MULTISCALE MAXIMUM LIKELIHOOD ANALYSIS OF A SEMIPARAMETRIC MODEL, WITH APPLICATIONS

by

Guenther Walther

April 2001

Department of Statistics
STANFORD UNIVERSITY
Stanford, California
MULTISCALE MAXIMUM LIKELIHOOD ANALYSIS OF A SEMIPARAMETRIC MODEL, WITH APPLICATIONS

by

Guenther Walther
Department of Statistics
Stanford University

April 2001

This research was supported in part by
National Science Foundation grants DMS-9704557
and DMS-9875598

Department of Statistics
Sequoia Hall
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://www-stat.stanford.edu
Multiscale Maximum Likelihood Analysis of a Semiparametric Model, with Applications\textsuperscript{1}

Guenther Walther  
Department of Statistics, 390 Serra Mall  
Stanford University, Stanford, CA 94305  
email walther@stat.stanford.edu  
last revised March 2001

Abstract

A special semiparametric model for a univariate density is introduced that allows to analyze 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. This technical report is a detailed version of Walther (2001).

\textsuperscript{1}Research partially supported by NSF Grants DMS-9704557 and DMS-9875598.

\textit{AMS 1991 subject classifications.} Primary 62G10; secondary 60F15.

\textit{Key words and phrases.} Multiscale analysis, Grenander estimator, minimax, adaptive, mixture, log-concave, failure rate, penalized maximum likelihood

Abbreviated title: Multiscale Maximum Likelihood
1 Introduction and overview

This paper is concerned with the semiparametric model

\[ f(x) = \exp(\phi(x) + cx), \]  

(1)

where \( f \) is a probability density with support in \([0, 1]\), \( \phi \) is a nonincreasing function taking values in \( \mathbb{R} \), 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 \( X_1, \ldots, X_n \) from \( f \), the problem is to test whether \( c = 0 \), i.e. \( f \) 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 et al. 1988,Ch.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 of bathtub-shape, i.e. 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,p.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,p.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,p.493) and Titterington et al. (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 sampling literature, see Gilks and Wild (1992), Dellaportas and Smith (1993), and Brooks (1998). It models single-component distributions as logarithmically concave densities, i.e. densities of the form $f(x) = e^{\psi(x)}$, where $\psi$ 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 $\psi$ 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 logconcave 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 logconcave 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 $\hat{F}_n$ of the $X_i$, see Shorack and Wellner (1986,p.86). The transformation $(X_i - d) \mapsto \hat{F}_n(X_i - d)$ 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 $\frac{1}{2}\text{gamma}(2,0.1)+\frac{1}{2}\text{gamma}(5,0.15)$. 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 Woodroo and Sun (1999) for the link and for an approach to detect a global trend.
Figure 1: Left: \( \{X_i - d\}_{i=1}^{49} \) are plotted on the horizontal axis, where \( X_i \sim \gamma(10, 0.1), d = 0.2 \). 
\( \{F_i(X_i - d)\}_{i=1}^{49} \) 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).

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.

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 et al. (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 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 argmax 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 to construct 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

$$H_0: \ c = 0$$

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 already noted that in the case $c = 0$ the MLE of $f$ is the Grenander estimator based on the observations $X_i$, i.e. the left-hand slope of the least concave majorant of the empirical cdf of the $X_i$, see e.g. Robertson et al. (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

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

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

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 3-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 \( D(c) := \{ f : \log f(y) - \log f(x) \leq c(y - x) \text{ for all } x < y \} \). But Green and Silverman (1994, p.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 to 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 et al. (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 \( \inf_{m \in Mon} \| (\log f - m)w \|_\infty \), where \( Mon \) is the class of nonincreasing functions, and \( w \) is a weight function that allows to downweight 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 et al. (1985, p.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 \( \sup_{y < x} \frac{3}{2}(f^{1/3}(y) - f^{1/3}(x)) \). This leads to the test statistics

\[
T_n(c) = \sup_{X(1) \leq x < y \leq X(n)} \frac{3}{2}(\hat{f}_n^c)^{1/3}(y) - (\hat{f}_n^c)^{1/3}(x))
\]

\[
= \max_{1 \leq i < j \leq n} \frac{3}{2}c^{1/3}(\hat{g}_n^c(e^{cX_i})e^{cX_j})^{1/3} - (\hat{g}_n^c(e^{cX_{i+1}})e^{cX_i})^{1/3})
\]

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 $U_n(a) = \sup \{ t \geq 0 : G_n(t) - at \text{ is maximal} \}$, 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 argmax 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 $[1, c^c]$, we have $T_n(c) \leq \sup_{c \in [0,1]} 3(\int_{n}^{c} x^{1/3}) 1 - 1 \approx \sup_{y \in [1, c^c]} |c \hat{g}_n^c(y) y - 1|$. Formally switching arguments by taking the sup over the range of $\hat{g}_n^c$ instead of its domain gives $\sup_c |caU_n(a) - 1| = \sup_c ca[U_n(a) - (g^c)^{-1}(a)|. This heuristic shows that one needs to control the argmax 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

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

$V_c(a)$ 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 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

$$
\lim_{n \to \infty} \sup_{c \in \epsilon_n^{1/4}, \log_{10} n} \frac{(\frac{n}{4c})^{1/3} T_n(c) - (\log(\sqrt{n}/2))^{1/3}}{(\log(\sqrt{n}/2))^{-2/3} \log \log(\sqrt{n})} \leq L \ a.s.
$$
Figure 2: ‘Excursions under parabolas’ at various scales of the Brownian Bridge

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 $\alpha$. 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 $\hat{\theta}_n$ can be computed using standard algorithms such as the pool-adjacent-violators algorithm (PAVA), see Robertson et al. (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 setup 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.
But 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 $C(\log n)^{1/3}$ with $C < (2c)^{1/3}$: Any test with asymptotic level $\alpha$ has asymptotically a type II error of at least $1 - \alpha$ for some alternative of the described form, i.e. 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. $(2c)^{1/3}$ is called the exact separation constant, and $(\log n)^{1/3}$ the minimax rate of testing, see Ingster (1993).

**Theorem 3** The minimax rate of testing $H_0$ versus the semiparametric alternative $H_c$ is $(\log n)^{1/3}$ 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 $d_n = C(\log n)^{1/3}$ with $C < (2c)^{1/3}$, then

$$\lim_{n \to \infty} \inf_{\psi_n} \sup_{f \in H_c: \delta(f, H_0) \geq d_n} P_f(\psi_n(X_n) = 0) \geq 1 - \alpha,$$

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

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

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

Theorem 3 makes precise the adaptivity property of the test $\phi_n$: In a certain sense it performs as well as any level $\alpha$ 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 $X_1, \ldots, X_n$ 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 $W_i := \sum_{j=1}^{i} D_j / \sum_{k=1}^{n} D_k$, $i = 1, \ldots, n$ ($X_{(0)} := 0$), see Robertson et al. (1988,Ch.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 et al. (1988,Ch.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 non-increasing failure rate on \([0, \infty)\), provided only that the statistic \(T_{n-1}\) is computed with the \(W_i, 1 \leq i \leq n - 1\), 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, i.e. the \(X_i\) are iid \(E(1)\), the assertion is an immediate consequence of the fact that the joint law of \((W_1, \ldots, W_{n-1})\) is the same as that of the order statistics of \(n-1\) iid \(U[0, 1]\) random variables, see Shorack and Wellner (1986, Ch.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 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 changepoint \(t_0\) 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 discretizations did not change the results much. The CTTT statistic was evaluated against its limiting normal distribution, see Robertson et al. (1988, Ch.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 changepoint \(t_0\) the slope \(s\) was chosen so that the powers obtained in the simulation fell into a nontrivial range. The case \(t_0 = 0\) is taylormade 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, 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 to localize the changepoint \(t_0\) 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 greatly impair the performance of the CTTT statistic, while the MSML statistic is designed to handle such a case.
<table>
<thead>
<tr>
<th>sample size</th>
<th>$t_0 = 0, s = 0.1$</th>
<th>$t_0 = 2.5, s = 0.25$</th>
<th>$t_0 = 5, s = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSML</td>
<td>CTHTT</td>
<td>MSML</td>
</tr>
<tr>
<td>250</td>
<td>0.607</td>
<td>0.883</td>
<td>0.785</td>
</tr>
<tr>
<td>500</td>
<td>0.898</td>
<td>0.991</td>
<td>0.990</td>
</tr>
<tr>
<td>750</td>
<td>0.978</td>
<td>0.999</td>
<td>1.000</td>
</tr>
</tbody>
</table>

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

5.2 Detecting the presence of mixing. Assume the $X_i$ to be ordered and set $X'_i := \tilde{F}_n(X_i - d)s_d$, where the factor $s_d := \#\{i : X_i \geq X_1 + d\}/(n\tilde{F}_n(X_n - d))$ scales the $X'_i$ linearly into an interval with length equal to the fraction of nonzero $X'_i$. Analogously to section 3 set $T_n(c) := \max_{i < \epsilon < X'_{i+1}} \frac{1}{\epsilon} (\hat{g}_n(e^{cX'_i})e^{cX'_i})^{1/3} - (\hat{g}_n(e^{cX'_{i+1}}))e^{cX'_i})^{1/3}$, with the $\hat{g}_n$ 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 $\tilde{F}_n$ is the MLE under the null model. The logconcave MLE $\hat{f}_n$ can be readily computed using the iterative convex minorant algorithm (Jongbloed 1998), see Walther (2000). The theoretical properties of a logconcave MLE are similar to those of the MLE of a concave density, and the arguments in Groeneboom et al. (2000) suggest that the uniform rate of convergence is $O((\log n/n)^{2/5})$. If the scales $c$ are contained in the interval $[(\log^2 n/n)^{1/5}, n/\log^{10} n]$, then a result analogous to Theorem 2 holds for $T'_n$.

Theorem 5 Let $X_1, \ldots, X_n$ be iid from a logconcave distribution. Then under the assumptions stated prior to the theorem, the assertion 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 lab in the Genetics Department at Stanford. Flow cytometry uses light-induced fluorescence to measure certain characteristics of cells. One goal of the analysis is to detect the presence of sub-populations 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 $\mathbb{R}^d$ with common support $S$ and $p_i > 0$, $i = 1, \ldots, m$. Then on any compact subset of $S^o$ the representation

$$\sum_{i=1}^{m} p_i f_i(x) = \exp(\phi(x) + c \|x\|^2)$$

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

Thus the proposition leads to a second-order version of the semiparametric model (1), where $\phi$ 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

The following lemma will be used below:

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

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

**Proof:** One checks that $U_f(c, \cdot)$ is upper semicontinuous, so $\sup_{x \in B} |U_f(c, x) - x|$ is attained at some $x_0 \in B$. Set $u_0 := U_f(c, x_0) \in A$. We may assume $\epsilon := \|f - g\|_{\infty} > 0$ and consider first the case where $u_0 \geq x_0$.

One verifies that $x < u_0 - \sqrt{\epsilon/c}$ implies $-c(x - (x_0 + \sqrt{\epsilon/c}))^2 < -c(u_0 - (x_0 + \sqrt{\epsilon/c}))^2 + c(u_0 - x_0)^2 - c(x - x_0)^2 - 2\epsilon$. Together with $f(x) - c(x - x_0)^2 \leq f(u_0) - c(u_0 - x_0)^2$ for $x \in A$ we get for all $x \in (-\infty, u_0 - \sqrt{\epsilon/c}) \cap A$:

$$g(x) - c(x - (x_0 + \sqrt{\epsilon/c}))^2 \leq f(x) - c(x - (x_0 + \sqrt{\epsilon/c}))^2 + \epsilon$$

$$< f(u_0) - c(u_0 - (x_0 + \sqrt{\epsilon/c}))^2 - \epsilon$$

$$\leq g(u_0) - c(u_0 - (x_0 + \sqrt{\epsilon/c}))^2.$$

Hence $U_g(c, x_0 + \sqrt{\epsilon/c}) \geq u_0 - \sqrt{\epsilon/c}$. So if $x_0 + \sqrt{\epsilon/c} \in B$, then $\Omega_g(c) \geq u_0 - x_0 - 2\sqrt{\epsilon/c} = \Omega_f(c) - 2\sqrt{\epsilon/c}$. If $x_0 + \sqrt{\epsilon/c} \notin B$, then $\Omega_f(c) = u_0 - x_0 \leq \sqrt{\epsilon/c}$ as $u_0 \in A$ cannot be larger than the right endpoint of $B$. So again $\Omega_g(c) \geq \Omega_f(c) - 2\sqrt{\epsilon/c}$. The case $u_0 < x_0$ can be treated analogously. The lemma follows. □

**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 $F^{-1}(0) := \inf\{x : F(x) > 0\}$ and $F^{-1}(1) := \sup\{x : F(x) < 1\}$. Thus the cdf $F_d$ of $F(X - d)$ is given on $[0, 1]$ by $F_d(t) = P(X \leq F^{-1}(t) + d) = F(F^{-1}(t) + d)$, which is continuously differentiable on $(0, 1) \setminus \{F(F^{-1}(1) - d)\}$ 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

$$f_d(t) = f(F^{-1}(t) + d)/f(F^{-1}(t)), \quad t \in (0, 1).$$

(3)

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 $(F^{-1}(0), F^{-1}(1))$. But 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 \( \phi_1 \) 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 \( \phi \) is concave and \( c \geq 0 \). Thus by (3), \( \log f_{d}(t) = \phi(F^{-1}(t) + d) - \phi(F^{-1}(t)) + 2cdF^{-1}(t) + cd^2 \). The stated increment of \( \phi \) is nonincreasing in \( t \) as \( \phi \) 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. p. 326 in Robertson et al. 1988) shows that the MLE must be of the form

\[
\log \hat{f}_n(x) = \hat{\phi}(x) + cx, \tag{4}
\]

where \( \hat{\phi}(x) \) equals a constant \( \hat{\phi}_i \) on \( (x_{i-1}, x_i] \), \( i = 1, \ldots, n \), and \( \hat{\phi}(x) = -\infty \) for \( x \in (-\infty, x_0] \cup (x_n, \infty) \). Here \( x_i := X_{(i)}, i = 1, \ldots, n \), and \( x_0 := 0 \). Thus \( \hat{\phi} \) is given by the solution of the optimization problem

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

Ex. 1.5.7 of Robertson et al. (1988) shows that \( (\exp(\hat{\phi}_i), 1 \leq i \leq n) \) is the antitonic regression of \( (g_1, \ldots, g_n) \) with weights \( (w_1, \ldots, w_n) \), where \( w_i = \int_{x_{i-1}}^{x_i} e^{ct} \, dt = (e^{cx_i} - e^{cx_{i-1}})/c \) and \( g_i = 1/(nw_i), i = 1, \ldots, n \). Thm. 1.4.4 in Robertson et al. (1988), applied for antitone instead of isotope regression, gives \( \exp(\hat{\phi}_i) = \min_{x \leq x_{i-1}} \max_{x \geq x_i} \left( \sum_{j=s+1}^t w_j/\sum_{j=s+1}^t w_j \right) \). But the last fraction equals \( \frac{c}{n} (t-s)/(e^{cx_i} - e^{cx_s}) = c(G_n(y_1) - G_n(y_s))/(y_t - y_s) \), where \( y_i := e^{cx_i}, i = 0, \ldots, n \), and \( G_n \) is the empirical cdf of the \( (y_i, i = 1, \ldots, n) \). 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 now follows from (4). ☐

**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, p. 536) or Dümbgen and Spokoiny (2000, proof of Thm. 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 \( \mathbb{R} \)). For \( k \geq 4 \) define the rectangle \( R_k := [0, 1] \times [2^k, 2^{k+1}] \) and the lattice \( L_k := \{(i/(k2^k), 2^k + j2^k/k), 1 \leq i \leq k2^k, 0 \leq j \leq k - 1\} \subset R_k \). Now consider an arbitrary pair \((a, c^{2/3}) \in R_k \) and let \((\tilde{a}, c^{2/3}) \in L_k \) such that \( a \leq \tilde{a}, a - \tilde{a} \leq 1/(k2^k) \), and \( c \geq \tilde{c}, c^{2/3} - \tilde{c}^{2/3} \leq 2^k/k \). Then \( c^{2/3}/(k2^k) \leq 2/k \) and \((\tilde{c}/c)^{2/3} \geq \tilde{c}^{2/3}/(c^{2/3} + 2^k/k) \geq 1 - 1/k \). Hence for \( \lambda \in \mathbb{R} \) the inequality \( c^{2/3}(V_c(a) - a) \geq \lambda \) together with above monotonicity properties
of \( V_c(a) \) imply

\[
\tilde{c}^{2/3}|V_c(\tilde{a}) - \tilde{a}| \geq \tilde{c}^{2/3}(V_c(\tilde{a}) - \tilde{a}) \\
\geq \tilde{c}^{2/3}(V_c(a) - \tilde{a}) \\
\geq (\tilde{c}/c)^{2/3}c^{2/3}(V_c(a) - a - 1/(k2^k)) \\
\geq (1 - 1/k)(\lambda - 2/k).
\]

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 \( 2^k \leq c^{2/3} \leq 2^{k+1} \) entails \((3/2 \log c^{2/3})^{1/3} + 16(3/2 \log c^{2/3})^{-2/3} \log \log c \geq (3/2 \log 2^k)^{1/3} + 6(3/2 \log 2^k)^{-2/3} \log \log 2^k + 2/k =: \lambda_k \). Hence \( A_k \) implies \( \tilde{c}^{2/3}V_c(\tilde{a}) - \tilde{a} | \geq (1 - 1/k)(\lambda_k - 2/k) \) for some \((\tilde{a}, c^{2/3}) \in L_k\).

It is helpful now to take for \( Y \) two-sided Brownian motion originating from the origin. Using Brownian scaling one sees that \( L(c^{2/3}V_c(a)) = L(V_1(ac^{2/3})) \). The process \( a \mapsto V_1(a) - a \) is stationary and the tail behavior of its marginal density is given by \( f_Z(t) \sim 4^{t^3/2|t|} \exp(-2/3|t|^3 + 2^{1/3}a_1|t|)/A \) (as \( |t| \to \infty \)), where \( a_1 \approx -2.3381 \) and \( A \approx 0.7022 \), see Groeneboom (1989, Cor. 3.4). Let the random variable \( Z \) have density \( f_Z \). Then for \( L \) large enough \( P(|Z| > L) \leq \int_{-L}^{L} C t \exp(-2/3t^3) dt \leq C \int_{-L}^{L} t^2 / L \exp(-2/3t^3) dt = C/(2L) \exp(-2/3L^3) \) for some constant \( C \). Together with \# \( L_k = k^22^k \) one obtains for \( k \) large enough:

\[
P(A_k) \leq \frac{Ck^22^k}{(1 - 1/k)(\lambda_k - 2/k)} \exp\left(-2/3(1 - 1/k)^3(\lambda_k - 2/k)^3\right) \\
\leq \frac{Ck^22^k}{(3/2 \log 2^k)^{1/3}} \exp\left(-2/3(1 - 3/k)(3/2 \log 2^k + 18 \log \log 2^k)\right) \\
\leq \frac{Ck^22^k}{(3/2 \log 2^k)^{1/3}} \exp\left(\log 2^{-k+3} - 3 \log \log 2^k\right) \\
\leq \frac{Ck^2}{(3/2 \log 2^k)^{10/3}} \\
\leq C/k^{4/3}.
\]

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 \( \sup_{a \leq c < C} \sup_{a \in [0,1]} c^{2/3}|V_c(a) - a| < \infty \) for all \( C > e^a \) by the monotonicity properties of \( V_c(a) \)).
Now define \( \tilde{V}_c(a) := \arg \max_{t \geq 0} \{W(t) - c(t - a)^2\} \), 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 \( \tilde{V}_c(a) = V_c(a) \) 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 \( \tilde{V}_c(a) \) 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 \([\arg \max_{t \geq a} \{W(t) - c(t - a)^2\} - a > L]\) and \([\arg \max_{t \leq a} \{W(t) - c(t - a)^2\} - a < -L]\) 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 \( a \geq 0 \) one finds \( P(|\tilde{V}_c(a) - a| > L) \leq 2p(L) \leq 2\sqrt{P(|\tilde{V}_c(a) - a| > L)} \), and thus \( P(c^{2/3}|\tilde{V}_c(a) - a| > L) \leq C/\sqrt{L} \exp(-\frac{1}{2}L^3) \) for large enough \( L \). The additional factor \( \frac{1}{2} \) in the exponent is compensated for by the fact that \( \#(L_k \cap [0,2^{-k/2}] \times [2^k,2^{k+1}]) \leq k^{2/2}k' \), which gives

\[
P\left( \frac{c^{2/3}}{\frac{3}{2} \log c^{2/3}} - \frac{\frac{3}{2} \log c^{2/3}}{\frac{3}{2} \log c^{2/3}} \right) > 32 \text{ for some } (a,c^2) \in [0,2^{-k}] \times [2^{k},2^{k+1}) \leq C/k^{4/3}.
\]

Finally, for a Brownian Bridge \( B(t) = W(t)-tW(1), \) 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 \( V_c^W(a - W(1)/(2c)) \leq V_c^W(a) \) for \( 0 \leq a \leq W(1)/(2c) \), and then sup \( a \in [0,1] \) \( c^{2/3} |V_c^B(a) - a| \leq \sup_{a \in [1,1+|W(1)/(2c)|]} c^{2/3} |V_c^W(a) - a| + |W(1)|/c^{1/3} \). The claim for the Brownian Bridge now follows because the theorem holds for \( V_c^W \): The restriction of the argmax to \([0,1]\) can be dealt with just as in the step from two-sided to one-sided Brownian motion above. \( \Box \)

**Proof of Theorem 2:** We will make use of the following extension of Theorem 1, which follows from the law of the iterated logarithm:

**Corollary 1** Let \( Y \) be as in Theorem 1 and \( p \geq 0 \). Then

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

**Proof of Corollary 1:** We will show that a.s.

\[
\sup_{a \in [-p,0]} c^{2/3} |V_c(a) - a| \leq (\log c)^{1/3} \vee pc^{2/3} + K(\log c)^{-2/3} \log \log c \tag{5}
\]

for all \( c \) large enough, where the random variable \( K \) equals the finite quantity asserted by Theorem 1. The proof for \( \sup_{a \in [1,1+p]} \) is analogous.
Set \( g(t) := \sqrt{2t \log \log(1/t)} \) and for \( a < 0 \) define
\[
r_a := \begin{cases} 
\frac{(\log c)^{1/3} + K(\log c)^{-2/3} \log \log c}{\sqrt{2/3}} - |a| & \text{if } |a| \leq (\log c)^{1/3}/c^{2/3} \\
K(\log c)^{-3/3} \log \log c & \text{if } |a| > (\log c)^{1/3}/c^{2/3}.
\end{cases}
\]

By the law of the iterated logarithm there exists a righthand neighborhood of 0, \( N = N(\omega) \), where \( Y(t) \geq 2g(t) \). We can restrict our analysis to \( N \), because \( V_c(a) \leq V_c(0) \downarrow 0 \) as \( c \to \infty \), uniformly in \( a \). Likewise, \( r_a \) will fall into \( N \) for large enough \( c \) (not depending on \( a \)). Note that \( V_c(a) \) is the location of the point where the parabola \( c(t - a)^2 + b \), sliding down along the line \( t = a \), hits \( Y \). Set \( f_{a, c}(t) = c(t - a)^2 - ca^2 \), so \( f_{a, c}(0) = 0 = g(0+) \) and \( f'_{a, c}(0) < \infty = g'(0+) \). Hence \( f_{a, c} < \frac{1}{2}g \) in a righthand neighborhood of 0, and so for all \( a < 0, c > 0 \) there exists a \( t \in (0, r_a) \) with \( f_{a, c}(t) < Y(t) \), again by the LIL. One verifies that \( f_{a, c}(r_a) > 2g(r_a) \) for large enough \( c \) depending only on \( K \). (Use \( f_{a, c}(r_a) > \max(r_a^2, 2cr_a|a|) \) and distinguish the three cases where \( |a| \) falls between \( 0, 4(\log c)^{1/3}/(\sqrt{K}c^{2/3}), (\log c)^{1/3}/c^{2/3} \) and \( \infty \).) But then \( f_{a, c} > Y \) on \( N \cap [r_a, \infty) \) as \( f_{a, c} \) is strictly convex and \( g \) is strictly concave. It follows that \( V_c(a) < r_a \) for \( c \) large enough, and (5) entails. \( \square \)

To fix notation, note that if the \( X_i, i = 1, \ldots, n \) are iid \( f \) with \( cdf \) \( F \) on \( [0, 1] \), then the \( Y_i := e^{cX_i} \) have density \( g^c(y) = f(\log y)/cy \) with \( cdf \) \( G^c(y) = F(\log y/c) \) 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.

Now let first \( f \) 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 \( \hat{g}_n^c \):
\[
\lim_{n \to \infty} \sup_{c \in \{e^{-\frac{3}{2} \log_2(n)}, \ldots, e^{\frac{3}{2} \log_2(n)}\}} \sup_{y \in D_c} \frac{(\sqrt{nc^2/2})^{3/2} y(\hat{g}_n^c(y) - g^c(y)) - (\log(\sqrt{nc^2/2}))^{1/3}}{(\log(\sqrt{nc^2/2}))^{-2/3} \log \log(\sqrt{nc})} \leq K \quad \text{a.s.} \tag{6}
\]

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 \( l_{c,n} = 8(K \vee 1) \log(\sqrt{nc})\left(\frac{e}{n \log^2(\sqrt{nc})}\right)^{1/3} \). The term \( l_{c,n} \) is necessary due to the fact that the Grenander estimator has an upward bias near the left endpoint of its support, see Woodroofe and Sun (1993). The assertion of the theorem follows from \( 3\left(\left(\hat{f}_n^c \log y/c\right)^{1/3} - 1\right) \leq \hat{f}_n^c(\log y/c) - 1 = cy(\hat{g}_n^c(y) - g^c(y)) \) together with (6), and by proceeding similarly with \( 3\left(1 - (\hat{f}_n^c)^{1/3}\right) \) using the companion result to (6) for \( -(\hat{g}_n^c(y) - g^c(y)) \). The nuisance of \( \hat{g}_n^c(1+) \) overestimating \( g^c(1+) \) does not affect this result because we consider a difference of \( \hat{f}_n^c \); this can be verified with a simple calculation using a property similar to (2) in Woodroofe and Sun (1993).
To prove (6), note that $G^c(y) = \frac{\log y}{c}$ on $D_c$ and $g^c(y) = \frac{1}{cy}$ has range $R_c := [\frac{1}{c}, \frac{1}{c^2}]$. For $a > 0$ define

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

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

$$\hat{U}_n(a) := G^c(U_n(a)) = \sup\{u \in [0,1] : IU_n(u) + \sqrt{n}(u - aG^{c^{-1}}(u)) \text{ is maximal}\}, \quad (7)$$

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

$$\hat{g}^c_n(y) \leq z \iff U_n(z) \leq y \text{ a.e. } (y, z), \quad (8)$$

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

$$\left(\frac{\sqrt{n}/c}{2}\right)^{2/3} y \left(\hat{g}^c_n(y) - \frac{1}{cy}\right) \leq x \iff \hat{g}^c_n(y) \leq \frac{1}{cy} + \frac{xn^{-1/3}(c/2)^{-2/3}}{y}$$
$$\iff U_n\left(\frac{1}{cy} + \frac{xn^{-1/3}(c/2)^{-2/3}}{y}\right) \leq y \text{ a.e. } y \text{ by (8)}$$
$$\iff U_n(a + ax(4c/n)^{1/3}) \leq \frac{1}{ac}$$
$$\iff \hat{U}_n(a(1 + x(4c/n)^{1/3})) \leq G^c\left(\frac{1}{ac}\right) \text{ by (7)}$$
$$\iff \hat{U}_n(a(1 + x(4c/n)^{1/3})) - u_a \leq \frac{1}{c} \log((1 + x(4c/n)^{1/3})$$

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 (6) will follow once we show that a.s.

$$\sup_{a,ac \in D_{c,n}} \left(\hat{U}_n(a(1 + x(4c/n)^{1/3})) - u_a\right) \leq \frac{1}{c} \log(1 + x(4c/n)^{1/3}) \text{ for all } c \in [\epsilon_0 n^{-1/2}, \frac{n}{\log^3 n}] \quad (9)$$

for $n$ large enough.

The plan is to show (9) by applying Theorem 1 and Corollary 1 to an approximation to $\hat{U}_n$ 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 \( \tilde{U}_n(a) := \sup\{u \in [0, 1] : B_n(u) - \sqrt{n} \xi_u(u - u_a)^2 \text{ is maximal}\} \), where \((B_n, n \in I\mathbb{N})\) is a sequence of Brownian Bridges constructed on the same probability space as \(I U_n\) such that

\[
\| I U_n - B_n \|_\infty = O\left(\frac{\log^2 n}{\sqrt{n}}\right) \quad \text{a.s.}
\]  

(10)

see Komlos et al. (1975). (Here the \(O(\cdot)\) a.s. notation means that the ratio of lefthand side and righthand side is bounded above by a constant that may depend on the realization \(I U_n\) and \(B_n\).) \((ac)^{-1} \in D_{c,n}\) implies \(u_a < 1\). It is helpful to let \(a\) vary over a larger set such that the range of \(u_a\) extends to 1. Define \(R_{c,n} := \{a \leq c^{-1} \left[1 + 8(K \lor 1) \log \log(\sqrt{n}c) \left(\frac{\log^2 \left(\frac{c}{\log(\sqrt{n}c)}\right)}{\log^2 \left(\frac{\sqrt{n}c}{2}\right)}\right)^{1/3}\right]^{-1} : u_a \leq 1\}\). While \(a \in R_{c,n}\) can be so large that \(u_a < 0\), one verifies that in that case \((\sqrt{n}c/2)^{2/3}\left|\bar{U}_n(a)\right| \leq (\log(\sqrt{n}c/2))^{1/3}\) for \(n\) large enough (depending only on \(K\)). So Corollary 1 gives a.s.

\[
\sup_{a \in R_{c,n}} (\sqrt{n}c/2)^{2/3}\left|\bar{U}_n(a) - u_a\right| \leq (\log(\sqrt{n}c/2))^{1/3} + K(\log(\sqrt{n}c/2))^{-2/3} \log \log(\sqrt{n}c)
\]  

(11)

for all \(c \in [e^6n^{-1/2}, n/\log^{10} n]\), if \(n\) large enough (depending only on \(K\)).

Now we account for the error incurred by employing \(\tilde{U}_n\) instead of \(\bar{U}_n\). Retracing the proof of Theorem 2.1 in Groeneboom et al. (1999) for our special case of the density \(g^c\) and employing some straightforward improvements for this case, one finds \(P(\eta^{1/3}|U_n(a) - \frac{1}{ac}| > z) \leq 2\exp(-\frac{z^2}{24c^2(\bar{a}c)^{-2}})\). The first Borel-Cantelli lemma yields

\[
|U_n(a) - (ac)^{-1}| = O\left(\frac{(\log n)^{1/3}}{a(c\sqrt{n})^{2/3}}\right) \quad \text{a.s.},
\]  

(12)

uniformly in \(c > 0\) and \(a \in R_c\). As \(G^c\) is concave we have \(|G^c(U_n(a)) - G^c((ac)^{-1})| \leq g^c(\min(U_n(a), (ac)^{-1})) \cdot |U_n(a) - (ac)^{-1}| = \frac{|U_n(a) - (ac)^{-1}|}{a^{-1} - (ac)^{-1} - U_n(a)|}\). Together with (12) we get

\[
|\tilde{U}_n(a) - G^c((ac)^{-1})| = O\left(\frac{(\log n)^{1/3}}{(c\sqrt{n})^{2/3}}\right) \quad \text{a.s., uniformly in } c \in (0, n/2^2 n] \text{ and } a \in R_c.
\]  

We now plug into this result \(\bar{a} := a(1 + x(4c/n)^{1/3})\) instead of \(a\) to obtain (note that \(G^c((\bar{a}c)^{-1}) = \bar{U}_n\))

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

(13)

uniformly in \(c \in [e^6n^{-1/2}, n/\log^{2} n]\) and \(a \in R_{c,n}\). This immediate for those \(a \in R_{c,n}\) for which \(\bar{a} \in R_c\), or equivalently, \(u_a \geq 0\). We already noted that if \(a \in R_{c,n}\) is such that \(u_a < 0\), then \((\sqrt{n}c/2)^{2/3}\left|u_a\right| \leq (\log(\sqrt{n}c/2))^{1/3} \leq 2(\log n)^{1/3}\). (13) then follows from the fact that \(\tilde{U}_n(\cdot)\) is nonincreasing.
On $[0, 1]$, $G^{c^{-1}}(u) = e^{cu}$. A Taylor series expansion gives $\sqrt{n}(u - a(1 + x(4c/n)^{1/3}))e^{cu} = -\sqrt{n}c/2(u - u_a)^2 - \sqrt{n}c/6 e^{-\xi}(u - u_a)^3 +$ terms not involving $u$, where $\xi$ lies between $0$ and $u - u_a$. For $u$ in the neighborhood of $u_a$ given by (13), the term $\sqrt{n}c/6 e^{-\xi}(u - u_a)^3$ is $O(\log n)$, uniformly in $c \in (0, n/\log n)$ and $a \in R_{c,n}$. Together with (7), (13) and the fact that $\arg \max_{u \in [0,1]} \{J_U(u) - \sqrt{n}c/2(u - u_a)^2\}$ also falls into the neighborhood of $u_a$ given by (13), one concludes that one can write

$$\bar{U}_n(a(1 + x(4c/n)^{1/3})) = \sup\{u \in [0, 1] : \bar{U}_n(u) - \sqrt{n}c/2(u - u_a)^2 + d_{n,a,c}(u) \text{ is maximal}\}, \quad (14)$$

where $\|d_{n,a,c}\|_\infty = O(\frac{\log n}{\sqrt{n}})$ a.s., uniformly in $c \in [e^{\epsilon}n^{-1/2}, n/\log^2 n]$ and $a \in R_{c,n}$. (10), (14) and Lemma 2 now yield

$$\left| \sup_{a \in R_{c,n}} |\bar{U}_n(a(1 + x(4c/n)^{1/3})) - u_a| - \sup_{a \in R_{c,n}} |\bar{U}_n(a) - u_a| \right| = O\left(\frac{\log^2 n}{nc}\right) \text{ a.s.}$$

uniformly in $c \in [e^{\epsilon}n^{-1/2}, n/\log^2 n]$. (9) and hence (6) now follow with (11) upon observing that $(ac)^{-1} \in D_{c,n}$ implies $a \in R_{c,n}$, $1/\epsilon \log(1 + x(4c/n)^{1/3}) \geq \frac{1}{\epsilon}(x(4c/n)^{1/3} - \frac{1}{2}x^2(4c/n)^{2/3}) \geq (\sqrt{n}c/2)^{-2/3}[x - \frac{1}{2}(\log(\sqrt{n}c/2))^{-2/3} \log(\sqrt{n}c)]$, and $O\left(\frac{\log^2 n}{nc}\right) \leq (\sqrt{n}c/2)^{-2/3}(\frac{1}{2^\delta}(\log(\sqrt{n}c/2))^{-2/3} \log(\sqrt{n}c))$ for $c \leq \frac{n}{\log^2 n}$ and $n$ large enough.

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 $f_{j,k} = f_0 + \phi_j + \psi_k, 1 \leq j, k \leq m$, where $\phi_j$, $\psi_k$ 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 $[b, b^']$. Now define for $j \leq m := \frac{1}{2b}$: $\phi_j(x) := \phi(x - \frac{1}{2} - (j - 1)b)$ and $\psi_j(x) := -\phi(x - (j - 1)b)$. So for $n$ large, $\phi_j$ looks like $c(x - \frac{1}{2} - (j - \frac{1}{2})b)$ on $[\frac{1}{2} + (j - \frac{1}{2})b, \frac{3}{2} + j\frac{b}{2}]$, and $\psi_j$ looks like $c(x - (j - \frac{1}{2})b)$ on $[(j - 1)b, (j - \frac{1}{2})b]$. There are $m = m(c, n) \approx (n/\log n)^{1/3} \delta$'s supported on $[0, 1]$. The dependence of $\phi_j, \psi_j, f_{j,k}, b$ and $m$ on $c$ and $n$ will be suppressed. One readily checks that $\int_0^b \phi = 0$ and $(\phi(b) + 1)^{1/3} - (\phi(0) + 1)^{1/3} \geq 2/3d_n$, whence $f_{j,k} \in H_c$ and $\delta(f_{j,k}, H_0) \geq d_n$ for $1 \leq j, k \leq m$. We will show

$$\frac{1}{m^2} \sum_{1 \leq j, k \leq m} e^{\Lambda_{j,k}} \to 1 \text{ in } P_0^n \text{-probability as } n \to \infty, \quad (15)$$

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

\[
\sup_{f \in H_{\epsilon,k}} P^n_f(\tau_n = 0) \geq \frac{1}{m^2} \sum_{1 \leq j, k \leq m} P^n_{j,k}(\tau_n = 0)
\]

\[
\geq E_0'(1(\tau_n = 1) + \frac{1}{m^2} \sum_{1 \leq j, k \leq m} e^{\Lambda_{n,j,k}}1(\tau_n = 0)) - \alpha_n
\]

\[
\geq (1 - \epsilon) P^n_0\left(\frac{1}{m^2} \sum_{1 \leq j, k \leq m} e^{\Lambda_{n,j,k}} \geq 1 - \epsilon\right) - \alpha_n.
\]

So assertion (a) of the theorem follows from (15).

Set \( S^n_j := \sum_{i=1}^n (\phi_j(X_i) - E_0\phi_j(X_i)) - \frac{1}{2} \sum_{i=1}^n E_0\phi_j^2(X_i) \) and define \( T^n_j \) analogously with \( \phi_j \) replaced by \( \psi_j \). We will prove (15) by showing

\[
E_0^n|e^{\Lambda_{n,j,k}} - e^{S^n_j + T^n_k}| \to 0 \text{ uniformly in } j, k. \tag{16}
\]

\[
\frac{1}{m^2} \sum_{1 \leq j, k \leq m} e^{S^n_j + T^n_k} \to 1 \text{ in } P^n_0\text{-probability.} \tag{17}
\]

As for (16), 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

\[
\sup_n E^n_{j,k} \exp\left(2 \sum_{i=1}^n Y_i\right) = \sup_n \int_0^\infty P^n_{j,k}(2 \sum_{i=1}^n Y_i \geq \log t)dt
\]

\[
\leq 1 + \sup_n \int_1^\infty \exp\left\{-2(\log t)^2/(144nd_n^4)\right\}dt < \infty
\]

as \( \sup_n nd_n^4 < \infty \). Further, \( |E_{j,k}(\phi_j(X_i) + \psi_k(X_i))^2 - E_0\phi_j^2(X_i) - E_0\psi_k^2(X_i)| = |E_0(\phi_j(X_i) + \psi_k(X_i))^3| \leq \frac{3}{c} d_n^4 \), and \( E_{j,k}R_i \leq 2d_n^4 \). Hence

\[
\sup_n E^n_{j,k}(e^{S^n_j + T^n_k - \Lambda^n_{j,k}})^2 = \sup_n E^n_{j,k} \exp\left\{2 \sum_{i=1}^n Y_i\right\} \exp\left\{\sum_{i=1}^n (E_{j,k}(\phi_j(X_i)
\]

\[
+ \psi_k(X_i))^2 - E_0\phi_j^2(X_i) - E_0\psi_k^2(X_i) - 2E_{j,k}R_i\right\} < \infty \tag{18}
\]

(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, \( Var^n_{j,k}(S^n_j + T^n_k - \Lambda^n_{j,k}) = Var^n_{j,k}(\sum_{i=1}^n Y_i) \leq 36nd_n^4 \to 0 \), as \( |Y_i| \leq 6d_n^2 \). Hence \( S^n_j + T^n_k - \Lambda^n_{j,k} \to 0 \) in \( P^n_{j,k}\)-probability uniformly in \( j, k \). Together with the uniform integrability condition (18), (16) follows by a standard argument, and by employing \( E_0^n|e^{\Lambda_{n,j,k}} - e^{S^n_j + T^n_k}| = E^n_{j,k}|1 - e^{S^n_j + T^n_k - \Lambda^n_{j,k}}| \).
To prove (17) we will show $\frac{1}{m} \sum_{j=1}^{m} e^{S_{j}} \rightarrow 1$ 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 $\lfloor m/n^{\epsilon} \rfloor \approx n^{1/3-\epsilon}/(\log n)^{1/3}$ intervals $I_{k}^{n}$, $k = 1, \ldots, \lfloor m/n^{\epsilon} \rfloor$, of equal length, where the small $\epsilon > 0$ will be specified later. Denote by $\nu_{k}$ the random number of $X_{i}$'s that fall into $I_{k}^{n}$. A standard calculation shows that the event $A_{n} := \left| |\nu_{k} - \frac{n^{\epsilon}}{2m} n| \leq \frac{m^{\epsilon}}{2} n n^{\epsilon/3} \right.$ for all $k = 1, \ldots, \lfloor m/n^{\epsilon} \rfloor$] satisfies $\lim_{n \to \infty} P_{0}^{n}(A_{n}) = 1$. Under $P_{0}^{n}$ and conditional on the vector $\nu$, the collections of $X_{i}$'s pertaining to different $I_{k}^{n}$ are independent, and within each $I_{k}^{n}$ the $X_{i}$'s are iid uniform. Each $I_{k}^{n}$ contains the supports of a block of $\approx n^{\epsilon}$ consecutive $\phi_{j}$'s. Thus the random variables $\frac{1}{n^{\epsilon}} \sum_{j \in \text{block } k} e^{S_{j}}, k = 1, \ldots, \lfloor m/n^{\epsilon} \rfloor$, are $P_{0}^{n}$-conditionally independent given $\nu$. The assertion will follow once we prove $\frac{1}{n^{\epsilon}} \sum_{j \in \text{block } k} e^{S_{j}} \rightarrow 1$ in conditional $P_{0}^{n}$-probability given $\nu$, uniformly in $\nu \in A_{n}$. The proof of Cor. 10.1.2 in Chow-Teicher (1988) shows that it is enough to prove

$$E_{0}^{n}(e^{S_{j}}|\nu) \rightarrow 1 \quad \text{uniformly in } j \in \{1, \ldots, m\} \text{ and in } \nu \in A_{n}, \quad (19)$$

$$E_{0}^{n}\left(\frac{1}{n^{\epsilon}} \sum_{j \in \text{block } k} e^{S_{j}} \cdot 1\left(\sum_{j \in \text{block } k} e^{S_{j}} \geq \frac{m}{n^{\epsilon}} \delta\right)|\nu\right) \rightarrow 0 \quad \text{uniformly in } \quad k \in \{1, \ldots, \lfloor m/n^{\epsilon} \rfloor\} \text{ and in } \nu \in A_{n}, \text{ for every } \delta > 0. \quad (20)$$

**Lemma 3** Let $t \geq 0$. Then

$$E_{0}^{n}(e^{tS_{j}}|\nu) = \exp\left(\frac{C^{3}}{6n^{\epsilon}} \log n \cdot (t^{2} - t) + o(1)\right),$$

uniformly in $j \in \{1, \ldots, m\}$ and in $\nu \in A_{n}$. Here $C$ is the constant given in the statement of the theorem.

**Proof:** Clearly $E_{0}^{n}(\phi_{j}^{2}(X_{1}) = O(d_{n}^{p+1})$, and if the index $j$ falls into block $k$, then $E_{0}^{n}(\phi_{j}^{2}(X_{1})|\nu) = \frac{\nu_{k}}{|I_{k}^{n}|} E_{0}^{n}(\phi_{j}^{2}(X_{1}), p \geq 1$. So a Taylor expansion gives $E_{0}^{n}(\exp(t\phi_{j}(X_{1}))|\nu) = 1 + \frac{\nu_{k}}{|I_{k}^{n}|} (t E_{0}^{n}(\phi_{j}(X_{1}) + \frac{t^{2}}{2} E_{0}^{n}(\phi_{j}^{2}(X_{1}) + O(d_{n}^{1})).$ Thus for $\nu \in A_{n}$

$$E_{0}^{n}(e^{tS_{j}}|\nu) = \left(E_{0}^{n}(e^{t\phi_{j}(X_{1})}|\nu)\right)^{n} \cdot \exp\left(-nt E_{0}^{n}(\phi_{j}(X_{1}) - \frac{nt}{2} E_{0}^{n}(\phi_{j}^{2}(X_{1}))\right)$$

$$\leq \max_{\nu \in A_{n}} \exp\left\{ n \frac{\nu_{k}}{|I_{k}^{n}|} \left(t E_{0}^{n}(\phi_{j}(X_{1}) + \frac{t^{2}}{2} E_{0}^{n}(\phi_{j}^{2}(X_{1}) + O(d_{n}^{1})\right)$$

$$- n t E_{0}^{n}(\phi_{j}(X_{1}) - \frac{nt}{2} E_{0}^{n}(\phi_{j}^{2}(X_{1})\right) \right\}$$

22
where we used $1 + x \leq e^x$. \( |F_k| = \frac{m}{2m} \), so \( \max_{x \in A_k} \frac{\nu_k}{|F_k|} \leq n + n^{2/3-\epsilon/6}/(\log n)^{1/6} \). Using $E_0 \phi_d^\circ(X_1) = O((\log n/n)^{(p+1)/3})$, $p \geq 1$, shows that the above max is not larger than \( \exp \left\{ \frac{1}{2} (t^2 - t) E_0 \phi_d^\circ(X_1) + o(1) \right\} \). One checks $E_0 \phi_d^\circ(X_1) = \frac{1}{3c} d_n^2 + O(d_n^4) = \frac{C^3}{3c} \cdot \frac{\log n}{n} + o(\frac{1}{n})$. The lower bound follows quite analogously via the inequality $1 + x \geq e^{x-1/2x^2}$, $x \geq 0$. The result is uniform due to the conditioning. \( \square \)

Now (19) follows by setting $t = 1$ in the lemma. Let $j$ be an index in block $k$. The expression in (20) is not larger than $E^n_0 \left( e^{S_j^k \cdot 1(e^{S_j^k} > \frac{m}{n} \delta \text{ for some } l \in \text{block } k)} \right)$

$$\leq \sum_{l \in \text{block } k} E^n_0 \left( e^{S_j^k \cdot 1(e^{S_j^k} > \frac{m}{n} \delta)} \right)$$

$$\leq \sum_{l \in \text{block } k} \frac{E^n_0(e^{S_j^k} e^{S_j^l} | \nu)}{(\frac{m}{n} \delta)^t} \quad \text{for any } t > 0.$$

As the conditional $P^n_{\nu}$-distribution of $e^{S_j^k}$ is the same as that of $e^{S_j^l}$, we get $E^n_0(e^{S_j^k} e^{S_j^l} | \nu) \leq E^n_0(e^{(1+t)S_j^l} | \nu)$ by Cauchy-Schwarz. Setting $t = \sqrt{2c/\lambda^3} - 1 > 0$ and applying the Lemma shows that the above sum is not larger than

$$n^\epsilon \frac{\exp \left( \frac{C^3}{3c} \log n \cdot (t+1)t \right)}{n(1/3-\epsilon)t} \cdot (\log n)^{t/3} \cdot O(1)$$

$$= n(\sqrt{C^3/(2c-1)}t^{1/3} (1+\epsilon)^t) \cdot (\log n)^{t/3} \cdot O(1)$$

$$\rightarrow 0$$

for $\epsilon$ small enough, as $\sqrt{C^3/(2c)} - 1 < 0$. The proof of (a) is complete.

For the proof of part (b) we start with the following lemma:

**Lemma 4** Let $g$ be a nonincreasing density on $[1, z]$ and set $g^{-1}(a) := \inf\{y : g(y) < a\}$ and $U_n(a) := \sup\{y \geq 1 : G_n(y) - ay \text{ is maximal}\}$. Assume there exist $y_1 \in (1, z)$ and $t > 0$ such that

$$g(y_1) > 0 \text{ and } g(y) \geq g(y_1) \frac{y_1}{y} \text{ for } y \in [y_1(1-t), y_1 + t] \subset [1, z]. \quad (21)$$

Then for all $x \in [0, n^{1/3}t(1+z)]$ and $a := g(y_1)(1 + xn^{-1/3})$ the inequality

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

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

23
The lemma is a generalization of Theorem 2.1 of Groeneboom et al. (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 inequality requires condition (21) and the link between $a$ and $x$. The bounded range for $x$ substitutes for the bounded range of $g$ in Groeneboom et al. (1996).

**Proof of Lemma 4:** We will employ the exponential bound for reverse time martingales used in the proof of Theorem 2.1 in Groeneboom et al. (1999). Let $x \in (0, n^{1/3} \varepsilon]$ and set $y_0 := g^{-1}(a)$. (21) gives $g(y_1(1 - x n^{-1/3})) \geq g(y_1)/(1 - x n^{-1/3}) \geq g(y_1)(1 + x n^{-1/3})$, whence $y_0 \geq y_1(1 - x n^{-1/3})$. So if $\delta_n := y_1 + x n^{-1/3} - y_0$, then $\delta_n n^{1/3} \leq x(1 + y_1) \leq x(1 + z)$. Writing $G(a, b) := \int_a^b g(y) \, dy$ we have $v_n := a \delta_n/G(y_0, y_0 + \delta_n) > 1$ as $a > g(y_1)$. Also $v_n \leq a/g(y_0 + \delta_n) \leq g(y_1)/(y_1 + x n^{-1/3}) \leq (1 + t^2)$, as $g$ is nonincreasing and because of (21) and $y_1 \geq 1$.

The arguments of Lemmata 2.1 and 2.3 in Groeneboom et al. (1999) now yield

\[
P\left(n^{1/3}(U_n(a) - g^{-1}(a)) > x(1 + z)\right) \leq P\left(n^{1/3}(U_n(a) - g^{-1}(a)) > \delta_n n^{1/3}\right) \leq \exp\{-nG(y_0, y_0 + \delta_n)h(v_n)\}, \tag{22}
\]

where $h(\cdot)$ is defined as in Groeneboom et al. (1999). The inequality for $h(\cdot)$ given there yields $h(v_n) \geq (v_n - 1)^2/(2(1 + t)^2)$. Using $a \delta_n - G(y_0, y_0 + \delta_n) \geq \int_{y_1}^{y_1 + x n^{-1/3}} (a - g(y)) \, dy \geq x n^{-1/3}(g(y_1)/(1 + x n^{-1/3}) - g(y_1)) = g(y_1)x^2 n^{-2/3}$ and $G(y_0, y_0 + \delta_n) \leq (y_1 - y_0)a + x n^{-1/3}g(y_1) \leq g(y_1)x n^{-1/3}(z + 1 + x t)$ (via $y_1 - y_0 \leq z x n^{-1/3}$), one obtains $|v_n - 1| = (a \delta_n - G(y_0, y_0 + \delta_n)/G(y_0, y_0 + \delta_n) \geq n^{-1/3}/(z(1 + t) + 1)$, and hence $h(v_n) \geq (x n^{-1/3})^2/(2(1 + t)^2(z(1 + t) + 1)^2)$. On the other hand, $G(y_0, y_0 + \delta_n) \geq g(y_1 + x n^{-1/3})x n^{-1/3} \geq g(y_1)y_1 x n^{-1/3}/(y_1 + x n^{-1/3}) \geq g(y_1)x n^{-1/3}/(z + t)$, where the second inequality follows from (21). So (22) is not larger than

\[
\exp\{-g(y_1)K((1 + z) x)^3\} = \exp\{-g(y_1)K((1 + z) x)^3\},
\]

completing the proof of Lemma 4. □

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 $1 \leq y_1 < y_2 \leq e^c$:

\[
g^c(y_2) - g^c(y_1) = f \left( \frac{\log y_2}{c} \right) \frac{1}{ty_2} - f \left( \frac{\log y_1}{c} \right) \frac{1}{ty_1} \leq \exp\{f \left( \frac{\log y_2}{c} \right) + c \left( \frac{\log y_2}{c} - \frac{\log y_1}{c} \right) \} \frac{1}{ty_2} - f \left( \frac{\log y_1}{c} \right) \frac{1}{ty_1} = f \left( \frac{\log y_1}{c} \right) \frac{y_2}{ty_1} \frac{1}{ty_2} - \frac{1}{ty_1} = 0.
\]

So $g^c$ is nonincreasing. As $\delta(f, H_0) \geq d_n$, there exist $0 < x_1 < x_2 \leq 1$ such that $f^{1/3}(x_2) - f^{1/3}(x_1) = \sup_{y_1 < x_2} (f^{1/3}(x_2) - f^{1/3}(s)) \geq 2/3 d_n$. Using (1) one readily deduces

\[
f(x_1) \geq \left( \frac{2/3 d_n}{\exp(c/3) - 1} \right)^3 \quad \text{and} \quad y_2 - y_1 \geq \frac{(\log n)^{1/5}}{f^{1/3}(x_1)} n^{-1/3}, \tag{23}
\]

24
where $y_i := e^{cxi}$.

To apply Lemma 4 we verify (21): By the definition of $x_1$ we have $g^c(y) = f(\log y)^{1/c} \geq f(\log y)^{1/c_y} = g^c(y_1)y/y$ for $y \in [1, y_2]$. By (23) 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 $g^c$ give

$$Pr \left\{ \frac{U_n(g^c(y_1)(1 + xn^{-1/3}))}{c} \leq y_1 + xn^{-1/3} \right\} \geq 1 - \exp(-K_1\sqrt{\log n}), \quad (24)$$

where $K_1 > 0$ does not depend on $f, n$ or $x_1$ (as $t_n \leq 1$ by (23)), and $U_n$ is taken with respect to the random variables $Y_i := e^{X_i}$. Using $b^{1/3} - a^{1/3} \leq \frac{1}{3}(b-a)a^{-2/3}$ for $a, b > 0$ we get

$$Pr \left\{ \frac{\left(\frac{n}{4c}\right)^{1/3} \left(\frac{\hat{f}_n^c(x_1)}{\hat{f}_n^c(x_1)} \left(\frac{\log(y_1 + xn^{-1/3})}{c} \right) - f^{1/3}(x_1) \right)}{c} \leq (\log n)^{1/6}d_n^{-1/3}(2e^c + 1)/3 \right\} \geq Pr \left\{ \left(\frac{n}{4c}\right)^{1/3} \left(\frac{\hat{f}_n^c(x_1)}{\hat{f}_n^c(x_1)} \left(\frac{\log(y_1 + xn^{-1/3})}{c} \right) - f^{1/3}(x_1) \right) \leq (\log n)^{1/6}d_n^{-1/3}(2e^c + 1) \right\}$$

$$\geq Pr \left\{ \hat{g}_n^c(y_1 + xn^{-1/3}) - g^c(y_1) \leq g^c(y_1)(xn^{-1/3} + y_1xn^{-1/3} + x^2n^{-2/3}) \right\}$$

$$\geq \exp(-K_1\sqrt{\log n}) \quad (25)$$

by (24) and (8). Similarly one finds

$$Pr \left\{ \left(\frac{n}{4c}\right)^{1/3} \left(\frac{\hat{f}_n^c(x_1)}{\hat{f}_n^c(x_1)} \left(\frac{\log(y_2 - x'n^{-1/3})}{c} \right) - f^{1/3}(x_2) \right) \geq -K_2(\log n)^{1/6} \right\} \geq 1 - \exp(-K_3\sqrt{\log n})$$

where $x' := (\log n)^{1/6}/f^{1/3}(x_2)$. Setting $\epsilon := C - (2c)^{1/3} > 0$ gives $(n/(4c))^{1/3}3/2(\log(n)/c)^{1/3}((\log n)^{1/6}(2c)^{1/3} + \epsilon)(\log n)^{1/3} - (1/2\log n + \log c/2)^{1/3} \geq \epsilon/(8c)^{1/3}(\log n)^{1/3}$ for $n$ large enough, depending only on $c$. Together with (25) and (26) we obtain for every constant $l$:

$$Pr \left\{ \left(\frac{n}{4c}\right)^{1/3} \left(\frac{\hat{f}_n^c(x_1)}{\hat{f}_n^c(x_1)} \left(\frac{\log(y_2 - x'n^{-1/3})}{c} \right) \right) \geq \left(\frac{\hat{f}_n^c(x_1)}{\hat{f}_n^c(x_1)} \left(\frac{\log(y_1 + xn^{-1/3})}{c} \right) \right) - \left(\frac{\sqrt{n\epsilon}}{2}\right)^{1/3} > l \right\} \rightarrow 1,$$

the convergence not depending on $f, y_1$ or $y_2$. As $y_1 + xn^{-1/3} < y_2 - x'n^{-1/3}$ by (23) and the critical value $l_n(1 - \alpha)$ is bounded by Theorem 2 and Fatou’s lemma, one obtains $Pr(\phi_n(T_n) = 1) \rightarrow 1$ uniformly in $\{f \in H_c : \delta(f, H_0) \geq d_n\}$, proving (b).

Note that the crucial feature of the asymptotically minimax adaptive test lies in comparing $(n/(4cc))^{1/3}T_n(c)$ to $(\log(n)^{1/6})(\sqrt{n\epsilon}/2)^{1/3}$ 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 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 (7). When employing $\tilde{F}_n$ in place of $F$ note that the Radon-Nikodym derivative $f(\tilde{F}_n^{-1}(t) + d)/\tilde{f}_n(\tilde{F}_n^{-1}(t))$ differs from the nonincreasing function $f(\tilde{F}_n^{-1}(t) + d)/f(\tilde{F}_n^{-1}(t))$ 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 logconcave $f$ using the fact that $f$ is bounded above and away from 0 on $(F^{-1}(\epsilon), F^{-1}(1-\epsilon))$.) 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


26


