applicability, as they use no parametric or known form of the alternative density other than saying that the density is monotone. The computer code for computing P- and D-tests and critical values is available upon request.

TABLE 4
Comparison with Moeller's results

Number n	1 5	2 22	3 28	4 14	5 13	6 29	7 26	8 23	9 8	10 5	2 11	13 15
$\hat{\gamma}^{-1}$	**0.699**	**0.601**	**1.15**	**1.148**	**0.823**	**0.687**	**1.051**	**0.959**	**0.882**	**0.823**	**0.549**	**0.826**
Tests	Significance level $\alpha = 0.05$											
L	*	S	*	*	*	S	*	*	*	*	*	*
Z	*	S	*	*	*	S	*	*	*	*	S	*
D	*	S	*	*	*	S	*	*	*	*	S	*
P	*	S	*	*	*	S	*	*	*	*	S	*
	Significance level $\alpha = 0.1$											
L	*	S	*	*	*	S	*	*	*	*	*	*
Z	*	S	*	*	*	S	*	*	*	*	S	*
D	S	S	*	*	*	S	*	*	*	*	S	*
P	*	S	*	*	*	S	*	*	*	*	S	*
	Significance level $\alpha = 0.2$											
L	*	S	*	*	S	S	*	*	*	*	S	*
Z	*	S	*	*	*	S	*	*	*	*	S	*
D	S	S	*	*	*	S	*	*	*	*	S	*
P	S	S	*	*	*	S	*	*	*	*	S	*
	Significance level $\alpha = 0.3$											
L	*	S	*	*	S	S	*	*	*	S	S	*
Z	S	S	*	*	S	S	*	*	*	*	S	S
D	S	S	*	*	*	S	*	*	S	*	S	*
P	S	S	*	*	*	S	*	*	S	*	S	*

Notes: Here n is the number of failure times; $\hat{\gamma}$ is Moeller's estimate under the Rasch–Weibull process; S indicates "significant" (i.e., reject H_0 in favor of H_1: $\lambda(t)$ is nondecreasing), and $*$ indicates "not significant." Equipment 11 is not included as there was only one failure time.

APPENDIX

Proofs.

OUTLINE OF THE PROOF OF PROPOSITION 1. Letting $f_k = e^{\theta_k}$ and introducing a Lagrange multiplier $n\gamma$ leads to consideration of

$$l^*(\theta) = l_{\alpha,\beta}(f) - n\gamma \sum_{i=1}^{n} (x_i - x_{i-1}) f_i.$$

Clearly, $\partial l^*(\theta)/\partial\theta_k = n(c_k - w_k f_k)$ for $k = 1, \dots, n$, where c_k and w_k are as in (1). A necessary and sufficient condition that θ maximize l^* is that

$$\sum_{k=1}^{n} (c_k - w_k f_k)(\xi_k - \theta_k) \le 0$$

for all $-\infty < \xi_n \le \cdots \le \xi_1 < \infty$. By Theorems 1.32 and 1.44 of RWD, the solution to this problem is given by (1) for a given γ. To satisfy the constraint, γ must be so chosen that $\sum_{i=1}^{n}(x_i - x_{i-1}) f_i(\gamma) = 1$. Using the relations $\sum_{i=1}^{n} w_i f_i(\gamma) = \sum_{i=1}^{n} c_i = 1 + \beta$ from Theorem 1.36 of RWD and $\sum_{i=1}^{n} w_i f_i(\gamma) = \gamma \sum_{i=1}^{n} (x_i - x_{i-1}) f_i(\gamma) + \alpha f_1(\gamma)$, the constraint may be written $\gamma = 1 + \beta - \alpha f_1(\gamma)$. The right side is a bounded, nondecreasing, concave function of γ. It vanishes at $\gamma = 0$ and its derivative there is $(1+\beta)x_n/\alpha$. Thus the equation $\gamma = 1 + \beta - \alpha f_1(\gamma)$ has a unique solution $\hat{\gamma}$ if $x_n > \alpha/(1+\beta)$.

When $\alpha = 0 = \beta$, Grenander's estimator is obtained. □

Least concave majorants. Let $\mathscr{H}$ denote the class of all bounded functions H: $[0,1] \to \Re$. Then $\mathscr{H}$ is a Banach space with pointwise linear operations and the supremum norm, $\|H\| = \sup_{0 \le t \le 1} |H(t)|$. For $H \in \mathscr{H}$, let $\tilde{H}$ denote the least concave majorant of H. Then $\|\tilde{H} - \tilde{K}\| \le \|H - K\|$, by Marshall's lemma. The following properties of least concave majorants are needed.

LEMMA 1. *Let $G, H \in \mathscr{H}$, let $B \subseteq [0,1]$ and let $\varepsilon > 0$. If $|G(t) - H(t)| \le \varepsilon$ for all $t \in B$ and $|\tilde{G}(t) - \tilde{H}(t)| \le \varepsilon$ for all $t \in B'$, then $\|\tilde{G} - \tilde{H}\| \le \varepsilon$.*

PROOF. If $t \in B$, then $G(t) \le H(t) + \varepsilon \le \tilde{H}(t) + \varepsilon$, and if $t \in B'$, then $G(t) \le \tilde{G}(t) \le \tilde{H}(t) + \varepsilon$. So, $G(t) \le \tilde{H}(t) + \varepsilon$ for all $0 \le t \le 1$. Since $\tilde{H} + \varepsilon$ is concave, $\tilde{G} \le \tilde{H} + \varepsilon$. The lemma then follows from this observation and its dual, in which the roles of G and H are reversed. □

LEMMA 2. *Let $G, H \in \mathscr{H}$ and suppose that $\tilde{G}(0) = \tilde{H}(0)$ and $\tilde{G}(1) = \tilde{H}(1)$. Let $\tilde{g}$ and $\tilde{h}$ denote the left-hand derivatives of $\tilde{G}$ and $\tilde{H}$. If $-\infty < \tilde{g}(1) \le \tilde{g}(0+) < \infty$, and $-\infty < \tilde{h}(1) \le \tilde{h}(0+) < \infty$, then*

$$\int_0^1 [\tilde{g}(t) - \tilde{h}(t)]^2 \, dt \le \|\tilde{G} - \tilde{H}\| \big[\tilde{g}(0+) - \tilde{g}(1) + \tilde{h}(0+) - \tilde{h}(1)\big].$$

PROOF. Integrating by parts from ε to $1-\varepsilon$ and then letting $\varepsilon \to 0$, yields

$$\int_0^1 [\tilde{g}(t) - \tilde{h}(t)]^2\,dt = \left| \int_0^1 [\tilde{G}(t) - \tilde{H}(t)][d\tilde{g}(t) - d\tilde{h}(t)] \right|$$
$$\leq \|\tilde{G} - \tilde{H}\| [\tilde{g}(0+) - \tilde{g}(1) + \tilde{h}(0+) - \tilde{h}(1)]$$

as asserted. □

LEMMA 3. *If H is upper semicontinuous at* 0, *respectively* 1, *then* $\tilde{H}(0) = H(0)$, *respectively* $\tilde{H}(1) = H(1)$.

The easy proof is left to the reader.

PROOF OF (10). The proof of (10) is described next.

LEMMA 4.

$$\sup_{t_{n1} \leq t \leq 1} \left| U_n(t) - (\alpha_n + \hat{\gamma}_n x_{nn})t + \alpha_n \right| = o_p\left(\frac{1}{\sqrt{n}}\right)$$

and

$$\sup_{0 \leq t \leq 1} |U_n(t) - t| = O_p\left(\frac{1}{\sqrt{n}}\right).$$

PROOF. The first assertion follows easily from (9) and the fact that $U_n(t) - (\alpha_n + \hat{\gamma}_n x_{nn})t + \alpha_n$ is piecewise linear on $[t_{n1}, 1]$. The second assertion follows similarly since $\alpha_n + t_{n1} = O_p(1/\sqrt{n})$ in probability. □

In the next lemma, let $\|G\|_\varepsilon = \sup_{\varepsilon \leq t \leq 1} |G(t)|$ for $G \in \mathscr{H}$ and $0 < \varepsilon < 1$.

LEMMA 5.

$$\lim_{n\to\infty} \left\| H_n - \mathbb{B}^{\varphi}_{c,n} \right\|_{t_{n1}} = 0$$

in probability.

PROOF. From (8),

$$H_n(t) - \mathbb{B}^{\varphi}_{c,n}(t) = \frac{\sqrt{n}}{\alpha_n + \hat{\gamma}_n x_{nn}} \left\{ U_n(t) - (\alpha_n + \hat{\gamma}_n x_{nn})t + \frac{1}{\sqrt{n}} \Phi \circ U_n(t) + \frac{1}{\sqrt{n}} \mathbb{F}_n \circ U_n(t) + \beta_n \delta_1(t) \right\}$$
$$- [\mathbb{B}_n(t) + \Phi(t) - c + c\delta_1(t)]$$
$$= \frac{\sqrt{n}}{\alpha_n + \hat{\gamma}_n x_{nn}} [U_n(t) - (\alpha_n + \hat{\gamma}_n x_{nn})t + \alpha_n]$$

$$+\frac{1}{\alpha_n+\hat{\gamma}_n x_{nn}}\big[\Phi\circ U_n(t)-\Phi(t)\big]$$
$$+\frac{1}{\alpha_n+\hat{\gamma}_n x_{nn}}\big[\mathbb{F}_n\circ U_n(t)-\mathbb{B}_n\circ U_n(t)\big]$$
$$+\frac{1}{\alpha_n+\hat{\gamma}_n x_{nn}}\big[\mathbb{B}_n\circ U_n(t)-\mathbb{B}_n(t)\big]$$
$$+\left(\frac{1}{\alpha_n+\hat{\gamma}_n x_{nn}}-1\right)\big[\mathbb{B}_n(t)+\Phi(t)-c+c\delta_1(t)\big]$$

for $t_{n1}\le t\le 1$. So,

$$\begin{aligned}\|H_n-\mathbb{B}^{\varphi}_{c,n}\|_{t_{n1}} &\le \sup_{t_{n1}\le t\le 1}\frac{\sqrt{n}}{\alpha_n+\hat{\gamma}_n x_{nn}}\big|U_n(t)-(\alpha_n+\hat{\gamma}_n x_{nn})t+\alpha_n\big|\\ &\quad+\frac{1}{\alpha_n+\hat{\gamma}_n x_{nn}}\big(\|\Phi\circ U_n-\Phi\|+\|\mathbb{F}_n-\mathbb{B}_n\|+\|\mathbb{B}_n\circ U_n-\mathbb{B}_n\|\big)\\ &\quad+\left|\frac{1}{\alpha_n+\hat{\gamma}_n x_{nn}}-1\right|\big(\|B_n\|+\|\Phi\|+2c\big)\\ &=o_p(1)\end{aligned}$$

as $n\to\infty$, by Lemma 4 and (7). □

LEMMA 6. *For any Brownian bridge* $\mathbb{B}$, $-\infty<\tilde{b}^{\varphi}_c(1+)\le\tilde{b}^{\varphi}_c(0+)<\infty$ *w.p.*1, *and*

$$\lim_{n\to\infty}\big[|\tilde{h}_n(0+)-\tilde{b}^{\varphi}_{c,n}(0+)|+|\tilde{h}_n(1)-\tilde{b}^{\varphi}_{c,n}(1)|\big]=0$$

in probability.

PROOF. First, it is clear that

$$0\le\tilde{b}^{\varphi}_c(0+)=\sup_{0<t<1}\frac{\mathbb{B}(t)+\Phi(t)-c}{t}\vee 0<\infty$$

w.p.1 for any Brownian bridge $\mathbb{B}$. Next, let $\varepsilon_n>0$, $n\ge 1$, be a sequence for which $\varepsilon_n\to 0$ and $\sqrt{n}\varepsilon_n\to\infty$ as $n\to\infty$. Then $\hat{f}_n(0+)=\sup_{t\ge\varepsilon_n}F^*_n(t)/(\alpha_n+\hat{\gamma}_n t)$ with probability approaching 1, since $\hat{f}_n(0+)\ge 1$ w.p.1 and

$$\begin{aligned}P\left\{\sup_{t\le\varepsilon_n}\frac{F^*_n(t)}{\alpha_n+\hat{\gamma}_n t}\ge 1\right\} &\le P\left\{\sup_{t\le\varepsilon_n}\sqrt{n}[F^*_n(t)-t]\ge c-\sqrt{n}|\hat{\gamma}_n-1|\right\}\\ &\to 0\end{aligned}$$

as $n\to\infty$, by elementary properties of empirical processes. Observe that $\varepsilon_n>t_{n1}$ with probability approaching 1, since $\sqrt{n}\varepsilon_n\to\infty$ as $n\to\infty$. So, with probability approaching 1,

$$\big|\tilde{h}_n(0+)-\tilde{b}^{\varphi}_{c,n}(0+)\big|\le\sup_{t\ge\varepsilon_n}\left|\frac{H_n(t)-\mathbb{B}^{\varphi}_{c,n}(t)}{t}\right|\le\varepsilon_n^{-1}\sup_{t\ge t_{n1}}\big|H_n(t)-\mathbb{B}^{\varphi}_{c,n}(t)\big|;$$

and the right side approaches zero in probability if $\varepsilon_n \to 0$ sufficiently slowly as $n \to \infty$. This establishes the two assertions made about the left endpoint, and the right endpoint may be handled similarly. □

LEMMA 7.

$$\|\tilde{H}_n - \tilde{\mathbb{B}}^{\varphi}_{c,n}\| \to 0$$

in probability as $n \to \infty$.

PROOF. By Lemmas 1 and 5, it suffices to show that $\sup_{0 \le t \le t_{n1}} |\tilde{H}_n(t) - \tilde{\mathbb{B}}^{\varphi}_{c,n}(t)| \to 0$ in probability. To see this, first observe that $\tilde{H}_n(0) = 0 = \tilde{\mathbb{B}}^{\varphi}_{c,n}(0)$, since $H_n(0) = 0 = \mathbb{B}^{\varphi}_{c,n}(t)$ and both processes are upper semicontinuous at 0. So,

$$|\tilde{H}_n(t)| \le [\tilde{h}_n(0) - \tilde{h}_n(1)]t_{n1},$$

$$|\tilde{B}^{\varphi}_{c,n}(t)| \le [\tilde{b}^{\varphi}_{c,n}(0) - \tilde{b}^{\varphi}_{c,n}(1)]t_{n1},$$

for $0 \le t \le t_{n1}$, and the right sides approach 0 in probability as $n \to \infty$. □

Recall

$$\Lambda_n = (\alpha_n + \hat{\gamma}_n x_{nn}) n \int_0^1 \log[\tilde{g}_n(t)]\, d\tilde{G}_n(t) - n\alpha_n [\hat{f}_n(x_{n1}) - 1].$$

THEOREM 1.

$$\lim_{n\to\infty} \left| \Lambda_n - \tfrac{1}{2}\int_0^1 \tilde{b}^{\varphi}_{c,n}(t)^2\, dt - n(1 - x_{nn}) \right| = 0$$

in probability.

PROOF. Let $0 < \varepsilon_n \to 0$ as $n \to \infty$ in such a manner that $\sqrt{n}\varepsilon_n \to \infty$, and let A_n be the event $A_n = \{1 - \varepsilon_n \le \tilde{g}_n(1) \le \tilde{g}_n(0+) \le 1 + \varepsilon_n\}$ for each $n \ge 1$. Then $P(A_n) \to 1$ as $n \to \infty$ and each A_n implies

$$\begin{aligned} 2[\tilde{g}_n(t) - 1] + \frac{1}{(1+\varepsilon_n)}[\tilde{g}_n(t) - 1]^2 &\le 2\tilde{g}_n(t)\log[\tilde{g}_n(t)] \\ &\le 2[\tilde{g}_n(t) - 1] + \frac{1}{(1-\varepsilon_n)}[\tilde{g}_n(t) - 1]^2 \end{aligned}$$

for all $0 \le t \le 1$. So, A_n implies

$$\begin{aligned} \Lambda_n \le{}& (\alpha_n + \hat{\gamma}_n x_{nn}) n \int_0^1 [\tilde{g}_n(t) - 1]\, dt - c\sqrt{n}[\hat{f}_n(x_{n1}) - 1] \\ &+ \frac{\alpha_n + \hat{\gamma}_n x_{nn}}{2(1-\varepsilon_n)} n \int_0^1 [\tilde{g}_n(t) - 1]^2\, dt \end{aligned} \tag{A.1}$$

and a similar lower bound. To estimate the two integrals on the right, observe that

$$\int_0^1 [\tilde{g}_n(t) - 1]\,dt = \frac{1+\beta_n}{\alpha_n + \hat{\gamma}_n x_{nn}} - 1 \tag{A.2}$$

$$= \frac{1 - \hat{\gamma}_n x_{nn}}{\alpha_n + \hat{\gamma}_n x_{nn}}$$

$$= \frac{\alpha_n[\hat{f}_n(x_{n1}) - 1]}{\alpha_n + \hat{\gamma}_n x_{nn}} + \frac{\hat{\gamma}_n(1 - x_{nn})}{\alpha_n + \hat{\gamma}_n x_{nn}} \tag{A.3}$$

and

$$\begin{aligned} &\int_0^1 \big[\tilde{h}_n(t) - \tilde{b}^{\varphi}_{c,n}(t)\big]^2\,dt \\ &\geq \big|H_n - \mathbb{B}^{\varphi}_{c,n}\big|\big[\tilde{h}_n(0+) - \tilde{h}_n(1) + \tilde{b}^{\varphi}_{c,n}(0+) - \tilde{b}^{\varphi}_{c,n}(1)\big] \\ &\to 0 \end{aligned} \tag{A.4}$$

in probability as $n \to \infty$, by Lemmas 2, 5 and 6. So, combining (A.1)–(A.4), A_n implies

$$\Lambda_n \leq \hat{\gamma}_n n(1 - x_{nn}) + \frac{\alpha_n + \hat{\gamma}_n x_{nn}}{2(1-\varepsilon_n)} \int_0^1 \tilde{h}_n(t)^2\,dt$$

and a similar lower bound. It follows that with probability approaching 1,

$$\begin{aligned} \left|\Lambda_n - \frac{1}{2}\int_0^1 \tilde{b}^{\varphi}_{c,n}(t)^2\,dt - n(1-x_{nn})\right| &\leq |\hat{\gamma}_n - 1| n(1 - x_{nn}) \\ &\quad + \left|\frac{1+\varepsilon_n}{1-\varepsilon_n}\int_0^1 \tilde{h}_n(t)^2\,dt - \int_0^1 \tilde{b}^{\varphi}_{c,n}(t)^2\,dt\right|, \end{aligned}$$

which approaches zero as $n \to \infty$ by Lemmas 2 and 7. □

For equation (10), it remains to show that $n(1 - x_{nn})$ has a limiting exponential distribution and is asymptotically independent of $\mathbb{B}^{\varphi}_{n,c}$. The first part is clear. The asymptotic independence is implied by asymptotic independence of F^*_n and x_{nn}, which may be shown easily by conditioning on x_{nn}. See a related calculation for a more complicated situation in McCormick and Sun (1993).

Acknowledgment. Thanks to Bob Keener for helpful comments on Example 1.

REFERENCES

ASCHER, H. E. and FEINGOLD (1984). *Repairable Systems Reliability*. Dekker, New York.

BAIN, L. J., ENGELHARDT, M. and WRIGHT, F. T. (1985). Tests for an increasing trend in the intensity of a Poisson process: a power study. *J. Amer. Statist. Assoc.* **80** 419–422.

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

BICKEL, P., KLAASSEN, C., RITOV, Y. and WELLNER, J. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press.

BOSWELL, M. T. (1966). Estimating and testing trend in a stochastic process of Poisson type. *Ann. Math. Statist.* **37** 1564–1573.

BREIMAN, L. (1992). *Probability*. SIAM, Philadelphia.

COHEN, A., PERLMAN, M. D. and SACKROWITZ, H. B. (1990). Unbiasedness of tests of homogeneity when alternatives are ordered. In *Topics in Statistical Dependence* IMS, Hayward, CA. 135–145.

COHEN, A. and SACKROWITZ, H. B. (1993). Evaluating tests for increasing intensity of a Poisson process. *Technometrics* **35** 446–448.

COX, D. R. and LEWIS, P. A. W. (1966). *The Statistical Analysis of Series of Events.* Methuen, London.

CROW, L. H. (1974). Reliability and analysis for complex repairable systems. In *Reliability and Biometry* (F. Proschan and R. J. Serfling, eds.) 379–410. SIAM, Philadelphia, PA.

CSÖRGŐ, M. and REVESZ, P. (1981). *Strong Approximation in Probability and Statistics.* Academic Press, New York.

GRENANDER, X. (1956). On the theory of mortality measurement. Part II. *Skand. Akt.* **39** 123–153.

GROENEBOOM, P. (1983). The concave majorant of Brownian motion. *Ann. Probab.* **11** 1016–1027.

GROENEBOOM, P. and PYKE, R. (1983). Asymptotic normality of statistics based on the convex minorants of empirical distribution functions. *Ann. Probab.* **11** 328–345.

HARTIGAN, J. A. and HARTIGAN, P. M. (1985). The Dip test of unimodality. *Ann. Statist.* **13** 70–84.

MCCORMICK, W. P. and SUN, J. (1993). Sums and maxima of discrete stationary processes. *J. Appl. Probab.* **30** 863–876. Correction: [*J. Appl. Probab.* **31** 867 (1994).]

MISRA, P. N. (1983). Software reliability analysis. *IBM Syst. J.* **22** 262–270.

MOELLER, S. K. (1976). The Rasch–Weibull process. *Scand. J. Statist.* **3** 107–115.

ROBERTSON, T., WRIGHT, F. and DYKSTRA R. (1988). *Order Restricted Inference*. Wiley, New York.

SUN, J. and WOODROOFE, M. (1997). Semi-parametric estimation for biased sampling models. *Statist. Sinica* **7** 545–575.

WANG, Y. (1994). A Bartlett-type adjustment for the likelihood ratio statistic with an ordered alternative. *Statist. Probab. Lett.* **20** 347–352.

WOODROOFE, M. and SUN, J. (1993). A penalized likelihood estimate of $f(0+)$ when f is nonincreasing. *Statist. Sinica* **3** 501–515.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109-1027
E-MAIL: michaelw@umich.edu.

DEPARTMENT OF STATISTICS
CASE WESTERN RESERVE UNIVERSITY
CLEVELAND, OHIO 44106
E-MAIL: jiayang@sun.cwru.edu.