The Annals of Statistics
2001, Vol. 29, No. 5, 1297–1319

MULTISCALE MAXIMUM LIKELIHOOD ANALYSIS OF A SEMIPARAMETRIC MODEL, WITH APPLICATIONS[1]

BY GUENTHER WALTHER

Stanford University

A special semiparametric model for a univariate density is introduced that allows analyzing a number of problems via appropriate transformations. Two problems treated in some detail are testing for the presence of a mixture and detecting a wear-out trend in a failure rate. The analysis of the semiparametric model leads to an approach that advances the maximum likelihood theory of the Grenander estimator to a multiscale analysis. The construction of the corresponding test statistic rests on an extension of a result on a two-sided Brownian motion with quadratic drift to the simultaneous control of "excursions under parabolas" at various scales of a Brownian bridge. The resulting test is shown to be asymptotically optimal in the minimax sense regarding both rate and constant, and adaptive with respect to the unknown parameter in the semiparametric model. The performance of the method is illustrated with a simulation study for the failure rate problem and with data from a flow cytometry experiment for the mixture analysis.

1. Introduction and overview. This paper is concerned with the semiparametric model

$$f(x) = \exp(\phi(x) + cx), \tag{1}$$

where f is a probability density with support in [0, 1], ϕ is a nonincreasing function taking values in $\overline{\mathbf{R}}$, and $c \geq 0$ is a real parameter. For uniqueness we will always take c to be the smallest value possible in the above representation. Given n iid observations $X_1, \dots, X_n$ from f, the problem is to test whether $c = 0$, that is, f is nonincreasing, while under the alternative $c > 0$, f is allowed to possess local stretches of exponential growth with unknown parameter c.

Besides the direct application to testing whether a density is monotone [see, e.g., Hildenbrand and Hildenbrand (1985)], a number of important problems from different areas in statistics can be reduced to the semiparametric model (1) by appropriate transformations. Two problems from reliability theory and the analysis of mixtures will be addressed in more detail in Section 5.

The shape of the failure rate function plays an important role in reliability theory. An increase in the failure rate marks a wear-out trend and can be used as a signal for preventive maintenance or replacement. The standard approach for testing a constant vs. a monotone failure rate is based on the cumulative total time on test statistic [see Robertson, Wright and Dykstra

Received September 1999; revised April 2001.
[1]Supported in part by NSF Grants DMS-97-04557 and DMS-98-75598.
AMS 2000 *subject classifications*. Primary 62G10, secondary 60F15.
Key words and phrases. Multiscale analysis, Grenander estimator, minimax, adaptive, mixture, log-concave, failure rate, penalized maximum likelihood.

(1988), Chapter 7.6], which looks for a global monotone trend in the normalized spacings. It exploits the fact that the latter are iid if the failure rate is constant (i.e., the distribution is exponential), whereas the sequence of normalized spacings is stochastically decreasing (increasing) when the failure rate is increasing (decreasing).

It is generally recognized, however, that many failure rates in practice are bathtub shaped, that is, are first decreasing (during a "burn-in" period), then are roughly constant, and are finally increasing (during the "wear-out" period) [see, e.g., Glaser (1980) or Miller (1981), page 15]. Statistical procedures looking for a global trend in the normalized spacings are clearly not well suited for detecting a wear-out (or burn-in) period for such failure rates, as the locally increasing and decreasing trends can cancel in the statistic. Rigorous statistical theory and procedures to detect these important alternatives seem not to have been developed yet, apparently due to the difficulties in modeling bathtub shaped failure rates parametrically (personal communication with Ingram Olkin). Section 5 will show how the theory to be developed for model (1) applies directly to this problem via the total time on test transformation.

The second application concerns detecting the presence of mixing in a distribution. That is, one wishes to decide whether a given sample is composed of observations from one population or from multiple subpopulations. The statistical theory has been developed with remarkable success in the case where the component distributions are from a one-parameter exponential family or from the two-parameter normal family [see, e.g., Lindsay and Roeder (1992, 1997) and Roeder (1994)]. There is also a considerable interest in a nonparametric approach to this problem, as the conclusions of a parametric approach can depend quite sensitively on the assumed model and skewed distributions in particular cause problems [see Roeder (1994), page 493]. However, the standard nonparametric approach to this problem is a test for unimodality, which is known not to be very sensitive to detect the presence of a subpopulation [see, e.g., Roeder (1994), page 493, and Titterington, Smith and Makov (1985)] for a judicious discussion on the use of modality in this context.

The approach taken here is based on a nonparametric model that is commonly employed in the MCMC and Gibbs sampling literature; see Gilks and Wild (1992), Dellaportas and Smith (1993) and Brooks (1998). It models single-component distributions as logarithmically concave densities, that is, densities of the form $f(x) = e^{\psi(x)}$, where ψ is a concave function. As detailed in the above references, this model is motivated by the fact that most commonly used parametric densities are log-concave, the prime example of course being the standard normal density where ψ is a quadratic. While this model is thus a quite natural choice for the mixture problem at hand, the requisite statistical methodology has not been developed yet. The following lemma shows how this approach can be subsumed under the general semiparametric model (1).

LEMMA 1. *Let F be the cdf of a univariate random variable X whose distribution is a finite mixture of log-concave distributions with common support. Then for all $d > 0$ the distribution of the random variable $F(X - d)$ is* *absolutely continuous on* (0, 1) (*and may have an atom at* 0). *The distribution of X is log-concave if and only if for all $d > 0$ the Radon–Nikodym derivative f_d of $F(X - d)$ is nonincreasing on* (0, 1). *Moreover, on every closed interval $I \subset (0, 1)$, f_d is of the form* (1).

The lemma is illustrated in Figure 1: forty observations $\{X_i\}_{i=1}^{40}$ were sampled from a gamma(10, 0.1) distribution. The shifted observations $\{X_i - d\}_{i=1}^{40}$ with $d = 0.2$ are plotted on the horizontal axis of the left figure. The solid line is a graph of the smoothed empirical cdf $\widetilde{F}_n$ of the X_i [see Shorack and Wellner (1986), page 86]. The transformation $(X_i - d) \mapsto \widetilde{F}_n(X_i - d)$ is delineated by dotted lines, and the transformed observations are plotted on the vertical axis. One clearly notes the decreasing frequency of the transformed observations in (0, 1), corresponding to the nonincreasing f_d. On the other hand, the solid line in the right figure shows the density of the mixture $\frac{1}{2}$gamma(2, 0.1) + $\frac{1}{2}$gamma(5, 0.15). For easier visualization f_d is also plotted in the figure, rather than a transformed sample. f_d is of the form (1), and the local increase indicates the presence of a mixture.

Another problem that can be subsumed under (1) is that of detecting a local trend in the intensity function of a nonhomogeneous Poisson process. See Woodroofe and Sun (1999) for the link and for an approach to detect a global trend.

The problem of testing whether a density is nonincreasing or related versions have been considered by Chaudhuri and Marron (1999), Dümbgen and Spokoiny (2000), who deal with the Gaussian white noise model and also give optimality results in that setting, and by Woodroofe and Sun (1999), who test uniformity versus a monotone density. The first two papers investigate shape properties of f with kernel estimates by simultaneously considering a range of bandwidths.

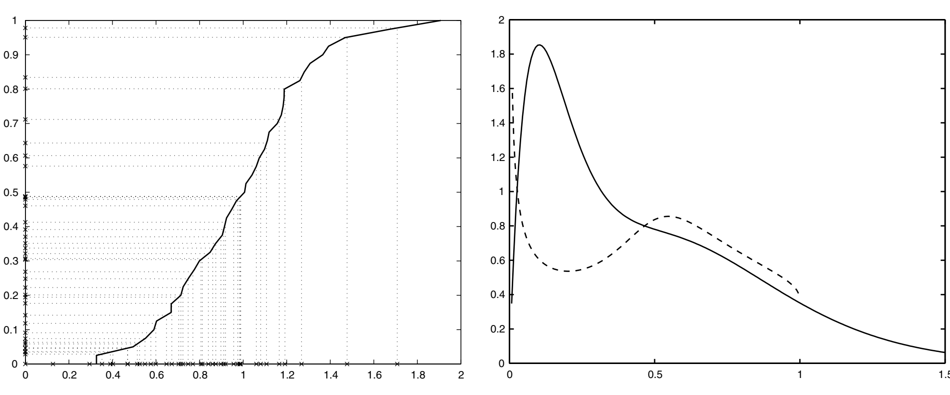

FIG. 1. *Left*: $\{X_i - d\}_{i=1}^{40}$ *are plotted on the horizontal axis, where* $X_i \sim \gamma(10, 0.1), d = 0.2$. $\{\widetilde{F}_n(X_i - d)\}_{i=1}^{40}$ *are plotted on the vertical axis. Right*: $f = \frac{1}{2}\gamma(2, 0.1) + \frac{1}{2}\gamma(5, 0.15)$ (*solid line*) *and Radon–Nikodym derivative of $F(X - d)$, where $X \sim f$* (*dashed line*).

For the analysis of properties of the logarithm of f, as required by model (1) and its extension (2), it is natural to use the method of maximum likelihood, which has an extensive history in order restricted inference, see Robertson, Wright and Dykstra (1988). Note that the MLE of c does not exist. The program is to compute the MLE of f for various fixed values of the unknown parameter c, and then to evaluate the evidence for $c > 0$ by combining the evidence obtained from the various MLEs. It will be shown how this approach advances the method of maximum likelihood to a multiscale analysis of the model (1) that enjoys certain adaptivity and optimality properties.

The MLE of f for the case $c = 0$ is the well-known Grenander estimate [see, e.g., Groeneboom (1985)]. We will call the model (1) when $c = 0$ the "null family," following terminology introduced by Silverman (1982) in a related situation. It will be seen that for a given $c > 0$ the MLE can be computed by solving a penalized ML problem, or equivalently, by transforming the data to a different scale, applying the estimator for the null family on that scale, and transforming the resulting estimator back to the original scale. The parameter c, or equivalently the Lagrange (tuning) parameter in the penalized ML problem, can thus be interpreted as a "scale" parameter that provides information about f on various scales. This approach avoids the usual problem of appropriately choosing the tuning parameter in a penalized ML problem. Rather, the analysis combines the information obtained on the various scales.

Deriving an optimal simultaneous testing procedure requires knowledge of the simultaneous null distribution of the Grenander estimators across scales. The pointwise limiting distribution of the Grenander estimator is related to an "excursion below a parabola" of a Brownian bridge and can be described in terms of the arg max of a two-sided Brownian motion with quadratic drift [see Prakasa Rao (1969) and Groeneboom (1985)], whose distribution is related to the solution of a heat equation [see Chernoff (1964) and Groeneboom (1989)]. To construct an appropriate test, this result is generalized by simultaneously considering such excursions at various scales of the Brownian bridge as well as across locations.

Section 2 shows how the analysis of the semiparametric model (1) gives rise to the multiscale procedure sketched above. In Section 3 the simultaneous behavior of "excursions below parabolas" of a Brownian bridge across scales and locations is derived, which allows constructing an appropriate test statistic. In Section 4 it is proved that the resulting procedure is adaptive with respect to the unknown parameter c and asymptotically minimax. In Section 5 the procedure is applied to two problems in reliability theory and the analysis of mixtures. Section 6 contains a brief outlook on further work. The proofs are deferred to Section 7.

2. The multiscale MLE.

The plan for testing

$$H_0\colon c = 0$$

in the model (1) is to compute the MLE of f for various values of the unknown parameter c, and then to extract and combine the evidence for $c > 0$ from the various estimates. We have already noted that in the case $c = 0$ the MLE of f is the Grenander estimator based on the observations X_i; that is, the left-hand slope of the least concave majorant of the empirical cdf of the X_i [see, e.g., Robertson, Wright and Dykstra (1988)]. A similar representation obtains for the case $c > 0$.

PROPOSITION 1. *Fix $c > 0$ in* (1). *Then the MLE $\hat{f}_n^c$ of f is given by*

$$\hat{f}_n^c(x) = \hat{g}_n^c(e^{cx})ce^{cx},$$

where $\hat{g}_n^c$ is the Grenander estimator based on the transformed observations $Y_i = e^{cX_i}$, $i = 1, \dots, n$.

The proof proceeds by showing that, just as in the case of the Grenander estimator, $\hat{f}_n^c$ is given via the solution to an isotonic regression problem, which now involves "exponentially tilted" weights. Proposition 1 shows that evaluating the MLE $\hat{f}_n^c$ for various values of c amounts to a multiscale analysis: The expression given by Proposition 1 is plainly the image density of $\hat{g}_n^c$ under the transformation $y \mapsto (\log y)/c$, hence $\hat{f}_n^c$ obtains by the following three-step procedure:

1. Map the observations X_i onto a different scale: $X_i \mapsto Y_i = e^{cX_i}$.
2. Compute the null estimate (the Grenander estimate) for the transformed data.
3. Transform the resulting estimate back onto the original scale with the inverse transformation $y \mapsto (\log y)/c$.

REMARK. Woodroofe and Sun (1999) treat a related problem using a penalized MLE. The above approach can also be put into a penalized ML framework. It is readily checked that $\hat{f}_n^c$ is the nonparametric MLE in the set $\mathscr{D}(c) := \{f\colon \log f(y) - \log f(x) \le c(y - x) \text{ for all } x < y\}$. But Green and Silverman [(1994), page 51] show that constrained maximum likelihood is just an alternative characterization of penalized maximum likelihood, with the Lagrange (tuning) multiplier of the appropriate penalty being a function of the constraint c. Thus the difference from the usual penalized ML approach is that we consider a range of Lagrange (tuning) parameters instead of trying to find an "optimal" one. This aspect is crucial for the optimality results derived in Section 4.

3. The test statistic and its null distribution. For the case where interest centers on local deviations from the null model, it was shown by Liero, Läuter and Konakov (1998), Neumann (1998) and Dümbgen and Spokoiny (2000) that it is advantageous to employ minimum distance goodness-of-fit statistics that are based on the supremum norm. Consequently we will measure the distance of f from H_0 by $\inf_{m \in \mathrm{Mon}} \| (\log f - m)w \|_\infty$, where Mon is the class of nonincreasing functions, and w is a weight function that allows downweighting the tails of the distribution, which is a desirable option in practice.

Here we use the notion of a distance in the usual loose sense [see, e.g., Titterington, Smith and Makov (1985), page 115]. We will treat in detail the case where $w = f^{1/3}$, as it can be shown that then, for the purpose of the following analysis, the above distance is equivalent to $\sup_{x<y} \frac{3}{2}(f^{1/3}(y) - f^{1/3}(x))$. This leads to the test statistics

$$\begin{aligned} T_n(c) &= \sup_{X_{(1)} \le x < y \le X_{(n)}} \tfrac{3}{2}((\hat{f}_n^c)^{1/3}(y) - (\hat{f}_n^c)^{1/3}(x)) \\ &= \max_{1 \le i < j \le n} \tfrac{3}{2} c^{1/3} \big((\hat{g}_n^c(e^{cX_j}) e^{cX_j})^{1/3} - (\hat{g}_n^c(e^{cX_{i+1}}) e^{cX_i})^{1/3} \big). \end{aligned}$$

Simultaneous use of the $T_n(c)$ across scales c requires an appropriate standardization of the T_n on each scale. Asymptotic considerations will yield a standardization that will be shown to result in a procedure that is adaptive and optimal in the asymptotic minimax sense.

To see heuristically how the distribution of T_n can be analyzed, it is informative to sketch Groeneboom's (1985) elegant derivation of Prakasa Rao's (1969) result on the pointwise limiting distribution of the Grenander estimator. Groeneboom noticed that this distribution can be derived from the limiting distribution of the process $U_n(a) = \sup\{t \ge 0\colon G_n(t) - at \text{ is maximal}\}$, where G_n denotes the empirical cdf. U_n can be interpreted as an inverse to the Grenander estimator $\hat{g}_n$. Writing $\sqrt{n}(G_n(t) - at) = \sqrt{n}(G_n(t) - G(t)) + \sqrt{n}(G(t) - at)$ and observing that the first term approximately equals a Brownian bridge while the second term behaves like a quadratic locally around $t_0 = g^{-1}(a)$, makes plausible that the limiting distribution of $\hat{g}_n(t_0)$, appropriately normalized, is given by the arg max of a two-sided Brownian motion with quadratic drift.

Considering now $T_n(c)$ in the case where $f = 1_{[0,1]}$, and thus the transformed density equals $g^c(y) = \frac{1}{cy}$ on $[1, e^c]$, we have $T_n(c) \le \sup_{x \in [0,1]} \times 3|(\hat{f}_n^c)^{1/3}(x) - 1| \approx \sup_{y \in [1, e^c]} |c\hat{g}_n^c(y)y - 1|$. Formally switching arguments by taking the sup over the range of $\hat{g}_n^c$ instead of its domain gives $\sup_a |caU_n(a) - 1| = \sup_a ca|U_n(a) - (g^c)^{-1}(a)|$. This heuristic shows that one needs to control the arg max of a Brownian bridge with quadratic drift, uniformly over varying centers of the quadratic. Furthermore, different values of c give rise to different curvatures of the quadratic. The following theorem gives the pertinent result for the limiting process.

THEOREM 1. *Let Y be either a standard Brownian motion or a standard Brownian bridge, and for $a \in [0,1]$, $c > 0$, define $V_c(a) := \arg\max_t \{Y(t) - c(t-a)^2\}$, where t ranges over the domain of Y. Then*

$$\sup_{c \ge e^e} \sup_{a \in [0,1]} \frac{c^{2/3}|V_c(a) - a| - (\log c)^{1/3}}{(\log c)^{-2/3} \log\log c} < \infty \quad a.s.$$

$V_c(a)$ is a.s. unique [see Kim and Pollard (1990)] and is the location on the t-axis of the point where the parabola $c(t-a)^2 + b$, sliding down along the line $t = a$, hits Y. Figure 2 provides an illustration of the situation for

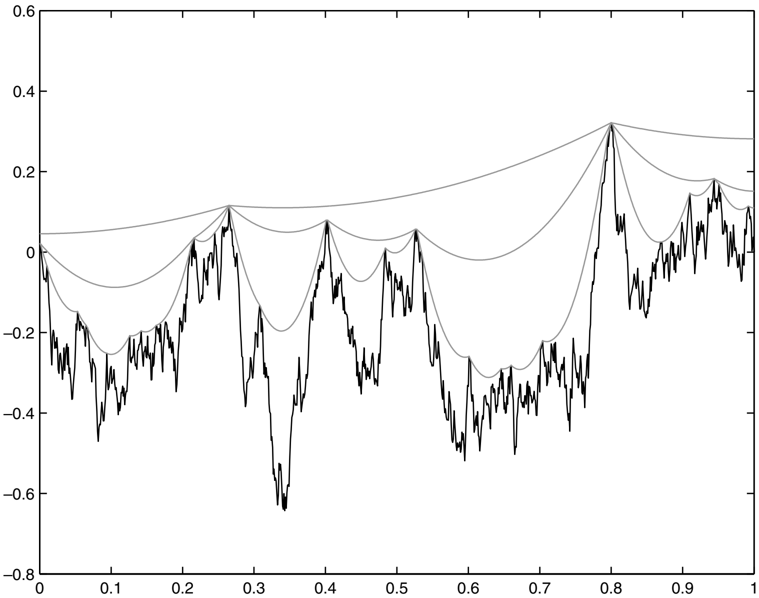

FIG. 2. *"Excursions under parabolas" at various scales of the Brownian bridge.*

three different scales c and a standard Brownian bridge Y. For each c only the lower envelope of all the parabolas with centers in $[0, 1]$ is shown. Note that apart from a vicinity of 0 and 1, this lower envelope coincides with a parabola that hits Y at least twice, and $\sup_a |V_c(a) - a|$ is attained for one of those parabolas. One sees that varying c corresponds to looking at "excursions under parabolas" at various scales of the Brownian bridge.

Theorem 1 yields the appropriate standardization of the test statistic $T_n(c)$ across scales.

THEOREM 2. *Let the X_i, $i \geq 1$, be iid $U[0, 1]$. Then*

$$\overline{\lim_{n\to\infty}} \sup_{c\in[e^e n^{-1/2},\, n/\log^{10} n]} \frac{(n/4c)^{1/3} T_n(c) - (\log(\sqrt{n}c/2))^{1/3}}{\log(\sqrt{n}c/2))^{-2/3} \log\log(\sqrt{n}c)} \leq L \quad a.s.$$

where L is a real random variable. If the X_i', $i \geq 1$, are iid $f \in H_0$ and generated by the inverse probability transform from the X_i, then the corresponding statistic is dominated by the above statistic eventually a.s.

Denote by $\phi_n(T_n)$ the test that rejects H_0 iff the sup in Theorem 2 exceeds a critical value $l_n(1-\alpha)$ to be specified. The recipe provided by Theorem 2 for obtaining $l_n(1-\alpha)$ is to evaluate the statistic for Monte Carlo samples of size n drawn from $U[0, 1]$. By Fatou's lemma, $\phi_n(T_n)$ has then asymptotically level α. Replacing the sup with the maximum over a finite grid of c-values should not have a large effect on the efficiency of the test. The Grenander estimator

$\hat{g}_n^c$ can be computed using standard algorithms such as the pool adjacent violators algorithm (PAVA) [see Robertson, Wright and Dykstra (1988)] or related versions which appear to run in $O(n \log n)$ time [see, e.g., Zhang and Newton (1997)].

4. Optimality. Theorem 3 below shows that the test $\phi_n(T_n)$ is asymptotically optimal in the minimax sense as described in the survey of Ingster (1993), and also adaptive with respect to the unknown parameter c. It is shown in Ingster (1993) that a meaningful set-up for such optimality results requires a restriction on the set of alternatives under consideration, usually via a smoothness assumption. Note that Lemma 1 states what kind of regularity will naturally be available for the mixture analysis: Any increase in the log-density in model (1) must satisfy a Lipschitz condition. However, the general semiparametric model (1) also allows for discontinuous decreases.

We denote by H_c the class of densities that satisfy (1), and by $\delta(f, H_0) = \sup_{x<y} \frac{3}{2}(f^{1/3}(y) - f^{1/3}(x))$ the distance of f from H_0 introduced in Section 3. Part (a) of the following theorem states that a meaningful test of H_0 is generally impossible if the alternative is in H_c and its distance from the null hypothesis is $C(\frac{\log n}{n})^{1/3}$ with $C < (2c)^{1/3}$: Any test with asymptotic level α has asymptotically a type II error of at least $1-\alpha$ for some alternative of the described form; that is, its asymptotic power is no larger than its significance level. On the other hand, part (b) of the theorem shows that if $C > (2c)^{1/3}$, then for the above test $\phi_n(T_n)$ the maximal type II error over the set of these alternatives goes to zero. Then $(2c)^{1/3}$ is called the exact separation constant, and $(\frac{\log n}{n})^{1/3}$ the minimax rate of testing; see Ingster (1993).

THEOREM 3. *The minimax rate of testing H_0 versus the semiparametric alternative H_c is $(\frac{\log n}{n})^{1/3}$ and the exact separation constant is $(2c)^{1/3}$. The test $\phi_n(T_n)$ with asymptotic level $\alpha \in (0,1)$ is asymptotically minimax and adaptive.*

(a) *If $d_n = C(\frac{\log n}{n})^{1/3}$ with $C < (2c)^{1/3}$, then*

$$\varliminf_{n\to\infty} \inf_{\psi_n} \sup_{f\in H_c:\, \delta(f,H_0)\ge d_n} P_f(\psi_n(\underline{X}_n) = 0) \ge 1-\alpha,$$

where $\inf_{\psi_n}$ denotes the infimum over all tests with level $\alpha_n \to \alpha$ that are based on an iid sample $\underline{X}_n = (X_1, \dots, X_n)$ from f.

(b) *If $d_n = C(\frac{\log n}{n})^{1/3}$ with $C > (2c)^{1/3}$, then*

$$\lim_{n\to\infty} \sup_{f\in H_c:\, \delta(f,H_0)\ge d_n} P_f(\phi_n(T_n) = 0) = 0.$$

Theorem 3 makes precise the adaptivity property of the test ϕ_n: in a certain sense it performs as well as any level α test possibly can, even if the latter were allowed to use knowledge of the unknown parameter c.

5. Applications.

5.1. *Detecting an increase in the failure rate.* Let $X_1, \dots, X_n$ denote the failure times from a continuous distribution on $[0, \infty)$. Then the normalized spacings are given by $D_i := (n-i+1)\,(X_{(i)} - X_{(i-1)})$ and the studentized total time on test statistics by $W_i := \sum_{j=1}^{i} D_j \big/ \sum_{k=1}^{n} D_k$, $i = 1, \dots, n$ ($X_{(0)} := 0$) [see Robertson, Wright and Dykstra (1988), Chapter 7]. The cumulative total time on test procedure (CTTT) described in Section 1 uses the statistic $\sum_{i=1}^{n-1} W_i$ [see Robertson, Wright and Dykstra (1988), Chapter 7.6].

Using the total time on test transformation, the multiscale maximum likelihood (MSML) procedure is immediately applicable to detect locally monotone parts in the failure rate.

THEOREM 4. *The assertion of Theorem* 2 *remains valid for testing the hypothesis of a nonincreasing failure rate on* $[0, \infty)$, *provided only that the statistic* T_{n-1} *is computed with the* W_i, $1 \le i \le n-1$, *in place of the* X_i, *and the exponential distribution takes the place of the uniform distribution.*

In the case where the failure rate is constant, that is, the X_i are iid $E(1)$, the assertion is an immediate consequence of the fact that the joint law of $(W_1, \dots, W_{n-1})$ is the same as that of the order statistics of $n-1$ iid $U[0, 1]$ random variables [see Shorack and Wellner (1986), Chapter 21.1].

Testing whether the failure rate is nondecreasing is analogous by changing T_{n-1} in an obvious way. Theorem 2 remains clearly also valid if the denominator is set to 1. This simplification was used in the following with hardly any effect on the simulation results, due to the fact that the denominator varies very slowly with c.

The performance of the MSML statistics will now be illustrated by a small simulation study. We will sample from a distribution whose failure rate is constant up to some point t_0, and linearly increasing thereafter. Thus the change-point t_0 marks the beginning of a wear-out period. The goal is to detect the presence (and location) of the increasing part. No decreasing stretch was built into the failure rate so that the CTTT is also applicable to this problem, thus allowing a comparison with the MSML procedure to show the limitations of the latter in this extreme case. The null distribution of the MSML statistic was obtained from 10,000 Monte Carlo samples using exponentially distributed random samples with the given sample size. The set of scales c was taken to be the integers from 1 to 10. Using finer disretizations did not change the results much. The CTTT statistic was evaluated against its limiting normal distribution [see Robertson, Wright and Dykstra (1988), Chapter 7.6]. Both tests were evaluated at the 5% significance level for 10,000 Monte Carlo samples of observations from distributions with failure rates $r(t) = 1/2 + s(t-t_0)^+$. For a given change-point t_0 the slope s was chosen so that the powers obtained in the simulation fell into a nontrivial range. The case $t_0 = 0$ is tailor made for the CTTT statistic, as the failure rate increases globally on the support. As expected, Table 1 shows that in this case it dominates the MSML statistic,

TABLE 1
Powers of the multiscale maximum likelihood (MSML) and cumulative total time on test (CTTT) procedures for alternatives with failure rates $r(t) = 0.5 + s(t - t_0)^+$

	$t_0 = 0$, $s = 0.1$		$t_0 = 2.5$, $s = 0.25$		$t_0 = 5$, $s = 2$	
Sample size	**MSML**	**CTTT**	**MSML**	**CTTT**	**MSML**	**CTTT**
250	0.607	0.883	0.785	0.799	0.426	0.414
500	0.898	0.991	0.990	0.976	0.892	0.700
750	0.978	0.999	1.000	0.997	0.993	0.852

which has to account for looking simultaneously over many substretches of the data. As t_0 increases, the trend becomes more local, modeling the onset of a wear-out period. The simulations show how the MSML statistic becomes the more powerful test for detecting the increase. The MSML statistic also allows localizing the change-point t_0 by retracing which stretch of the data results in the largest value of the statistic. Note that the simulations treat an extreme case that is unfavorable for the MSML statistic: using bathtub-shaped failure rates (i.e., a local decrease is present) would gravely impair the performance of the CTTT statistic, while the MSML statistic is designed to handle such a case.

5.2. *Detecting the presence of mixing.* Assume the X_i to be ordered and set $X'_i := \widetilde{F}_n(X_i - d)s_d$, where the factor $s_d := \#\{i\colon X_i \geq X_1 + d\}/(n\widetilde{F}_n(X_n - d))$ scales the X'_i linearly into an interval with length equal to the fraction of nonzero X'_i. Analogously to Section 3, set $T_n(c) := \max_{\varepsilon < X'_i < X'_j < 1-\varepsilon} \frac{3}{2} c^{1/3} \times \big((\hat{g}_n^c\,(e^{cX'_j})e^{cX'_j})^{1/3} - (\hat{g}_n^c(e^{cX'_{i+1}}))e^{cX'_i})^{1/3}\big)$, with the $\hat{g}_n^c$ computed using the X'_i instead of the X_i. The relevant statistic for this problem is then $T'_n(c) := \sup_{d>0} T_n(c)$. To avoid the lengthy analysis for the case where $\tilde{F}_n$ equals the smoothed ecdf, we give a result for the case where $\widetilde{F}_n$ is the MLE under the null model. The log-concave MLE $\tilde{f}_n$ can be readily computed using the iterative convex minorant algorithm [Jongbloed (1998); see also Walther (2000b)]. The theoretical properties of a log-concave MLE are similar to those of the MLE of a concave density, and the arguments in Groeneboom, Jongbloed and Wellner (2001) suggest that the uniform rate of convergence is $O((\log n/n)^{2/5})$. If the scales c are contained in the interval $\big[(\log^2 n/n)^{1/5}, n/\log^{10} n\big]$, then a result analogous to Theorem 2 holds for T'_n.

THEOREM 5. *Let $X_1, \dots, X_n$ be iid from a log-concave distribution. Then under the assumptions stated prior to the theorem, the assertions of Theorem 2 hold for T'_n in place of T_n.*

Figure 3 shows a histogram, plotted using the default settings of the Matlab histogram command, of flow cytometry measurements on 270 cells that were obtained in the Herzenberg Laboratory in the Genetics Department at Stanford. Flow cytometry uses light-induced fluorescence to measure certain

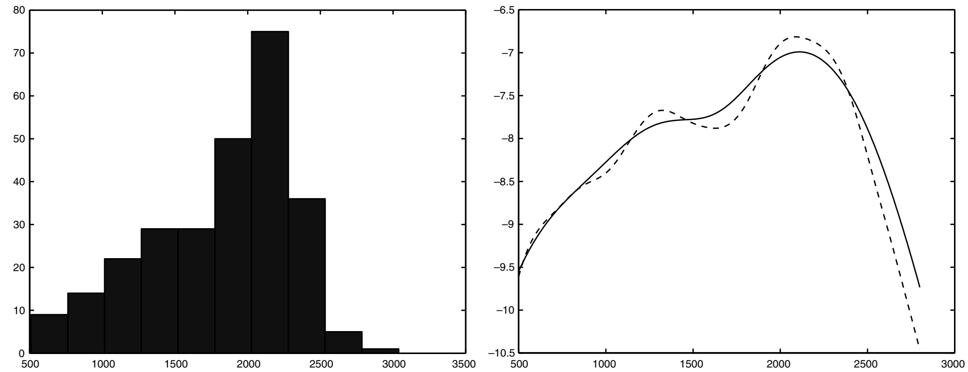

FIG. 3. *Left*: *Histogram of the flow cytometry data*. *Right*: *Logarithm of the kernel density estimates using the rule-of-thumb bandwidth* (*solid line*) *and the Sheather–Jones bandwidth* (*dashed line*).

characteristics of cells. One goal of the analysis is to detect the presence of subpopulations of cells, which could signal the presence of certain diseases. Also shown in Figure 3 are the logarithms of kernel density estimates evaluated with the rule-of-thumb bandwidth based on the interquartile range [see Section 3.4.2 of Silverman (1986)], as well as the Sheather–Jones bandwidth [see Sheather and Jones (1991)]. Both of these estimates confirm that the log-concave model is plausible for the component distributions. Both estimates also show a violation of concavity in the center of the data, suggesting that a mixture is present.

The null distribution of the above test was obtained from 1,000 Monte Carlo samples from the uniform distribution. A grid of ten equally spaced values between 0 and 5 was used for the range of c, and 20 equally spaced values between 0 and the standard deviation of the data for the range of d. The results were not sensitive to these choices. The resulting p-value was 0.037, indicating the presence of a mixture.

6. Outlook. The mixture problem can alternatively be analyzed without employing a transformation. A key to that approach is the following proposition, which will be stated without proof.

PROPOSITION 2. *Let the f_i be log-concave densities on $\mathbf{R}^d$ with common support S and $p_i > 0$; $i = 1, \ldots, m$. Then on any compact subset of S^o the representation*

$$\sum_i^m p_i f_i(x) = \exp(\phi(x) + c \parallel x \parallel^2) \tag{2}$$

holds for a concave function ϕ on $\mathbf{R}^d$ and a constant $c \geq 0$.

Thus the proposition leads to a second-order version of the semiparametric model (1), where ϕ is concave instead of monotone, and the quadratic cx^2 takes the place of the linear term cx. Further motivation for studying this model derives from the direct and important extension to the multivariate setting given in Proposition 2.

7. Proofs.

PROOF OF LEMMA 1. The assertion is trivial if F is degenerate. Otherwise F has a density f that is positive and continuous on $(F^{-1}(0), F^{-1}(1))$, where $F^{-1}(0) := \inf\{x\colon F(x) > 0\}$ and $F^{-1}(1) := \sup\{x\colon F(x) < 1\}$. Thus the cdf F_d of $F(X-d)$ is given on $[0, 1]$ by $F_d(t) = P(X \le F^{-1}(t)+d) = F(F^{-1}(t)+d)$, which is continuously differentiable on $(0, 1)\backslash\{F(F^{-1}(1)-d)\}$ and continuous on $(0, 1)$. Thus the distribution of $F(X-d)$ is absolutely continuous on $(0, 1)$ (and possibly has an atom at 0). Differentiating F_d shows that the Radon–Nikodym derivative is $f_d(t) = f(F^{-1}(t)+d)/f(F^{-1}(t))$; $t \in (0, 1)$. Taking logs and using the fact that $F^{-1}(\cdot)$ increases continuously from $F^{-1}(0)$ to $F^{-1}(1)$ shows that f_d is nonincreasing on (0,1) iff $u \mapsto \log f(u+d) - \log f(u)$ is nonincreasing on $(F^{-1}(0), F^{-1}(1))$. However, validity of this property for all $d > 0$ is equivalent to $\log f$ being concave on $(F^{-1}(0), F^{-1}(1))$ as $\log f$ is measurable.

Finally, if $I \subset (0,1)$ is a closed interval, then $M := \sup_{t\in I} \frac{d}{dt}F^{-1}(t) = \sup_{t\in I} \frac{1}{f(F^{-1}(t))} < \infty$ as f is positive and continuous on $(F^{-1}(0), F^{-1}(1))$. Thus $F^{-1}(t) = Mt + \phi_1(t)$ for a nonincreasing function ϕ_1 on I. Proposition 2 gives the representation $f(F^{-1}(t)) = \exp(\phi(F^{-1}(t)) + c(F^{-1}(t))^2)$, $t \in I$, where ϕ is concave and $c \ge 0$. Thus $\log f_d(t) = \phi(F^{-1}(t)+d) - \phi(F^{-1}(t)) + 2cdF^{-1}(t) + cd^2$. The stated increment of ϕ is nonincreasing in t as ϕ is concave. Substituting $F^{-1}(t) = Mt + \phi_1(t)$ in the term $2cdF^{-1}(t)$ proves the lemma. □

PROOF OF PROPOSITION 1. A simple argument [see, e.g., page 326 in Robertson, Wright and Dykstra (1988)] shows that the MLE must be of the form $\log \hat{f}_n^c(x) = \hat{\phi}(x) + cx$, where $\hat{\phi}(x)$ equals a constant $\hat{\phi}_i$ on $(x_{i-1}, x_i]$, $i = 1, \dots, n$, and $\hat{\phi}(x) = -\infty$ for $x \in (-\infty, x_0] \cup (x_n, \infty)$. Here $x_i := X_{(i)}$, $i = 1, \dots, n$, and $x_0 := 0$. Thus $\hat{\phi}$ is given by the solution of the optimization problem

$$\max \sum_{i=1}^{n} \hat{\phi}_i$$

$$\text{subject to} \quad \hat{\phi}_1 \ge \hat{\phi}_2 \ge \cdots \ge \hat{\phi}_n \quad \text{and} \quad \sum_{i=1}^{n} \exp(\hat{\phi}_i) \int_{x_{i-1}}^{x_i} e^{ct}\, dt = 1.$$

Example 1.5.7 of Robertson, Wright and Dykstra (1988) shows that $(\exp(\hat{\phi}_i), 1 \le i \le n)$ is the antitonic regression of $(g_1, \dots, g_n)$ with weights $(w_1, \dots, w_n)$, where $w_i = \int_{x_{i-1}}^{x_i} e^{ct}\, dt = (e^{cx_i} - e^{cx_{i-1}})/c$ and $g_i = 1/(nw_i)$, $i = 1, \dots, n$. Theorem 1.4.4 in Robertson, Wright and Dykstra (1988), applied for antitone instead of isotone regression, gives $\exp(\hat{\phi}_i) = \min_{s\le i-1} \max_{t\ge i}(\sum_{j=s+1}^{t} w_i g_i / \sum_{j=s+1}^{t} w_i)$. But the last fraction equals $\frac{c}{n}(t-s)/(e^{cx_t} - e^{cx_s}) = c(G_n(y_t) - G_n(y_s))/(y_t - y_s)$, where $y_i := e^{cx_i}$, $i = 0, \dots, n$, and G_n is the empirical cdf of the $(y_i,\ i = 1, \dots, n)$. Hence $\exp(\hat{\phi}_i)/c$ is again the least concave majorant of an empirical cdf, but this time of G_n and evaluated at y_i. The proposition follows. □

PROOF OF THEOREM 1. The proof employs some special properties of the process V together with a covering argument and an exponential inequality. See Shorack and Wellner [(1986), page 536] or Dümbgen and Spokoiny [(2000) proof of Theorem 6.1] for related approaches to derive results on the modulus of continuity of Brownian motion.

If follows from the definition of $V_c(a)$ that $a \mapsto V_c(a)$ is nondecreasing and $c \mapsto |V_c(a) - a|$ is nonincreasing (for any function Y on $\mathbf{R}$). For $k \ge 4$ define the rectangle $R_k := [0,1] \times [2^k, 2^{k+1})$ and the lattice $L_k := \{(i/(k2^k), 2^k + j2^k/k),\ 1 \le i \le k2^k,\ 0 \le j \le k-1\} \subset R_k$. Now consider an arbitrary pair $(a, c^{2/3}) \in R_k$ and let $(\tilde{a}, \tilde{c}^{2/3}) \in L_k$ be such that $a \le \tilde{a}$, $a - \tilde{a} \ge -1/(k2^k)$, and $c \ge \tilde{c}$, $c^{2/3} - \tilde{c}^{2/3} \le 2^k/k$. Then $c^{2/3}/(k2^k) \le 2/k$ and $(\tilde{c}/c)^{2/3} \ge \tilde{c}^{2/3}/(\tilde{c}^{2/3} + 2^k/k) \ge 1 - 1/k$. Hence for $\lambda \in \mathbf{R}$ the inequality $c^{2/3}(V_c(a) - a) \ge \lambda$ together with the above monotonicity properties of $V_c(a)$ imply $\tilde{c}^{2/3}|V_{\tilde{c}}(\tilde{a}) - \tilde{a}| \ge (\tilde{c}/c)^{2/3} c^{2/3}(V_c(a) - a - 1/(k2^k)) \ge (1 - 1/k)(\lambda - 2/k)$. Now define the event

$$A_k := \left[\frac{c^{2/3}(V_c(a) - a) - (3/2\ \log c^{2/3})^{1/3}}{(3/2 \log c^{2/3})^{-2/3} \log\log c} > 16 \text{ for some } (a, c^{2/3}) \in R_k\right],$$

and observe that $2^k \le c^{2/3} \le 2^{k+1}$ entails $(3/2 \log c^{2/3})^{1/3} + 16(3/2 \log c^{2/3})^{-2/3} \log\log c \ge (3/2 \log 2^k)^{1/3} + 6(3/2 \log 2^k)^{-2/3} \log\log 2^k + 2/k =: \lambda_k$. Hence A_k implies $\tilde{c}^{2/3}|V_{\tilde{c}}(\tilde{a}) - \tilde{a}| \ge (1 - 1/k)(\lambda_k - 2/k)$ for some $(\tilde{a}, \tilde{c}^{2/3}) \in L_k$.

It is helpful now to take for Y the two-sided Brownian motion originating from the origin. Using Brownian scaling one sees that $\mathscr{L}(c^{2/3}V_c(a)) = \mathscr{L}(V_1(ac^{2/3}))$. The process $a \mapsto V_1(a) - a$ is stationary and the tail behavior of its marginal density is given by $f_z(t) \sim 4^{4/3}/2|t| \exp(-2/3|t|^3 + 2^{1/3}a_1|t|)/A$ (as $|t| \to \infty$), where $a_1 \approx -2.3381$ and $A \approx 0.7022$ [see Groeneboom (1989), Corollary 3.4]. Let the random variable Z have density f_Z. Then for L large enough $P(|Z| > L) \le \int_L^\infty Ct \exp(-2/3t^3)\,dt \le C \int_L^\infty t^2/L \exp(-2/3t^3)\,dt = C/(2L)\exp(-2/3L^3)$ for some constant C. Together with $\#L_k = k^2 2^k$ one obtains for k large enough,

$$\begin{aligned} P(A_k) &\le \frac{Ck^2 2^k}{(1 - 1/k)(\lambda_k - 2/k)} \exp(-2/3(1 - 1/k)^3(\lambda_k - 2/k)^3) \\ &\le \frac{Ck^2 2^k}{(3/2 \log 2^k)^{1/3}} \exp(-2/3(1 - 3/k)(3/2 \log 2^k + 18 \log\log 2^k)) \end{aligned}$$

$$\le \frac{Ck^2 2^k}{(3/2\log 2^k)^{1/3}} \exp(\log 2^{-k+3} - 3\log\log 2^k)$$

$$\le \frac{Ck^2}{(3/2\log 2^k)^{10/3}} \le C/k^{4/3}.$$

Thus $\sum_k P(A_k) < \infty$. An analogous proof shows that this result is also true with $c^{2/3}(V_c(a) - a)$ replaced by $-c^{2/3}(V_c(a) - a)$, so the first Borel–Cantelli lemma establishes the theorem in the case where Y is a two-sided Brownian motion [note that $\sup_{e^e \le c < C} \sup_{a\in[0,1]} c^{2/3}|V_c(a) - a| < \infty$ for all $C > e^e$ by the monotonicity properties of $V_c(a)$].

Now define $\widetilde{V}_c(a) := \arg\max_{t\ge 0}\{W(t) - c(t-a)^2\}$, where W is two-sided Brownian motion. The proof above shows that a.s. $\inf_{a\in[c^{-1/3},1]} V_c(a) > 0$ for c large enough, whence $\widetilde{V}_c(a) = V_c(a)$ a.s. for $a \in [c^{-1/3}, 1]$ and c large enough. Thus, to prove the theorem for one-sided Brownian motion, it is enough to prove it for $\widetilde{V}_c(a)$ with a ranging only over $[0, c^{-1/3}]$. This proof proceeds just as before, the important difference being the tail estimate: The two events $[\arg\max_{t\ge a}\{W(t)-c(t-a)^2\} - a > L]$ and $[\arg\max_{t\le a}\{W(t)-c(t-a)^2\} - a < -L]$ are independent, have the same probability $p(L)$ (as $W(t)$ run backwards from a is one-sided Brownian motion) and jointly imply $|V_c(a) - a| > L$. So for $a \ge 0$ one finds $P(|\widetilde{V}_c(a) - a| > L) \le 2p(L) \le 2\sqrt{P(|V_c(a) - a| > L)}$, and thus $P(c^{2/3}|\widetilde{V}_c(a) - a| > L) \le C/\sqrt{L}\exp(-\frac{1}{2}\frac{2}{3}L^3)$ for large enough L. The additional factor $\frac{1}{2}$ in the exponent is compensated for by the fact that $\#(L_k \cap ([0, 2^{-k/2}] \times [2^k, 2^{k+1}])) \le k^2 2^{k/2}$, which gives

$$P\left(\frac{c^{2/3}\left(\widetilde{V}_c(a) - a\right) - \left(\frac{3}{2}\log c^{2/3}\right)^{1/3}}{\left(\frac{3}{2}\log c^{2/3}\right)^{-2/3}\log\log c} > 32 \right.$$

$$\left.\text{for some } (a, c^{2/3}) \in \left[0, 2^{-k/2}\right] \times \left[2^k, 2^{k+1}\right]\right) \le C/k^{4/3}.$$

Finally, for a Brownian bridge B write $B(t) = W(t) - tW(1)$, where W is one-sided Brownian motion. Then $V_c^B(a) := \arg\max_{t\in[0,1]}\{B(t) - c(t-a)^2\} = \arg\max_{t\in[0,1]}\{W(t) - c(t - a + W(1)/(2c))^2\} =: V_c^W(a - W(1)/(2c))$. One checks that $V_c^W(a - W(1)/(2c)) \le V_c^W(a)$ for $0 \le a \le W(1)/(2c)$, and then $\sup_{a\in[0,1]} c^{2/3}|V_c^B(a) - a| \le \sup_{a\in[0,1+|W(1)|/(2c)]} c^{2/3}|V_c^W(a) - a| + |W(1)|/c^{1/3}$. The claim for the Brownian bridge now follows because the theorem holds for V_c^W. The restriction of the arg max to [0, 1] can be dealt with just as in the step from two-sided to one-sided Brownian motion above. □

PROOF OF THEOREM 2. The proof will make use of the following corollary to Theorem 1, which is a consequence of the law of the iterated logarithm, as well as the subsequent lemma. See Walther (2000a) for details of the proofs.

COROLLARY 1. *Let Y be as in Theorem* 1 *and* $p \geq 0$. *Then*

$$\sup_{c\geq e^e}\ \sup_{a\in[-p,1+p]} \frac{c^{2/3}|V_c(a)-a| - (\log c)^{1/3} \vee pc^{2/3}}{(\log c)^{-2/3}\log\log c} < \infty \quad a.s.$$

LEMMA 2. *Let $A \subset B$ be compact intervals on the real line, and $f, g: A \to \mathbf{R}$ be upper semicontinuous. For $c > 0$ set $U_f(c,x) := \sup\{t \in A: f(t) - c(t-x)^2$ is maximal$\}$ and $\Omega_f(c) := \sup_{x\in B}|U_f(c,x) - x|$. Then for all $c > 0$,*

$$|\Omega_g(c) - \Omega_f(c)| \leq 2\sqrt{\|f-g\|_\infty/c}.$$

To fix notation, note that if the X_i, $i = 1, \dots, n$ are iid f with cdf F on [0, l], then the $Y_i := e^{cX_i}$ have density $g^c(y) = f(\frac{\log y}{c})/cy$ with cdf $G^c(y) = F(\frac{\log y}{c})$ on $D_c = [1, e^c]$. F_n and G_n^c denote the empirical cdf of the X_i's and Y_i's, respectively, and $\hat{f}_n^c$ and $\hat{g}_n^c$ are defined in Proposition 1.

First let f be the uniform density on [0, 1], so $g^c(y) = \frac{1}{cy}$ on $D_c = [1, e^c]$. The central step of the proof consists of employing Theorem 1 and Corollary 1 to show the following multiscale result for the Grenander estimator $\hat{g}_n^c$:

$$\limsup_{n\to\infty}\ \sup_{c\in\left[e^e n^{-1/2},\frac{n}{\log^{10} n}\right]}\ \sup_{y\in D_{c,n}} \frac{(\sqrt{n}c/2)^{2/3}y(\hat{g}_n^c(y) - g^c(y)) - (\log(\sqrt{n}c/2))^{1/3}}{(\log(\sqrt{n}c/2))^{-2/3}\log\log(\sqrt{n}c)} \leq K \quad \text{a.s.,} \tag{3}$$

where $D_{c,n} := [1 + l_{c,n}, e^c]$, the random variable K is the a.s. finite value asserted by Corollary 1, and one can take $l_{c,n} = 8(K \vee 1)\log\log(\sqrt{n}c)\times(\frac{c}{n\log^2(\sqrt{n}c/2)})^{1/3}$. The assertion of the theorem follows from $3((\hat{f}_n^c)^{1/3}(\frac{\log y}{c}) - 1) \leq \hat{f}_n^c(\frac{\log y}{c}) - 1 = cy(\hat{g}_n^c(y) - g^c(y))$ together with (3), and by proceeding similarly with $3(1 - (\hat{f}_n^c)^{1/3}(\cdot\cdot))$ using the companion result to (3) for $-(\hat{g}_n^c(y) - g^c(y))$.

To prove (3), note that $G^c(y) = \frac{\log y}{c}$ on D_c and $g^c(y) = \frac{1}{cy}$ has range $R_c := [\frac{1}{ce^c}, \frac{1}{c}]$. For $a > 0$ define

$$\begin{aligned} U_n(a) &:= \sup\{y \geq 1: G_n^c(y) - ay \text{ is maximal}\} \\ &= \sup\{y \in D_c: \sqrt{n}(G_n^c(y) - G^c(y)) + \sqrt{n}(G^c(y) - ay) \text{ is maximal}\} \end{aligned}$$

[note that $U_n(a) \in D_c$; see the picture in Groeneboom (1985), page 541]. The dependence of U_n on c will be suppressed. Setting $y = G^{c^{-1}}(u)$ in the above definition, one gets

$$\begin{aligned} \widetilde{U}_n(a) &:= G^c(U_n(a)) \\ &= \sup\{u \in [0,1]: IU_n(u) + \sqrt{n}(u - aG^{c^{-1}}(u)) \text{ is maximal}\}, \end{aligned} \tag{4}$$

where IU_n denotes the uniform empirical process. As the Grenander estimator is the left-continuous slope of the least concave majorant of the empirical cdf, one has

$$\hat{g}_n^c(y) \leq z \quad \Leftrightarrow \quad U_n(z) \leq y \quad \text{a.e. } (y, z) \tag{5}$$

[see (2.2) in Groeneboom (1985)]. Set $x := (\log(\sqrt{n}c/2))^{1/3} + (K+\delta)\times (\log(\sqrt{n}c/2))^{-2/3}\log\log(\sqrt{n}c)$, where δ is an arbitrary positive number, and observe that for a generic $y \in D_{c,n}$ and $a := \frac{1}{cy}$,

$$\begin{aligned}
&\left(\sqrt{n}c/2)^{2/3} y(\hat{g}_n^c(y) - \frac{1}{cy}\right) \le x \\
&\quad\Leftrightarrow \hat{g}_n^c(y) \le \frac{1}{cy} + \frac{xn^{-1/3}(c/2)^{-2/3}}{y} \\
&\quad\Leftrightarrow U_n\left(\frac{1}{cy} + \frac{xn^{-1/3}(c/2)^{-2/3}}{y}\right) \le y \quad \text{a.e. } y \text{ by (5)} \\
&\quad\Leftrightarrow \widetilde{U}_n(a(1+x(4c/n)^{1/3})) \le G^c\left(\frac{1}{ac}\right) \text{ by (4)} \\
&\quad\Leftrightarrow \widetilde{U}_n(a(1+x(4c/n)^{1/3})) - u_a \le \frac{1}{c}\log((1+x(4c/n)^{1/3}),
\end{aligned}$$

where $u_a := -\frac{1}{c}\log(ac(1+x(4c/n)^{1/3}))$; the dependence of u_a on c and n will be suppressed. Thus (3) will follow once we show that a.s.,

$$\sup_{a:\ \frac{1}{ac}\in D_{c,n}} \left(\widetilde{U}_n(a(1+x(4c/n)^{1/3})) - u_a\right) \le \frac{1}{c}\log(1+x(4c/n)^{1/3}) \tag{6}$$

for all $c \in [e^e n^{-1/2}, \frac{n}{\log^{10} n}]$, if n is large enough.

The plan is to show (6) by applying Theorem 1 and Corollary 1 to an approximation to $\widetilde{U}_n$ obtained by replacing the empirical process and the function $\sqrt{n}(u - aG^{c^{-1}}(u))$ by a Brownian bridge and a parabola, respectively, and then to incorporate the approximation error into this result. To this end, set $\widetilde{U}'_n(a) := \sup\{u \in [0,1]\colon B_n(u) - \sqrt{n}\frac{c}{2}(u-u_a)^2 \text{ is maximal}\}$, where $(B_n, n \in IN)$ is a sequence of Brownian bridges constructed on the same probability space as IU_n such that $\|IU_n - B_n\|_\infty = O(\log^2 n/\sqrt{n})$ a.s. [see Komlós, Major and Tusnády (1975)]. Corollary 1 gives

$$\begin{aligned}
&\sup_{a\in R_{c,n}} (\sqrt{n}c/2)^{2/3}|\widetilde{U}'_n(a) - u_a| \\
&\qquad \le (\log(\sqrt{n}c/2))^{1/3} + K(\log(\sqrt{n}c/2))^{-2/3}\log\log(\sqrt{n}c)
\end{aligned}\tag{7}$$

a.s. for all $c \in [e^e n^{-1/2}, n/\log^{10} n]$, if n is large enough (depending only on K). Here the set $R_{c,n}$ can be chosen such that $(ac)^{-1} \in D_{c,n}$ implies $a \in R_{c,n}$, which in turn implies $u_a \le 1$; see Walther (2000a) for details.

Now we account for the error incurred by employing $\widetilde{U}'_n$ instead $\widetilde{U}_n$. Retracing the proof of Theorem 2.1 in Groeneboom, Hooghiemstra and Lopuaä (1999) for our special case of the density g^c and employing some straightforward improvements for this case, one finds $P(n^{1/3}|U_n(a) - \frac{1}{ac}| > z) \le 2\exp(-z^3/24c(ac)^{-3})$. The first Borel–Cantelli lemma yields $|U_n(a) - (ac)^{-1}| = O((\log n)^{1/3}/a(c\sqrt{n})^{2/3})$ a.s. uniformly in $c > 0$ and $a \in R_c$. Together with the concavity of G^c one deduces $|\widetilde{U}_n(a)-G^c((ac)^{-1})| = O(\frac{(\log n)^{1/3}}{(c\sqrt{n})^{2/3}})$ a.s. uniformly in $c \in (0, n/\log^2 n]$ and $a \in R_c$. We now plug into this result $\bar{a} := a(1+x(4c/n)^{1/3})$ instead of a to obtain [note that $G^c((\bar{a}c)^{-1}) = u_a$]

$$|\widetilde{U}_n(a(1+x(4c/n)^{1/3})) - u_a| = O\left(\frac{(\log n)^{1/3}}{(c\sqrt{n})^{2/3}}\right) \quad \text{a.s.} \tag{8}$$

uniformly in $c \in [e^e n^{-1/2}, n/\log^2 n]$ and $a \in R_{c,n}$; see Walther (2000a) for the details when $u_a < 0$.

On $[0,1]$, $G^{c^{-1}}(u) = e^{cu}$. A Taylor series expansion gives $\sqrt{n}(u - a(1 + x(4c/n)^{1/3})e^{cu}) = -\sqrt{n}c/2(u-u_a)^2 - \sqrt{n}\frac{c^2}{6}e^{c\xi}(u-u_a)^3$ plus terms not involving u, where ξ lies between 0 and $u-u_a$. For u in the neighborhood of u_a given by (8), the term $\sqrt{n}\frac{c^2}{6}e^{c\xi}(u-u_a)^3$ is $O(\frac{\log n}{\sqrt{n}})$, uniformly in $c \in (0, n/\log n]$ and $a \in R_{c,n}$. Together with (4), (8) and the fact that $\arg\max_{u\in[0,1]}\{IU_n(u) - \sqrt{n}c/2(u-u_a)^2\}$ also falls into the neighborhood of u_a given by (8), one concludes that one can write $\widetilde{U}_n(a(1+x(4c/n)^{1/3})) = \sup\{u \in [0,1]\colon IU_n(u) - \sqrt{n}c/2(u-u_a)^2 + d_{n,a,c}(u) \text{ is maximal}\}$, where $\|d_{n,a,c}\|_\infty = O(\frac{\log n}{\sqrt{n}})$ a.s., uniformly in $c \in [e^e n^{-1/2}, n/\log^2 n]$ and $a \in R_{c,n}$. Together with Lemma 2 this yields $|\sup_{a\in R_{c,n}} |\widetilde{U}_n(a(1+x(4c/n)^{1/3})) - u_a| - \sup_{a\in R_{c,n}} |\widetilde{U}'_n(a) - u_a|| = O(\sqrt{\log^2 n/nc})$ a.s. uniformly in $c \in [e^e n^{-1/2}, n/\log^2 n]$. Now (6) and hence (3) follow from (7).

Finally, the case of a general $f \in H_0$ can be dealt with by appropriate modifications to the above proof, using the fact that F is concave, as well as some additional technical arguments. The details are omitted. □

PROOF OF THEOREM 3. For part (a), set $f_0 \equiv 1_{[0,1]}$. We will consider a uniform prior on the alternatives $f_{j,k} = f_0 + \phi_j + \psi_k$, $1 \le j, k \le m$, where ϕ_j, ψ_k and m are defined as follows: set $b := 2d_n(1+3d_n)/c$ and $\phi(x) := \frac{cb}{\exp(cb)-1}e^{cx} - 1$ on $[0,b]$, $\phi \equiv 0$ on $[0,b]^c$. Now define for $j \le m := \lfloor \frac{1}{2b} \rfloor$: $\phi_j(x) := \phi^+(x - \frac{1}{2} - (j-1)b)$ and $\psi_j(x) := -\phi^-(x - (j-1)b)$. So for n large, ϕ_j looks like $c(x - \frac{1}{2} - (j - \frac{1}{2})b)$ on $[\frac{1}{2} + (j-\frac{1}{2})b, \frac{1}{2} + jb]$, and ψ_j looks like $c(x - (j - \frac{1}{2})b)$ on $[(j-1)b, (j-\frac{1}{2})b]$. There are $m = m(c,n) \asymp (n/\log n)^{1/3}$ ϕ_j's supported on $[0,1]$. The dependence of ϕ_j, ψ_j, $f_{j,k}$, b and m on c and n will be suppressed. One readily checks that $\int_0^b \phi = 0$ and $(\phi(b)+1)^{1/3} - (\phi(0)+1)^{1/3} \ge 2/3d_n$, whence $f_{j,k} \in H_c$ and $\delta(f_{j,k}, H_0) \ge d_n$ for $1 \le j, k \le m$. We will show

$$\frac{1}{m^2}\sum_{1\le j,k\le m} e^{\Lambda^n_{j,k}} \to 1 \quad \text{in } P_0^n\text{-probability as } n \to \infty, \tag{9}$$

where $P^n_{j,k}$ denotes the probability measure pertaining to a sample $(X_1, \dots, X_n)$ drawn independently from $f_{j,k}$, and $\Lambda^n_{j,k} = \sum_{i=1}^n \log(1 + \phi_j(X_i) + \psi_k(X_i))$ denotes the log-likelihood ratio. Hence if $\{\tau_n(\underline{X}_n), n \ge 1\}$ is any sequence of tests with level $\alpha_n \to \alpha$, then for arbitrary $\varepsilon > 0$,

$$\begin{aligned}\sup_{f\in H_c:\ \delta(f,H_0)\geq d_n} P_f^n(\tau_n = 0) &\geq \frac{1}{m^2}\sum_{1\leq j,k\leq m} P_{j,k}^n(\tau_n = 0)\\ &\geq E_0^n\bigg(1(\tau_n = 1) + \frac{1}{m^2}\sum_{1\leq j,k\leq m} e^{\Lambda_{j,k}^n}1(\tau_n = 0)\bigg) - \alpha_n\\ &\geq (1-\varepsilon)P_0^n\bigg(\frac{1}{m^2}\sum_{1\leq j,k\leq m} e^{\Lambda_{j,k}^n} \geq 1-\varepsilon\bigg) - \alpha_n.\end{aligned}$$

So assertion (a) of the theorem follows from (9).

Set $S_j^n := \sum_{i=1}^n(\phi_j(X_i) - E_0\phi_j(X_i)) - \frac{1}{2}\sum_{i=1}^n E_0\phi_j^2(X_i)$ and define T_j^n analogously with ϕ_j replaced by ψ_j. We will prove (9) by showing

$$E_0^n|e^{\Lambda_{j,k}^n} - e^{S_j^n+T_k^n}| \to 0 \quad \text{uniformly in } j,k. \tag{10}$$

$$\frac{1}{m^2}\sum_{1\leq j,k\leq m} e^{S_j^n+T_k^n} \to 1 \quad \text{in } P_0^n\text{-probability.} \tag{11}$$

As for (10), fix j and k and use a Taylor expansion to write $\sum_{i=1}^n \log(1 + \phi_j(X_i) + \psi_k(X_i)) = \sum_{i=1}^n(\phi_j(X_i) + \psi_k(X_i)) - 1/2\sum_{i=1}^n(\phi_j(X_i) + \psi_k(X_i))^2 + \sum_{i=1}^n R_i$. Set $Y_i := 1/2(\phi_j(X_i) + \psi_k(X_i))^2 - 1/2E_{j,k}(\phi_j(X_i) + \psi_k(X_i))^2 - R_i + E_{j,k}R_i$. Then $E_{j,k}Y_i = 0$ and $|Y_i| \leq 6d_n^2$ for all i. Hoeffding's inequality gives $\sup_n E_{j,k}^n \exp(2\sum_{i=1}^n Y_i) = \sup_n \int_0^\infty P_{j,k}^n(2\sum_{i=1}^n Y_i \geq \log t)\,dt \leq 1 + \sup_n \int_1^\infty \exp\{-2(\log t)^2/(144nd_n^4)\}\,dt < \infty$ as $\sup_n nd_n^4 < \infty$. Further, $|E_{j,k}(\phi_j(X_i) + \psi_k(X_i))^2 - E_0\phi_j^2(X_i) - E_0\psi_k^2(X_i)| \leq \frac{3}{c}d_n^4$ and $E_{j,k}R_i \leq 2d_n^4$. Hence

$$\begin{aligned}&\sup_n E_{j,k}^n\big(e^{S_j^n+T_k^n-\Lambda_{j,k}^n}\big)^2\\ &\quad= \sup_n E_{j,k}^n \exp\bigg\{2\sum_{i=1}^n Y_i\bigg\}\\ &\qquad\times \exp\bigg\{\sum_{i=1}^n(E_{j,k}(\phi_j(X_i) + \psi_k(X_i))^2 - E_0\phi_j^2(X_i)\\ &\qquad\qquad - E_0\psi_k^2(X_i) - 2E_{j,k}R_i)\bigg\} < \infty\end{aligned} \tag{12}$$

[use $E_0\phi_j(X_i) = -E_0\psi_k(X_i)$]. It is readily seen that this bound holds uniformly in j,k. Next, $\mathrm{Var}_{j,k}^n(S_j^n + T_k^n - \Lambda_{j,k}^n) = \mathrm{Var}_{j,k}^n(\sum_{i=1}^n Y_i) \leq 36nd_n^4 \to 0$, as $|Y_i| \leq 6d_n^2$. Hence $S_j^n + T_k^n - \Lambda_{j,k}^n \to 0$ in $P_{j,k}^n$-probability uniformly in j,k. Together with the uniform integrability condition (12), (10) follows by a standard argument, and by employing $E_0^n|e^{\Lambda_{j,k}^n} - e^{S_j^n+T_k^n}| = E_{j,k}^n|1 - e^{S_j^n+T_k^n-\Lambda_{j,k}^n}|$.

To prove (11) we will show $\frac{1}{m}\sum_{j=1}^m e^{S_j^n} \to 1$ in P_0^n-probability. The proof with T_j^n in place of S_j^n is analogous. Note that the S_j^n are not independent. We will employ the following refinement of a conditioning idea which was used by Korostelev and Nussbaum (1995) in a related situation: Partition $[\frac{1}{2}, 1]$ into $[m/n^\varepsilon] \asymp n^{1/3-\varepsilon}/(\log n)^{1/3}$ intervals I_k^n, $k = 1,\dots,[m/n^\varepsilon]$, of equal length, where the small $\varepsilon > 0$ will be specified later. Denote by ν_k the random number of X_i's that fall into I^n_k. A standard calculation shows that the event $\mathscr{A}_n := [|\nu_k - \frac{n^\varepsilon}{2m}n| \le \sqrt{\frac{n^\varepsilon}{m}n}n^{\varepsilon/3}$ for all $k = 1, \dots, \lfloor m/n^\varepsilon \rfloor]$ satisfies $\lim_{n\to\infty} P^n_0(\mathscr{A}_n) = 1$. Under P^n_0 and conditional on the vector $\underline{\nu}$, the collections of X_i's pertaining to different I^n_k are independent, and within each I^n_k the X_i's are iid uniform. Each I^n_k contains the supports of a block of $\asymp n^\varepsilon$ consecutive ϕ_j's. Thus the random variables $\frac{1}{n^\varepsilon}\sum_{j\in \text{block } k} e^{S^n_j}$, $k = 1, \dots, \lfloor m/n^\varepsilon \rfloor$, are P^n_0-conditionally independent given $\underline{\nu}$. The assertion will follow once we prove $\frac{1}{m/n^\varepsilon}\sum_{k=1}^{m/n^\varepsilon}(1/n^\varepsilon \sum_{j\in\text{block } k} e^{S^n_j}) \to 1$ in conditional P^n_0-probability given $\underline{\nu}$, uniformly in $\underline{\nu} \in \mathscr{A}_n$. The proof of Corollary 10.1.2 in Chow and Teicher (1988) shows that it is enough to prove

$$E^n_0(e^{S^n_j}|\underline{\nu}) \to 1 \text{ uniformly in } j \in \{1, \dots, m\} \text{ and in } \underline{\nu} \in \mathscr{A}_n, \tag{13}$$

$$E^n_0\biggl(\frac{1}{n^\varepsilon}\sum_{j\in\text{block } k} e^{S^n_j}\cdot 1\biggl(\frac{1}{n^\varepsilon}\sum_{j\in\text{block } k} e^{S^n_j} > \frac{m}{n^\varepsilon}\delta\biggr)|\underline{\nu}\biggr) \to 0 \tag{14}$$
$$\text{uniformly in } k \in \{1, \dots, \lfloor m/n^\varepsilon \rfloor\} \text{ and in } \underline{\nu} \in \mathscr{A}_n, \text{ for every } \delta > 0.$$

LEMMA 3. *Let $t \ge 0$. Then*

$$E^n_0(e^{tS^n_j}|\underline{\nu}) = \exp\biggl(\frac{C^3}{6c}\log n \cdot (t^2 - t) + o(1)\biggr),$$

uniformly in $j \in \{1, \dots, m\}$ and in $\underline{\nu} \in \mathscr{A}_n$. Here C is the constant given in the statement of the theorem.

The proof of the lemma is based on a Taylor series expansion and standard inequalities. See Walther (2000a) for details. Now (13) follows by setting $t = 1$ in the lemma. Let j be an index in block k. The expression in (14) is not larger than $E^n_0(e^{S^n_j}\cdot 1(e^{S^n_l} > \frac{m}{n^\varepsilon}\delta$ for some $l \in$ block $k|\underline{\nu}) \le \sum_{l\in\text{block } k} E^n_0(e^{S^n_j}\cdot 1(e^{S^n_l} > \frac{m}{n^\varepsilon}\delta)|\underline{\nu}) \le \sum_{l\in\text{block } k} E^n_0(e^{S^n_j}e^{tS^n_l}|\underline{\nu})/(\frac{m}{n^\varepsilon}\delta)^t$ for any $t > 0$. As the conditional P^n_0-distribution of $e^{S^n_j}$ is the same as that of $e^{S^n_l}$, we get $E^n_0(e^{S^n_j}e^{tS^n_l}|\underline{\nu}) \le E^n_0(e^{(1+t)S^n_j}|\underline{\nu})$ by Cauchy–Schwarz. Setting $t = \sqrt{2c/C^3} - 1 > 0$ and applying the lemma shows that the above sum is not larger than $n^\varepsilon \exp(\frac{C^3}{6c}\log n \cdot (t+1)t)n^{-(1/3-\varepsilon)t}(\log n)^{t/3}\cdot O(1) = n^{(\sqrt{C^3/(2c)}-1)t/3+\varepsilon(1+t)}\cdot(\log n)^{t/3}\cdot O(1) \to 0$ for ε small enough, as $\sqrt{C^3/(2c)} - 1 < 0$. The proof of (a) is complete.

The proof of part (b) will make use of the following lemma.

LEMMA 4. *Let g be a nonincreasing density on $[1, z]$ and set $g^{-1}(a) := \inf\{y\colon g(y) < a\}$ and $U_n(a) := \sup\{y \ge 1\colon G_n(y) - ay$ is maximal$\}$. Assume there exist $y_1 \in (1, z)$ and $t > 0$ such that $g(y_1) > 0$ and $g(y) \ge g(y_1)\frac{y_1}{y}$ for $y \in [y_1(1-t), y_1 + t] \subset [1, z]$. Then for all $x \in [0, n^{1/3}t(1+z)]$ and $a := g(y_1)(1 + xn^{-1/3})$ the inequality*

$$P(n^{1/3}(U_n(a) - g^{-1}(a)) > x) \le \exp(-Kg(y_1)x^3)$$

holds for a constant K that does not depend on g. One can take $K = (1+z)^{-3}(2(z+t))^{-1}[(1+t)(z(1+t)+1)]^{-2}$.

The lemma is a generalization of Theorem 2.1 of Groeneboom, Hooghiemstra and Lopuaä (1999) in that g is not required to be smooth and the exponential bound is uniform in g apart from the factor $g(y_1)$. In turn, the lemma requires the link between a and x and that g satisfies the stated inequality. See Walther (2000a) for a proof of the lemma.

Now let $f \in H_c$ such that $\delta(f, H_0) \geq d_n$. Then (1) implies that the function g^c, defined via f at the beginning of the proof of Theorem 2, satisfies for all $1 \leq y_1 < y_2 \leq e^c$,

$$\begin{aligned} g^c(y_2) - g^c(y_1) &= f\left(\frac{\log y_2}{c}\right)\frac{1}{cy_2} - f\left(\frac{\log y_1}{c}\right)\frac{1}{cy_1} \\ &\leq f\left(\frac{\log y_1}{c}\right)\left(\frac{y_2}{y_1}\frac{1}{cy_2} - \frac{1}{cy_1}\right) = 0. \end{aligned}$$

So g^c is nonincreasing. As $\delta(f, H_0) \geq d_n$ there exist $0 \leq x_1 < x_2 \leq 1$ such that $f^{1/3}(x_2) - f^{1/3}(x_1) = \sup_{s<x_2}(f^{1/3}(x_2) - f^{1/3}(s)) \geq 2/3d_n$. Using (1) one readily deduces

$$f(x_1) \geq \left(\frac{2/3d_n}{\exp(c/3)-1}\right)^3 \quad \text{and} \quad y_2 - y_1 \geq \frac{(\log n)^{1/5}}{f^{1/3}(x_1)} n^{-1/3}, \tag{15}$$

where $y_i := e^{cx_i}$.

g^c satisfies the condition of Lemma 4: By the definition of x_1 we have $g^c(y) = f(\frac{\log y}{c})\frac{1}{cy} \geq f(\frac{\log y_1}{c})\frac{1}{cy} = g^c(y_1)y_1/y$ for $y \in [1, y_2]$. By (15) the last interval contains $[y_1(1-t_n), y_1+t_n] \subset [1, e^c]$, where $t_n = (\log n)^{1/6}n^{-1/3}/f^{1/3}(x_1)$, provided $x_1 \geq 2t_n/c$, which will be assumed from now on (otherwise one has to take $x_1 := 2t_n/c$ and use an additional argument). Applying Lemma 4 with $z := e^c$ and $x = x(f, n) := (\log n)^{1/6}/f^{1/3}(x_1)$ and using the monotonicity of g^c gives

$$P_f\left\{U_n\left(g^c(y_1)(1+xn^{-1/3})\right) \leq y_1 + xn^{-1/3}\right\} \geq 1 - \exp(-K_1\sqrt{\log n}), \tag{16}$$

where $K_1 > 0$ does not depend on f, n or x_1 [as $t_n \leq 1$ by (15)], and U_n is taken with respect to the random variables $Y_i := e^{cX_i}$. Using $b^{1/3} - a^{1/3} \leq \frac{1}{3}(b-a)a^{-2/3}$ for $a, b > 0$ we get

$$\begin{aligned} P_f\Bigg\{&\left(\frac{n}{4a}\right)^{1/3}\left((\hat{f}_n^c)^{1/3}\left(\frac{\log(y_1+xn^{-1/3})}{c}\right) - f^{1/3}(x_1)\right) \\ &\qquad\qquad \leq (\log n)^{1/6}4^{-1/3}(2e^c+1)/3\Bigg\} \\ \geq P_f\Bigg\{&\left(\frac{n}{4c}\right)^{1/3}\left(\hat{f}_n^c\left(\frac{\log(y_1+xn^{-1/3})}{c}\right) - f(x_1)\right)f^{-2/3}(x_1) \\ &\qquad\qquad \leq (\log n)^{1/6}4^{-1/3}(2e^c+1)\Bigg\} \end{aligned} \tag{17}$$

$$\begin{aligned}
&\geq P_f\{\hat{g}_n^c(y_1+xn^{-1/3})(y_1+xn^{-1/3})-g^c(y_1)y_1 \\
&\qquad \leq g^c(y_1)(xn^{-1/3}+y_1xn^{-1/3}+x^2n^{-2/3})\} \\
&\qquad \text{as } y_1xn^{-1/3}+x^2n^{-2/3}\leq 2e^c xn^{-1/3} \qquad\qquad \text{by (15)} \\
&= P_f\{\hat{g}_n^c(y_1+xn^{-1/3})\leq g^c(y_1)(1+xn^{-1/3})\} \\
&\geq 1-\exp\Big(-K_1\sqrt{\log n}\Big)
\end{aligned}$$

by (16) and (5). Similarly, one finds

$$\begin{aligned}
&P_f\bigg\{\Big(\frac{n}{4c}\Big)^{1/3}\bigg(\Big(\hat{f}_n^c\Big)^{1/3}\Big(\frac{\log(y_2-x'n^{-1/3})}{c}\Big)-f^{1/3}(x_2)\bigg)\geq -K_2(\log n)^{1/6}\bigg\} \\
&\qquad \geq 1-\exp\Big(-K_3\sqrt{\log n}\Big),
\end{aligned} \tag{18}$$

where $x' := (\log n)^{1/6}/f^{1/3}(x_2)$. Set $\varepsilon := C-(2c)^{1/3} > 0$, then $(n/(4c))^{1/3}\frac{3}{2}(f^{1/3}(x_2) - f^{1/3}(x_1)) - (\log(\sqrt{n}c/2))^{1/3} = (1/(4c))^{1/3}((2c)^{1/3} + \varepsilon)(\log n)^{1/3} - (1/2\log n + \log c/2)^{1/3} \geq \varepsilon/(8c)^{1/3}(\log n)^{1/3}$ for n large enough, depending only on c. Together with (17), (18) we obtain for every constant l,

$$\begin{aligned}
&P_f\bigg\{\Big(\frac{n}{4c}\Big)^{1/3}\frac{3}{2}\bigg(\Big(\hat{f}_n^c\Big)^{1/3}\Big(\frac{\log(y_2-x'n^{-1/3})}{c}\Big) \\
&\qquad -(\hat{f}_n^c)^{1/3}\Big(\frac{\log(y_1+xn^{-1/3})}{c}\Big)\bigg)-\Big(\log\frac{\sqrt{n}c}{2}\Big)^{1/3}>l\bigg\}\to 1,
\end{aligned}$$

the convergence not depending on f, y_1 or y_2. As $y_1 + xn^{-1/3} < y_2 - x'n^{-1/3}$ by (15) and the critical value $l_n(1-\alpha)$ is bounded by Theorem 2 and Fatou's lemma, one obtains $P_f(\phi_n(T_n) = 1) \to 1$ uniformly in $\{f \in H_c\colon \delta(f, H_0) \geq d_n\}$, proving (b).

Note that the crucial feature of the asymptotically minimax adaptive test lies in comparing $(n/(4c))^{1/3}T_n(c)$ to $(\log(\sqrt{n}c/2))^{1/3}$ across scales c. The exact choice of the denominator of the rescaling for the test statistic (cf. Theorem 2) is less important. The effort that went into deriving the rescaling sequence in the denominator in Theorem 2 reflects the desire to give equitable weights to all scales. □

PROOF OF THEOREM 5. To avoid technicalities the main arguments of the proof will be sketched. Denote by f and F the density and cdf of the X_i, respectively. The proof of Lemma 1 shows that if the cdf G has the same support as F and a density g that is positive and continuous in its interior, then $G(X_i - d)$ has Radon–Nikodym derivative $f(G^{-1}(t)+d)/g(G^{-1}(t))$, $t \in (0,1)$. If $G = F$, then the derivative is nonincreasing, again by the proof of Lemma 1. We will consider the least favorable case where F is the cdf of $U[0,1]$, and the above Radon–Nikodym derivative equals $1_{[0,1-d]}(t)$. To see how taking the sup over $d > 0$ is incorporated into the statement of Theorem 2, note that if $X_i > d$ then $F(X_i - d)$ is just a shift of X_i by an amount d. Thus the statistic looks at a subset of the same stretches of data considered in the context of Theorem 2, and this fact is readily incorporated into equation (4). When employing $\widetilde{F}_n$ in place of F note that the Radon–Nikodym derivative $f(\widetilde{F}_n^{-1}(t)+d)/\tilde{f}_n(\widetilde{F}_n^{-1}(t))$ differs from the nonincreasing function $f(\widetilde{F}_n^{-1}(t)+d)/f_n(\widetilde{F}_n^{-1}(t))$ by not more than $O((\log n/n)^{2/5})$ a.s. as $\| f - \tilde{f}_n \| = O((\log n/n)^{2/5})$ a.s. [and in the case of a general log-concave f using the fact that f is bounded above and away from 0 on $(F^{-1}(\varepsilon), F^{-1}(1-\varepsilon))$]. By Theorem 3(a) and the proof of Theorem 3(b), the statistic will not be sensitive to a perturbation of order $(\log n/n)^{2/5}$ if $c \gg (\log n/n)^{1/5}$. □

REFERENCES

BROOKS, S. P. (1998). MCMC convergence diagnosis via multivariate bounds on log-concave densities. *Ann. Statist.* **26** 398–433.

CHAUDHURI, P. and MARRON, J. S. (1999). SiZer for exploration of structures in curves. *J. Amer. Statist. Assoc.* **94** 807–823.

CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41.

CHOW, Y. S. and TEICHER, H. (1988). *Probability Theory*, 2nd ed. Springer, New York.

DELLAPORTAS, P. and SMITH, A. F. M. (1993). Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. *J. Roy. Statist. Soc. Ser. C* **42** 443–460.

DÜMBGEN, L. and SPOKOINY, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124–152.

GILKS, W. R. and WILD, P. (1992). Adaptive rejection sampling for Gibbs sampling. *J. Roy. Statist. Soc. Ser. C* **41** 337–348.

GLASER, R. E. (1980). Bathtub and related failure rate characteristics. *J. Amer. Statist. Assoc.* **75** 667–672.

GREEN, P. J. and SILVERMAN, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman and Hall, London.

GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. Le Cam and R. A. Olshen, eds.) **2** 539–555. Wadsworth, Belmont, CA.

GROENEBOOM, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory related Fields* **81** 79–109.

GROENEBOOM, P., HOOGHIEMSTRA, G. and LOPUAÄ, H. P. (1999). Asymptotic normality of the L1 error of the Grenander estimator. *Ann. Statist.* **27** 1316–1347.

GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). Estimation of a convex function: characterizations and asymptotic theory. *Ann. Statist.* To appear.

HILDENBRAND, K. and HILDENBRAND W. (1985). On the mean income effect: a data analysis of the U.K. family expenditure survey. In *Contributions to Mathematical Economics* (W. Hildenbrand and A. Mas-Colell, eds.) 247–263. North-Holland, Amsterdam.

INGSTER, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives I–III. *Math. Methods Statist.* **2** 85–114, 171–189, 249–337.

JONGBLOED, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *J. Comput. Graph. Statist.* **7** 310–321.

KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219.

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent r. v.'s and the sample d.f. *Z. Wahrsch. Verw. Gebiete* **32** 111–131.

KOROSTELEV, A. and NUSSBAUM, M. (1995). Density estimation in the uniform norm and white noise approximation. Preprint 154, Weierstrass Institute, Berlin.

LIERO, H., LÄUTER, H. and KONAKOV, V. (1998). Nonparametric versus parametric goodness of fit. *Statistics* **31** 115–149.

LINDSAY, B. G. and ROEDER, K. (1992). Residual diagnostics for mixture models. *J. Amer. Statist. Assoc.* **87** 785–794.

LINDSAY, B. G. and ROEDER, K. (1997). Moment-based oscillation properties of mixture models. *Ann. Statist.* **25** 378–386.

MILLER, R. G. (1981). *Survival Analysis*. Wiley, New York.

NEUMANN, M. H. (1998). Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. *Ann. Statist.* **26** 2014–2048.

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36.

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

ROEDER, K. (1994). A graphical technique for determining the number of components in a mixture of normals. *J. Amer. Statist. Assoc.* **89** 487–495.

SHEATHER, S. J. and JONES, M. C. (1991). A reliable data-based bandwidth selection method for kernel density-estimation. *J. Roy. Statist. Soc. Ser. B* **53** 683–690.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with applications to statistics*. Wiley, New York.

SILVERMAN, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *Ann. Statist* **10** 795–810.

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

TITTERINGTON, D. M., SMITH, A. F. M. and MAKOV, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. Wiley, New York.

WALTHER, G. (2000a). Multiscale maximum likelihood analysis of a semiparametric model, with applications. Technical report, Dept. Statistics, Stanford Univ.

WALTHER, G. (2000b). Detecting the presence of mixing with multiscale maximum likelihood. *J. Amer. Statist. Assoc.* To appear.

WOODROOFE, M. and SUN, J. (1999). Testing uniformity versus a monotone density. *Ann. Statist.* **27** 338–360.

ZHANG, Y. L. and NEWTON, M. A. (1997). On calculating the nonparametric maximum likelihood estimator of a distribution given interval censored data. *J. Comput. Graph. Statist.* To appear.

DEPARTMENT OF STATISTICS
390 SERRA MALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305
E-MAIL: walther@stat.stanford.edu