University of Bern
Institute of Mathematical Statistics and Actuarial Science

Technical Report 56

Multiscale Inference about a Density

Lutz Dümbgen* (Bern) and Günther Walther** (Stanford)

January 2006, revised June 2007

Abstract

We introduce a multiscale test statistic based on local order statistics and spacings that provides simultaneous confidence statements for the existence and location of local increases and decreases of a density or a failure rate. The procedure provides guaranteed finite-sample significance levels, is easy to implement and possesses certain asymptotic optimality and adaptivity properties.

Keywords and phrases. exponential inequality, modes, monotone failure rate, multiple test, order statistics, spacings, subexponential increments.

AMS 2000 subject classification. 62G07, 62G10, 62G15, 62G20, 62G30

* Work supported by Swiss National Science Foundation

** Work supported by NSF grants DMS-9875598, DMS-0505682
and NIH grant 5R33HL068522

1 Introduction

An important aspect in the analysis of univariate data is inference about qualitative characteristica of their distribution function F or density f, such as the number and location of monotone or convex regions, local extrema or inflection points. This issue has been addressed in the literature using a variety of methods. Silverman (1981), Mammen et al. (1992), Minnotte and Scott (1993), Fisher et al. (1994), Minnotte (1997), Cheng and Hall (1999) and Chaudhuri and Marron (1999, 2000) use kernel density estimates. Excess masses and related ideas are employed by Hartigan and Hartigan (1985), Hartigan (1987), Müller and Sawitzky (1991), Polonik (1995) and Cheng and Hall (1998). Good and Gaskins (1980) and Walther (2001) use maximum likelihood methods, whereas Davies and Kovac (2004) employ the taut string method. In the present paper, a qualitative analysis of a density f means simultaneous confidence statements about regions of increase and decrease as well as local extrema. Such simultaneous inference has been treated in the literature only sparingly. Also, the methods available so far provide only approximate significance levels as the sample size tends to infinity and rely on certain regularity conditions about f.

In this paper we introduce and analyze a procedure that provides simultaneous confidence statements with guaranteed given significance level for arbitrary sample size. The approach is similar to Dümbgen (2002), who used local rank tests in the context of nonparametric regression, or Chaudhuri and Marron's (1999, 2000) SiZer, where kernel estimators with a broad range of bandwidths are combined. Here we utilize test statistics based on local order statistics and spacings. The use of spacings for nonparametric inference about densities has a long history. For instance, Pyke (1965) describes various goodness-of-fit tests based on spacings, and Roeder (1992) uses such tests for inference about normal mixtures. Confidence bands for an antitonic density on $[0, \infty)$ via uniform order statistics and spacings have been constructed by Hengartner and Stark (1995) and Dümbgen (1998).

In Section 2 we define local spacings and related test statistics which indicate isotonic or antitonic trends of f on certain intervals. Then a deterministic inequality (Proposition 1) relates the joint distribution of all these test statistics in general to the distribution in the special case of a uniform density. This enables us to define a multiple test about monotonicity properties of f. Roughly speaking, we consider all intervals whose endpoints are observations. The rationale for using and combining statistics corresponding to such a large collection of (random) intervals is that the power for detecting an increase or decrease of f is maximized when the tested interval is close to an interval on which f has such a trend. In that context we also discuss two important differences to Chaudhuri and Marron's SiZer map.

In Section 3 we describe a particular way of calibrating and combining the single test statistics. Optimality results in Section 4 show that in many relevant situations, the resulting multiscale test is asymptotically as powerful in the minimax sense as any procedure can essentially be for detecting increases and decreases of f on small intervals as well as on large intervals. Thus neither the guaranteed confidence level nor the consideration of many intervals simultaneously results in a substantial loss of power. In addition we prove that our procedure is able to detect and localize an arbitrary number of local extrema under weak assumptions on the strength of these effects.

In Section 5 we consider a density f on $(0, \infty)$ and modify our multiple test in order to analyze monotonicity properties of the failure rate $f/(1-F)$. It is well-known that spacings are a useful object in this context; see e.g. Proschan and Pyke (1967), Bickel and Doksum (1969) and Barlow and Doksum (1972). While these authors use global test statistics, Gijbels and Heckman (2004) localize, standardize and combine such tests, albeit without calibrating the various scales. Hall and Van Keilegom (2002) use resampling from an appropriately calibrated null distribution in order to achieve better sensitivity to detecting local effects, which leads to an asymptotically valid test procedure without explicit information about the location of these effects. Walther (2001) uses a multiscale maximum likelihood analysis to detect local effects.

Section 6 illustrates the multiscale procedures with two examples and introduces a graphical display. In Section 7 we derive auxiliary results about weighted maxima and moduli of continuity of stochastic processes. These results generalize Theorem 6.1 of Dümbgen and Spokoiny (2001) and are of independent interest. Further proofs and technical arguments are deferred to Section 8.

To fix notation for the sequel, suppose that $Y_1, Y_2, \ldots, Y_m$ are independent random variables with unknown distribution function F and (Lebesgue) density f on the real line. In order to infer properties of f from these data we consider the corresponding order statistics $Y_{(1)} < Y_{(2)} < \cdots < Y_{(m)}$. In some applications, F is known to be supported by an interval $[a, \infty)$, $(-\infty, b]$ or $[a, b]$, where $-\infty < a < b < \infty$. In that case we add the point $Y_{(0)} := a$ or $Y_{(m+1)} := b$ or both to our ordered sample, respectively. This yields a data vector $\boldsymbol{X} = (X_{(i)})_{i=0}^{n+1}$ with real components $X_{(0)} < X_{(1)} < \cdots < X_{(n+1)}$, where $n \in \{m-2, m-1, m\}$. For $0 \leq j < k \leq n+1$ with $k - j > 1$, the conditional joint distribution of $X_{(j+1)}, \ldots, X_{(k-1)}$, given $X_{(j)}$ and $X_{(k)}$, coincides with the joint distribution of the order statistics of $k - j - 1$ independent random variables with density

$$f_{jk}(x) := \frac{1\{x \in \mathcal{I}_{jk}\} f(x)}{F(X_{(k)}) - F(X_{(j)})},$$

where $\mathcal{I}_{jk}$ stands for the interval

$$\mathcal{I}_{jk} \ := \ (X_{(j)}, X_{(k)}).$$

Thus $(X_{(j+i)})_{i=0}^{k-j}$ is useful in order to infer properties of f on $\mathcal{I}_{jk}$. The multiple tests to follow are based on all such tuples.

2 Local spacings and monotonicity properties of f

Let us consider one particular interval $\mathcal{I}_{jk}$ and condition on its endpoints. In order to test whether f is non-increasing or non-decreasing on $\mathcal{I}_{jk}$ we introduce the local order statistics

$$X_{(i;j,k)} \ := \ \frac{X_{(i)} - X_{(j)}}{X_{(k)} - X_{(j)}}, \quad j \le i \le k,$$

and the test statistic

$$T_{jk}(\boldsymbol{X}) \ := \ \sum_{i=j+1}^{k-1} \beta(X_{(i;j,k)}),$$

where

$$\beta(x) \ := \ 1\{x \in (0,1)\}(2x-1).$$

This particular test statistic $T_{jk}(\boldsymbol{X})$ appears as a locally most powerful test statistic for the null hypothesis "$\lambda \le 0$" versus "$\lambda > 0$" in the parametric model, where

$$f_{jk}(x) \ = \ \frac{1\{x \in \mathcal{I}_{jk}\}}{X_{(k)} - X_{(j)}} \Big(1 + \lambda\Big(\frac{x - X_{(j)}}{X_{(k)} - X_{(j)}} - \frac{1}{2}\Big)\Big).$$

Elementary algebra yields an alternative representation of our single test statistics:

(2.1)
$$T_{jk}(\boldsymbol{X}) \ = \ -(k-j) \sum_{i=j+1}^{k} \beta\Big(\frac{i-j-1/2}{k-j}\Big)\Big(X_{(i;j,k)} - X_{(i-1;j,k)}\Big).$$

Thus $T_{jk}(\boldsymbol{X})$ is a weighted average of the local spacings $X_{(i;j,k)} - X_{(i-1;j,k)}$, $j < i \le k$.

Suppose that f is constant on $\mathcal{I}_{jk}$. Then the random variable $T_{jk}(\boldsymbol{X})$ is distributed (conditionally) as

(2.2)
$$\sum_{i=1}^{k-j-1} \beta(U_i)$$

with independent random variables U_i having uniform distribution on $[0,1]$. Note that the latter random variable has mean zero and variance $(k-j-1)/3$. However, if f is non-decreasing or non-increasing on $\mathcal{I}_{jk}$, then $T_{jk}(\boldsymbol{X})$ tends to be positive or negative, respectively. The following proposition provides a more general statement, which is the key to our multiple test.

Proposition 1 *Define $\boldsymbol{U} = (U_{(i)})_{i=0}^{n+1}$ with components $U_{(i)} := F_o(X_{(i)})$, where F_o is the distribution function corresponding to the density $f_{0,n+1}$. Then $U_{(1)}, \ldots, U_{(n)}$ are distributed as the order statistics of n independent random variables having uniform distribution on $[0,1]$, while $U_{(0)} = 0$ and $U_{(n+1)} = 1$. Moreover, for arbitrary integers $0 \le j < k \le n+1$ with $k - j > 1$,*

$$T_{jk}(\boldsymbol{X}) \begin{cases} \ge \ T_{jk}(\boldsymbol{U}) & \text{if } f \text{ is non-decreasing on } \mathcal{I}_{jk}, \\ \le \ T_{jk}(\boldsymbol{U}) & \text{if } f \text{ is non-increasing on } \mathcal{I}_{jk}. \end{cases}$$

This Proposition suggests the following multiple test: Suppose that for a given level $\alpha \in (0,1)$ we know constants $c_{jk}(\alpha)$ such that

$$\mathbb{P}\Big\{ |T_{jk}(\boldsymbol{U})| \le c_{jk}(\alpha) \text{ for all } 0 \le j < k \le n+1, k-j > 1 \Big\} \ \ge \ 1 - \alpha. \tag{2.3}$$

Let

$$\mathcal{D}^{\pm}(\alpha) \ := \ \Big\{ \mathcal{I}_{jk} : \pm T_{jk}(\boldsymbol{X}) > c_{jk}(\alpha) \Big\}.$$

Then one can claim with confidence $1-\alpha$ that f must have an increase on every interval in $\mathcal{D}^+(\alpha)$, and it must have a decrease on every interval in $\mathcal{D}^-(\alpha)$. In other words, with confidence $1-\alpha$ we may claim that for every $\mathcal{I} \in \mathcal{D}^{\pm}(\alpha)$ and for every version of f there exist points $x, y \in \mathcal{I}$ with $x < y$ and $\pm(f(y) - f(x)) > 0$.

Combining the two families $\mathcal{D}^{\pm}(\alpha)$ properly allows to detect and localize local extrema as well: Suppose for instance that $I_1, I_2, \ldots, I_m \in \mathcal{D}^+(\alpha)$ and $D_1, D_2, \ldots, D_m \in \mathcal{D}^-(\alpha)$ such that $I_1 \le D_1 \le I_2 \le D_2 \le \cdots \le I_m \le D_m$, where the inequalities are to be understood elementwise. Under the weak assumption that f is continuous, one can conclude with confidence $1-\alpha$ that f has at least m different local maxima and $m-1$ different local minima.

Note that our multiscale test allows to combine test statistics $T_{jk}(\boldsymbol{X})$ with arbitrary 'scales' $k-j$. This is an advantage over Chaudhuri and Marron's (1999, 2000) SiZer map, where statements about *multiple* increases and decreases are available only at a common bandwidth. This is due to the fact that these authors use kernels with unbounded support and rely on a particular variation reducing property of the gaussian kernel which holds only for an arbitrary but global bandwidth. Another consequence of the kernel's unbounded support is that localizing trends of f itself is not possible.

3 Combining the single test statistics T_{jk}

It remains to define constants $c_{jk}(\alpha)$ satisfying (2.3). Note first that $T_{jk}(\boldsymbol{U})$ has mean zero and standard deviation $\sqrt{(k-j-1)/3}$. Motivated by recent results of Dümbgen and Spokoiny

(2001) about multiscale testing in gaussian white noise models we consider the test statistic

$$T_n(\boldsymbol{X}) \;:=\; \max_{0\le j<k\le n+1\,:\,k-j>1}\Bigl(\sqrt{\frac{3}{k-j-1}}\,|T_{jk}(\boldsymbol{X})| - \Gamma\Bigl(\frac{k-j}{n+1}\Bigr)\Bigr),$$

where $\Gamma(\delta) := (2\log(e/\delta))^{1/2}$. This particular additive calibration for various scales is necessary for the optimality results to follow. Without the term $\Gamma((k-j)/(n+1))$, the null distribution would be dominated by small scales, as there are many more local test statistics on small scales than on large scales, with a corresponding loss of power at large scales. The next theorem states that our particular test statistic $T_n(\boldsymbol{U})$ converges in distribution. Unless stated differently, asymptotic statements in this paper refer to $n \to \infty$.

Theorem 2

$$T_n(\boldsymbol{U}) \;\to_{\mathcal{L}}\; T(W) \;:=\; \sup_{0\le u<v\le 1}\Bigl(\frac{|Z(u,v)|}{\sqrt{v-u}} - \Gamma(v-u)\Bigr),$$

where

$$Z(u,v) \;:=\; 3^{1/2}\int_u^v \beta\Bigl(\frac{x-u}{v-u}\Bigr)\,dW(x),$$

and W is a standard Brownian motion on $[0,1]$. Moreover, $0\le T<\infty$ almost surely.

Consequently, if $\kappa_n(\alpha)$ denotes the $(1-\alpha)$–quantile of $\mathcal{L}(T_n(\boldsymbol{U}))$, then $\kappa_n(\alpha) = O(1)$, and the constants

$$c_{jk}(\alpha) \;:=\; \sqrt{\frac{k-j-1}{3}}\Bigl(\Gamma\Bigl(\frac{k-j}{n+1}\Bigr) + \kappa_n(\alpha)\Bigr)$$

satisfy requirement (2.3). For explicit applications we do not use the limiting distribution in Theorem 2 but rely on Monte-Carlo simulations of $T_n(\boldsymbol{U})$ which are implemented easily.

4 Power considerations

Throughout this section we focus on the detection of increases of f by means of $\mathcal{D}^+(\alpha)$. Analogous results hold true for decreases of f and $\mathcal{D}^-(\alpha)$.

For any bounded open interval $I \subset \mathbb{R}$ we quantify the isotonicity of f on I by

$$\begin{aligned}\inf_I f' \;&:=\; \inf_{x,y\in I\,:\,x<y} \frac{f(y)-f(x)}{y-x}\\ &=\; \inf_{x\in I} f'(x) \quad \text{if } f \text{ is differentiable on } I.\end{aligned}$$

Now we analyze the difficulty of detecting intervals I with $\inf_I f' > 0$. An appropriate measure of this difficulty turns out to be

$$H(f,I) \;:=\; \inf_I f' \cdot |I|^2/\sqrt{F(I)},$$

where $|I|$ denotes the length of I. Note that this quantity is affine equivariant in the sense that it does not change when f and I are replaced by $\sigma^{-1}f(\sigma^{-1}(\cdot - \mu))$ and $\{\mu + \sigma x : x \in I\}$, respectively, with $\mu \in \mathbb{R}$, $\sigma > 0$. For given numbers $\delta \in (0,1]$ and $\eta \in \mathbb{R}$, we define

$$\mathcal{F}(I,\delta,\eta) \ := \ \big\{ f : F(I) = \delta, H(f,I) \geq \eta \big\}$$

and

$$\mathcal{F}(\delta,\eta) \ := \bigcup_{\text{bounded intervals } I} \mathcal{F}(I,\delta,\eta) \, .$$

Note that $f(x) \geq \inf_I f' \cdot (x - \inf(I))$ on I, so that $F(I) \geq \inf_I f' \cdot |I|^2/2$. Hence

$$H(f,I) \ \leq \ 2\sqrt{F(I)}. \tag{4.1}$$

Thus $\mathcal{F}(I,\delta,\eta)$ and $\mathcal{F}(\delta,\eta)$ are nonvoid if, and only if, $\eta \leq 2\sqrt{\delta}$.

Theorem 3 *Let $\delta_n \in (0,1]$ and $0 < c_n < \sqrt{24} < C_n$.*

(a) *Let I_n be a bounded interval and f_n a density in $\mathcal{F}\Big(I_n, \delta_n, C_n\sqrt{\log(e/\delta_n)/n}\Big)$. Then*

$$\mathbb{P}_{f_n}\Big(\mathcal{D}^+(\alpha) \text{ contains an interval } J \subset I_n\Big) \ \to \ 1 \, ,$$

provided that $\big(C_n - \sqrt{24}\big)\sqrt{\log(e/\delta_n)} \to \infty$.

(b) *Let $\phi_n(\boldsymbol{X})$ be any test with level $\alpha \in (0,1)$ under the null hypothesis that $\boldsymbol{X}$ is drawn from a nonincreasing density. If $(\log n)^2/n \leq \delta_n \to 0$, then*

$$\inf_{f \in \mathcal{F}\left(\delta_n, c_n\sqrt{\log(e/\delta_n)/n}\right)} \mathbb{E}_f \phi_n(\boldsymbol{X}) \ \leq \ \alpha + o(1) \, ,$$

provided that $\big(\sqrt{24} - c_n\big)\sqrt{\log(e/\delta_n)} \to \infty$.

(c) *Let I_n be any interval and b_n some number in $[0, 2\sqrt{n\delta_n}]$. If $\phi_n(\boldsymbol{X})$ is any test with level $\alpha \in (0,1)$ under the null hypothesis that the density is nonincreasing on I_n, then*

$$\inf_{f \in \mathcal{F}\left(I_n, \delta_n, b_n/\sqrt{n}\right)} \mathbb{E}_f \phi_n(\boldsymbol{X}) \ \to \ 1$$

implies that $b_n \to \infty$ and $n\delta_n \to \infty$.

Analogous results hold true for detecting a decrease of f. Theorem 3 establishes that our multiscale statistic is optimal in the asymptotic minimax sense for detecting an increase on an unknown interval, both in the case of an increase occuring on a small scale ($\delta_n \searrow 0$) and when the increase occurs on a large scale ($\liminf \delta_n > 0$).

In the case of small scales, a comparison of (a) and (b) shows that there is a cut-off for the quantity $H(f, I)$ at $\sqrt{24 \log(e/\delta_n)/n}$: If one replaces the factor 24 with $24 + \epsilon_n$ with $\epsilon_n \searrow 0$ sufficiently slowly, then the multiscale test will detect and localize such an increase with asymptotic power one, whereas in the case $24 - \epsilon_n$ no procedure can detect such an increase with nontrivial asymptotic power.

In the case of large scales, one may replace $\mathcal{F}\left(I_n, \delta_n, C_n\sqrt{\log(e/\delta_n)/n}\right)$ in (a) with the family $\mathcal{F}\left(I_n, \delta_n, \tilde{C}_n/\sqrt{n}\right)$, where $\tilde{C}_n \to \infty$. Then a comparison of (a) and (c) shows again our multiscale test to be optimal, even in comparison to tests using a priori knowledge of the location and scale of the potential increase. Hence searching over over all (large and small) scales does not incur a serious drawback. In the case of small scales, (a) and (c) together show that ignoring prior information about the location of the potential increase leads to a penalty factor of order $o\big(\sqrt{\log(e/\delta_n)}\big) = o\big(\sqrt{\log n}\big)$.

Example 1. Let us first illustrate the theorem in the special case of a fixed continuous density f and a sequence of intervals I_n converging to a given point x_o, where we use the abbreviation

$$\rho_n \ := \ \log(n)/n.$$

Example 1a. Let f be continuously differentiable in a neighborhood of x_o such that $f(x_o) > 0$ and $f'(x_o) > 0$. If $|I_n| = D_n \rho_n^{1/3}$ with $D_n \to D > 0$, then $\delta_n := F(I_n)$ is equal to $D_n f(x_o)\rho_n^{1/3}(1 + o(1))$ and $\inf_{I_n} f' = f'(x_o) + o(1)$. Hence the quantity $H(f, I_n)$ may be written as $D_n^{3/2} f'(x_o) f(x_o)^{-1/2} \rho_n^{1/2}(1 + o(1))$, while $\sqrt{24 \log(e/\delta_n)/n} = 8^{1/2}\rho_n^{1/2} + o(1)$. Consequently, the conclusion of Theorem 3 (a) is correct if

$$D_n \ \searrow \ \big(8 f(x_o)/f'(x_o)^2\big)^{1/3}$$

sufficiently slowly.

Example 1b. Let f be differentiable on (x_o, ∞) with $f(x_o) = 0$ and $f'(x_o + h) = \gamma h^{\kappa-1}(1 + o(1))$ as $h \searrow 0$, where $\gamma, \kappa > 0$. If $I_n = \left[x_o + C_1 \rho_n^{1/(\kappa+1)}, x_o + C_2 \rho_n^{1/(\kappa+1)}\right]$ with $0 \le C_1 < C_2$, then the conclusion of Theorem 3 (a) is correct, provided that $\min(C_1^{\kappa-1}, C_2^{\kappa-1})$ and C_2/C_1 are sufficiently large.

Example 1c. Let f be twice continuously differentiable in a neighborhood of x_o such that $f(x_o) > 0$, $f'(x_o) = 0$ and $\pm f''(x_o) \neq 0$. Now take the two intervals $I_n^{(\ell)} := \left[x_o - C_2 \rho_n^{1/5}, x_o - C_1 \rho_n^{1/5}\right]$ and $I_n^{(r)} := \left[x_o + C_1 \rho_n^{1/5}, x_o + C_2 \rho_n^{1/5}\right]$ with $0 < C_1 < C_2$. If C_1 and C_2/C_1 are sufficiently large, then it follows from Theorem 3 (a) and its extension to locally decreasing densities that

$$\mathbb{P}\big(\mathcal{D}^{\pm} \text{ contains some } J \subset I_n^{(\ell)} \text{ and } \mathcal{D}^{\mp} \text{ contains some } J \subset I_n^{(r)}\big) \ \to\ 1\,.$$

Thus our multiscale procedure will detect the presence of the mode with asymptotic probability one and furthermore localize it with precision $O_p\left((\log(n)/n)^{1/5}\right)$. Up to the logarithmic factor, this is the optimal rate for estimating the mode (cf. Hasminskii 1979).

Example 2. Now let I be a fixed bounded interval, and consider a sequence of densities f_n such that $\sup_{x\in I} |f_n(x) - f_o| \to 0$ for some constant $f_o > 0$. Here the conclusion of Theorem 3 (a) is correct, provided that

$$\sqrt{n} \cdot \inf_I f_n' \ \to\ \infty.$$

The next theorem is about the simultaneous detection of several increases of f.

Theorem 4 *Let $f = f_n$, and let $\mathcal{I}_n$ be a collection of non-overlapping bounded intervals such that for each $I \in \mathcal{I}_n$,*

(4.2)
$$H(f_n, I) \ \geq\ C\big(\sqrt{\log(e/F_n(I))} + b_n\big)/\sqrt{n}$$

with constants $0 \leq b_n \to \infty$ and $C \geq \sqrt{24}$. Then

$$\mathbb{P}_{f_n}\Big(\textit{for each } I \in \mathcal{I}_n,\ \mathcal{D}^+(\alpha) \textit{ contains an interval } J \subset I\Big) \ \to\ 1$$

in each of the following three settings, where $\delta_n := \min_{I\in\mathcal{I}_n} F_n(I)$:

(i) $C \geq 34$.

(ii) $C > 2\sqrt{24}$ *and* $n\delta_n/\log(e\#\mathcal{I}_n) \to \infty$.

(iii) $C = \sqrt{24}$ *and* $n\delta_n/\log(e\#\mathcal{I}_n) \to \infty,\ \log\#\mathcal{I}_n = o(b_n^2)$.

It will be shown in Section 8 that (4.2) entails $n\delta_n \geq \left(C^2/4 + o(1)\right)\log n$. In particular, $\#\mathcal{I}_n = o(n)$. Moreover, Theorem 3 (a) follows from Theorem 4 by considering setting (iii) with $\mathcal{I}_n$ consisting of a single interval I_n.

A comparison with Theorem 3 (a) shows that the price for the simultaneous detection of an increasing number of increases or decreases is essentially a potential increase of the constant $\sqrt{24}$.

The proof of Theorem 4 rests on an inequality involving the following auxiliary functions: For $c \in [-2, 2]$ and $u \in [0, 1]$ let

$$g_c(u) \;:=\; 1 + c(u - 1/2).$$

This defines a probability density on $[0, 1]$ with distribution function

$$G_c(u) \;:=\; u - cu(1 - u)/2.$$

Proposition 5 *Define $\boldsymbol{U} = (U_{(i)})_{i=0}^{n+1}$ as in Proposition 1. For arbitrary integers $0 \le j < k \le n + 1$ with $k - j > 1$ it follows from $\inf_{\mathcal{I}_{jk}} f' \ge 0$ that*

$$T_{jk}(\boldsymbol{X}) \;\ge\; \sum_{i=j+1}^{k-1} \beta\big(G_S^{-1}(U_{(i;j,k)})\big) \quad \textit{with} \quad S \;:=\; \frac{H(f, \mathcal{I}_{jk})}{\sqrt{F(\mathcal{I}_{jk})}}.$$

Moreover, for any fixed $c \in [-2, 2]$ and $U \sim \mathrm{Unif}[0, 1]$,

$$\mathbb{E}\beta(G_c^{-1}(U)) \;=\; c/6\,, \quad \mathrm{Var}\big(\beta(G_c^{-1}(U))\big) \;\le\; 1/3\,,$$

while

$$\mathbb{E}\exp\big(t\beta(G_c^{-1}(U))\big) \;\le\; \exp(ct/6 + t^2/6) \quad \textit{for all } t \in \mathbb{R}\,.$$

5 Monotonicity of the failure rate of f

To investigate local monotonicity properties of the failure rate $f/(1 - F)$, such as the presence of a ‘burn-in’ period or a ‘wear-out’ period, we consider

$$W_i \;:=\; \sum_{k=1}^{i} D_k \Big/ \sum_{k=1}^{n+1} D_k, \quad i = 0, \ldots, n + 1,$$

where $D_i := (n - i + 2)(X_{(i)} - X_{(i-1)})$, $i = 1, \ldots, n + 1$, are the normalized spacings. Here $X_{(0)} < X_{(1)} < \cdots < X_{(n+1)}$ are the order statistics of $n + 2$ or $n + 1$ i.i.d. observations from F, in the latter case with $X_{(0)}$ being the left endpoint of the support of F. The next proposition shows that the problem can now be addressed by applying the methodology of Section 2 to the transformed data vector $\boldsymbol{W} = (W_i)_{i=0}^{n+1}$.

Proposition 6 *Set $X'_{(i)} := -\log(1 - F(X_{(i)}))$, $i = 0, \ldots, n + 1$, and define $\boldsymbol{W}' = (W'_i)_{i=0}^{n+1}$ analogously as above with $\boldsymbol{X}'$ in place of $\boldsymbol{X}$. Then $\boldsymbol{W}' =_{\mathcal{L}} \boldsymbol{U}$, and for arbitrary integers $0 \le j < k \le n + 1$ with $k - j > 1$,*

$$T_{jk}(\boldsymbol{W}) \begin{cases} \ge \; T_{jk}(\boldsymbol{W}') & \textit{if the failure rate of } f \textit{ is non-decreasing on } \mathcal{I}_{jk}, \\ \le \; T_{jk}(\boldsymbol{W}') & \textit{if the failure rate of } f \textit{ is non-increasing on } \mathcal{I}_{jk}. \end{cases}$$

6 Graphical displays and examples

We first illustrate the methodology with a sample of size $m = 300$ from the mixture distribution

$$F \ = \ 0.3 \cdot \mathrm{Gamma}(2) + 0.2 \cdot \mathcal{N}(5, 0.1) + 0.5 \cdot \mathcal{N}(11, 9),$$

where $\mathrm{Gamma}(2)$ denotes the gamma distribution with density $g(x) = xe^{-x}$ on $(0, \infty)$. Figure 1 depicts the density f of F.

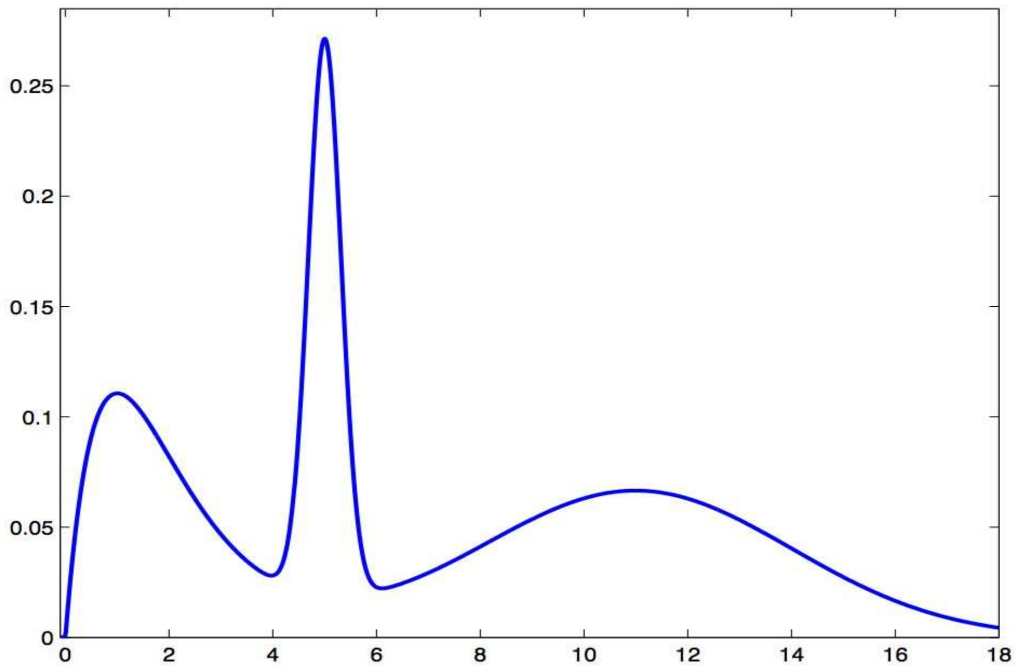

Figure 1: Density of $0.3 \cdot \mathrm{Gamma}(2) + 0.2 \cdot \mathcal{N}(5, 0.1) + 0.5 \cdot \mathcal{N}(11, 9)$

Figure 2 gives a line plot of the data and a visual display of the multiscale analysis: The horizontal line segments above the line plot depict all minimal intervals in $\mathcal{D}^+(0.1)$, those below the line plot depict all minimal intervals in $\mathcal{D}^-(0.1)$. Here we estimated the quantile $\kappa_{m-2}(0.1)$ to be 1.518 in 9999 Monte Carlo Simulations, where we restricted (j, k) in the definition of T to index pairs (j, k) such that $(k - j)/(m + 1) \leq 0.34$. For example, we can conclude with simultaneous confidence 90% that each of the intervals $(0.506, 3.887)$ and $(5.022, 5.841)$ contains a decrease, and each of the intervals $(3.983, 4.882)$ and $(5.841, 10.307)$ contains an increase. As these four intervals are disjoint, we can conclude with confidence 90% that the density has at least three modes.

A referee reports that the taut string method of Davies and Kovac (2004) found three modes in about 82% of the cases. Our method finds three modes in about 39% and exactly two modes in

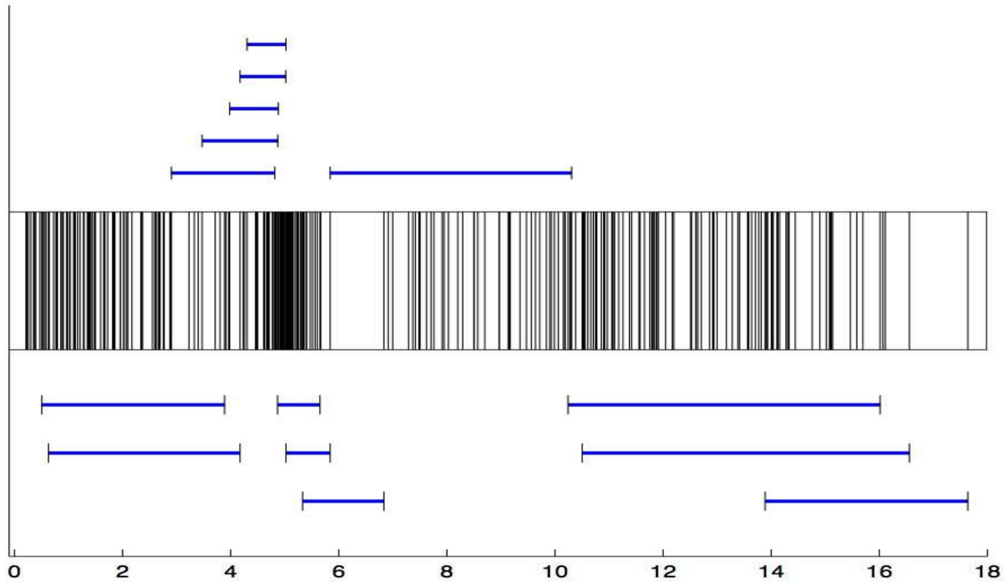

Figure 2: Minimal intervals in $\mathcal{D}^+(0.1)$ (top) and $\mathcal{D}^-(0.1)$ (bottom).

about 50% of the cases. However, the latter method also allows to localize the modes. Figure 3 provides a diagnostic tool for this type of inference. Each horizontal line segment, annotated by '+' or '-', depicts an interval in some $\mathcal{D}^+(\alpha)$, resp. $\mathcal{D}^-(\alpha)$. In each row, the depicted intervals are disjoint with an alternating sequence of signs. The number in the first colunm gives the smallest significance level at which this sequence of alternating signs obtains, and the plot shows all such sequences that have a significance level of 10% or less. The intervals depicted in a given row are chosen to have the smallest right endpoint among the minimal intervals at the stated level. Consecutive intervals are plotted with a small vertical offset to better visualize their endpoints. For example, figure 3 implies a p-value of less than 1% for the existence of at least two modes, and a p-value of 7.33% for the existence of at least three modes.

Our second example concerns the detection of an increase in a failure rate. Gijbels and Heckman (2004) compare a global test and four versions of a localized test in a simulation study. A sample of size $m = 50$ is drawn from a distribution whose hazard rate $h(t)$ is modeled via $\log h(t) = a_1 \log t + \beta(2\pi\sigma^2)^{-1/2} \exp\{-(t-\mu)^2/(2\sigma^2)\}$. Table 1 shows the power of our procedure from Section 5 for the choices of parameters a_1, β, σ used by Gijbels and Heckman (2004). The cases with $\beta = 0, a_1 \leq 0$ pertain to the null hypothesis of a non-increasing failure rate, whereas $\beta = 0, a_1 = 0.01$ implies an increasing failure rate. The other eight cases result in a failure rate with a local increase. The power of the test introduced in Section 5 exceeds those of the five tests examined by Gijbels and Heckman (2004) in four of the nine cases that involve an

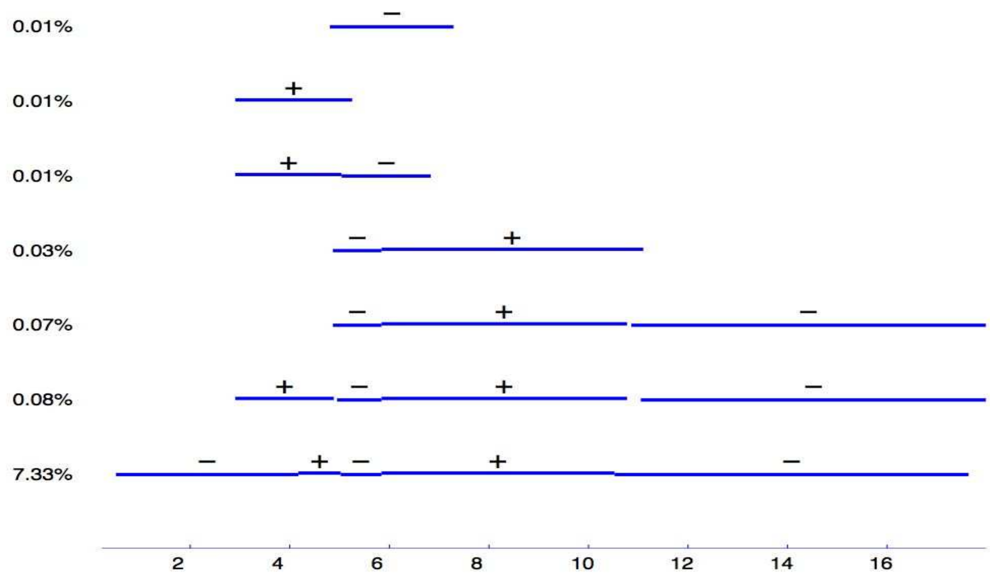

Figure 3: Alternating sequences of minimal intervals in $\mathcal{D}^+(\alpha)$ and $\mathcal{D}^-(\alpha)$ with the corresponding p-values α.

increase in the failure rate.

a_1	-0.2	-0.1	0	0.01
$\beta = 0$	0.014	0.026	0.049	0.052
$\beta = 0.3$, $\sigma = 0.2$	0.066	0.115	0.215	0.224
$\beta = 0.3$, $\sigma = 0.1$	0.188	0.301	0.439	0.451

Table 1: Proportion of rejections of the null hypothesis at the 5% significance level in 10,000 simulations.

7 Auxiliary results about stochastic processes

Throughout this section let $Z = (Z(t))_{t \in \mathcal{T}}$ be a stochastic process with continuous sample paths on a totally bounded metric space $(\mathcal{T}, \rho)$, where $\rho \leq 1$. 'Totally bounded' means that for arbitrary $u > 0$ the capacity number

$$D(u) = D(u, \mathcal{T}, \rho) \ := \ \max\Big\{\#\mathcal{T}_o : \mathcal{T}_o \subset \mathcal{T}, \rho(s,t) > u \text{ for different } s, t \in \mathcal{T}_o\Big\}$$

is finite. Moreover let $Z = (Z(t))_{t \in \mathcal{T}}$ be a stochastic process on $\mathcal{T}$ with continuous sample paths. We analyze the modulus of continuity of Z with respect to ρ. In addition we consider a function $\sigma : \mathcal{T} \to (0, 1]$, where $\sigma(t)$ may be viewed as measure of spread for the distribution of $Z(t)$. We assume that

(7.1) $$\big|\sigma(s) - \sigma(t)\big| \ \leq \ \rho(s,t) \quad \text{for all } s, t \in \mathcal{T},$$

and that $\{t \in \mathcal{T} : \sigma(t) \geq \delta\}$ is compact for any $\delta \in (0, 1]$.

We start with a version of Chaining which is simlar to Lemma VII.9 of Pollard (1984) and was used by Dümbgen (1998). For the reader's convenience a proof is given below.

Theorem 7 *Let K be some positive constant, and for $\delta > 0$ let $G(\cdot, \delta)$ a nondecreasing function on $[0, \infty)$ such that for all $\eta \geq 0$ and $s, t \in \mathcal{T}$ with $\rho(s, t) \geq \delta$,*

$$\mathbb{P}\Big\{\frac{|Z(s) - Z(t)|}{\rho(s,t)} > G(\eta, \delta)\Big\} \ \leq \ K \exp(-\eta). \tag{7.2}$$

Then for arbitrary $\delta > 0$ and $a \geq 1$,

$$\mathbb{P}\Big\{|Z(s) - Z(t)| \geq 12 J(\rho(s,t), a) \text{ for some } s, t \in \mathcal{T} \text{ with } \rho(s,t) \leq \delta\Big\} \ \leq \ \frac{K\delta}{2a},$$

where

$$J(\epsilon, a) \ := \ \int_0^\epsilon G(\log(aD(u)^2/u), u)\, du.$$

Remark 1. If we apply the preceding inequality to $\delta = 2^{-k}$ with $k = 0, 1, 2, \ldots$, then it follows from the Borel-Cantelli-Lemma that

$$\limsup_{\delta \searrow 0} \sup_{s,t \in \mathcal{T}_* : \rho(s,t) \leq \delta} \frac{|Z(s) - Z(t)|}{J(\rho(s,t), 1)} \ \leq \ 12 \quad \text{almost surely}.$$

Remark 2. Suppose that the process Z has sub–Weibull increments in the sense that for some constant $\kappa > 0$ and arbitrary $s, t \in \mathcal{T}$, $\eta \geq 0$,

$$\mathbb{P}\big\{|Z(s) - Z(t)| > \rho(s,t)\eta\big\} \ \leq \ 2\exp(-(\eta/\kappa)^\kappa).$$

Then the exponential inequality (7.2) is satisfied with $G(\eta, \delta) = (\kappa\eta)^{1/\kappa}$. This includes the situation of processes with subgaussian ($\kappa = 2$) and subexponential ($\kappa = 1$) increments.

Remark 3. Suppose that $G(\eta, \delta) = \tilde{q}\,\eta^q$ for some constants $\tilde{q}, q > 0$. In addition let

$$D(u) \ \leq \ Au^{-B} \quad \text{for } 0 < u \leq 1$$

with constants $A \geq 1$ and $B > 0$. Then elementary calculations show that for $0 < \epsilon \leq 1$ and $a \geq 1$,

$$J(\epsilon, a) \ \leq \ C\,\epsilon\, \log(e/\epsilon)^q$$

with $C = \tilde{q}\, \max\big(1 + 2B, \log(aA^2)\big)^q \int_0^1 \log(e/z)^q\, dz$.

With the conclusion of Theorem 7 in mind, we prove a result about the standardized process $Z/\sigma = \big(Z(t)/\sigma(t)\big)_{t \in \mathcal{T}}$.

Theorem 8 *Suppose that the following two conditions are satisfied:*

(i) *There is a function $G : [0,\infty) \times (0,1] \to [0,\infty)$ such that for arbitrary $\eta \ge 0$, $\delta \in (0,1]$ and $t \in \mathcal{T}$ with $\sigma(t) \ge \delta$,*

$$\mathbb{P}\Big\{|Z(t)| \ge \sigma(t) G(\eta,\delta)\Big\} \ \le\ 2\exp(-\eta).$$

Moreover,

$$G_o := \sup_{\eta\ge 0, 0<\delta\le 1} \frac{G(\eta,\delta)}{1+\eta} \ <\ \infty\,.$$

(ii) *There are positive constants A, B, V such that*

$$D\Big(u\delta, \{t \in \mathcal{T} : \sigma(t) \le \delta\}, \rho\Big) \ \le\ A u^{-B} \delta^{-V} \quad \text{for all } u, \delta \in (0,1].$$

For constants $q, Q > 0$ define the events

$$\mathcal{A}(q,Q,\delta) \ :=\ \Big\{ \sup_{s,t\in\mathcal{T}\,:\,\rho(s,t)\le\delta} \frac{|Z(s)-Z(t)|}{\rho(s,t)\log(e/\rho(s,t))^q} \le Q \Big\}, \quad \delta > 0.$$

Then there exists a constant $C = C(G_o, A, B, V, q, Q) > 0$ such that for $0 < \delta \le 1$ the probability of the event

$$\Big\{ |Z| \ \le\ \sigma\, G\big(V\log(1/\sigma) + C\log\log(e/\sigma), \sigma\big) + C\sigma\log(e/\sigma)^{-1} \text{ on } \{t : \sigma(t) \le \delta\} \Big\}$$

is at least $\mathbb{P}(\mathcal{A}(q,Q,2\delta)) - C\log(e/\delta)^{-1}$.

Remark. In case of $G(\eta,\delta) = (\kappa\eta)^{1/\kappa}$ with $\kappa > 1$,

$$\begin{aligned} & G\Big(V\log(1/\delta) + C\log\log(e/\delta), \delta\Big) + C\log(e/\delta)^{-1} \\ &= (\kappa V\log(1/\delta))^{1/\kappa} + O\Big(\log\log(e/\delta)\log(e\delta)^{1/\kappa-1}\Big) \\ &= (\kappa V\log(1/\delta))^{1/\kappa} + o(1) \quad \text{as } \delta \searrow 0. \end{aligned}$$

The preceding two theorems and remarks entail the following corollary which extends Theorem 6.1 of Dümbgen and Spokoiny (2001). The main difference is that we don't need to assume subgaussian increments of our stochastic process.

Corollary 9 *Suppose that the following three conditions are satisfied:*

(i) *There exist constants $A, B, V > 0$ such that for arbitrary $u, \delta \in (0,1]$,*

$$D\big(u\delta, \{t \in \mathcal{T} : \sigma(t) \le \delta\}, \rho\big) \ \le\ A u^{-B} \delta^{-V}.$$

(ii) *There exists a constant* $K \geq 1$ *such that for arbitrary* $s, t \in \mathcal{T}$ *and* $\eta \geq 0$*,*

$$\mathbb{P}\bigl(|Z(s) - Z(t)| \geq K\rho(s,t)\eta\bigr) \ \leq \ K \exp(-\eta).$$

(iii) *For arbitrary* $t \in \mathcal{T}$ *and* $\eta \geq 0$*,*

$$\mathbb{P}\bigl(|Z(t)| \geq \sigma(t)\eta\bigr) \ \leq \ 2\exp(-\eta^2/2).$$

Then

$$\mathbb{P}\Bigl(\sup_{s,t\in\mathcal{T}} \frac{|Z(s)-Z(t)|}{\rho(s,t)\log(e/\rho(s,t))} \geq \eta\Bigr) \ \leq \ p_1(\eta \,|\, A, B, K),$$

$$\mathbb{P}\Bigl(\sup_{t\in\mathcal{T}} \frac{|Z(t)|/\sigma(t) - \sqrt{2V\log(1/\sigma(t))}}{D(\sigma(t))} \geq \eta\Bigr) \ \leq \ p_2(\eta \,|\, A, B, V, K)$$

with $D(\delta) := \log(e/\delta)^{-1/2}\log(e\log(e/\delta))$*, where* $p_1(\cdot \,|\, A, B, K)$ *and* $p_2(\cdot \,|\, A, B, V, K)$ *are universal functions such that* $\lim_{\eta\to\infty} p_1(\eta \,|\, A, B, K) = \lim_{\eta\to\infty} p_2(\eta \,|\, A, B, V, K) = 0$.

Proof of Theorem 7. Since Z is assumed to have continuous sample paths, it suffices to verify the assertion on some dense subset $\mathcal{T}_*$ of $\mathcal{T}$. We choose inductively maximal subsets $\mathcal{T}_1 \subset \mathcal{T}_2 \subset \mathcal{T}_3 \subset \cdots$ of $\mathcal{T}$ such that

$$\rho(s,t) \ > \ \delta_k := 2^{-k}\delta \quad \text{for different } s, t \in \mathcal{T}_k.$$

In particular, for any $t \in \mathcal{T}$ and $k \geq 1$ there is a point $\pi_k(t) \in \mathcal{T}_k$ with $\rho(t, \pi_k(t)) \leq \delta_k$. Hence $\mathcal{T}_* := \bigcup_{k\geq 1}\mathcal{T}_k$ is a dense subset of $\mathcal{T}$. Furthermore, $\#\mathcal{T}_k \leq D(\delta_k)$. Now define

$$\eta_k \ := \ G\Bigl(\log(aD(\delta_k)^2/\delta_k), \delta_k\Bigr).$$

Then the event $\mathcal{A} := \bigcup_{k\geq 1}\bigl\{|Z(s) - Z(t)| > \rho(s,t)\eta_k \text{ for some } s, t \in \mathcal{T}_k\bigr\}$ has probability

$$\begin{aligned}
\mathbb{P}(\mathcal{A}) \ &\leq \ \sum_{k\geq 1}\sum_{\{s,t\}\subset\mathcal{T}_k} \mathbb{P}\bigl\{|Z(s)-Z(t)| > \rho(s,t)\eta_k\bigr\} \\
&\leq \ K\sum_{k\geq 1} 2^{-1}D(\delta_k)^2\exp\Bigl(-\log(aD(\delta_k)^2/\delta_k)\Bigr) \\
&= \ K\delta/(2a).
\end{aligned}$$

For $s, t \in \mathcal{T}_*$ there exist integers $1 \leq \ell < m$ with $\delta_{\ell-1} \geq \rho(s,t) > \delta_\ell$ and $s, t \in \mathcal{T}_m$ (where $\delta_0 := \delta$). Define $s_m := s, t_m := t$ and inductively $s_k := \pi_k(s_{k+1}), t_k := \pi_k(t_{k+1})$ for $k = m-1, m-2, \ldots, \ell$. Then one can conclude that

$$\rho(s_\ell, t_\ell) \ \leq \ \rho(s,t) + \sum_{k=\ell}^{m-1}\Bigl(\rho(s_k, s_{k+1}) + \rho(t_k, t_{k+1})\Bigr) \ \leq \ 6\delta_\ell.$$

Thus outside of the event $\mathcal{A}$,

$$\begin{aligned}
|Z(s) - Z(t)| &\le |Z(s_\ell) - Z(t_\ell)| + \sum_{k=\ell}^{m-1} \big(|Z(s_k) - Z(s_{k+1})| + |Z(t_k) - Z(t_{k+1})| \big) \\
&\le \rho(s_\ell, t_\ell)\eta_\ell + 2 \sum_{k=\ell}^{m-1} \delta_k \eta_{k+1} \\
&\le 12(\delta_\ell - \delta_{\ell+1})\eta_\ell + 8 \sum_{k \ge \ell} (\delta_{k+1} - \delta_{k+2})\eta_{k+1} \\
&< 12 \sum_{k=\ell}^{\infty} (\delta_k - \delta_{k+1})\eta_k \\
&\le 12 J(\delta_\ell, a) \\
&< 12 J(\rho(s,t), a)\,.
\end{aligned}$$

When bounding the series by an integral, we tacitly assumed that $G(\eta, \delta)$ is non-decreasing in $\eta \ge 0$ and non-increasing in $\delta > 0$. This may be assumed without loss of generality, because otherwise one could replace $G(\eta, \delta)$ in (7.2) with

$$\tilde{G}(\eta, \delta) := \inf_{\eta' \ge \eta, 0 < \delta' \le \delta} G(\eta', \delta') \;\le\; G(\eta, \delta)\,. \qquad \Box$$

Proof of Theorem 8. The idea is to prove the assertion on some countable subset $\mathcal{T}^*$ of $\mathcal{T}$ by means of conditions (i) and (ii), and then to use the modulus of continuity of Z on the events $\mathcal{A}(q, Q, \cdot)$.

The set $\mathcal{T}^*$ is constructed inductively as follows: Let t_1 be any point in $\mathcal{T}$ maximizing σ. Next let u be some continuous, non-decreasing function from $(0, 1]$ into itself to be specified later. Suppose that we picked already $t_1, \ldots, t_m$. If the set

$$\Big\{ t \in \mathcal{T} : \min_{i=1,\ldots,m} \rho(t, t_i) \ge u(\sigma(t))\sigma(t) \Big\} \tag{7.3}$$

is nonvoid, then let t_{m+1} be an element of it with maximal value $\sigma(t)$. Since the displayed set is closed and $\{\sigma \ge \delta\}$ is compact for any $\delta > 0$, the point t_{m+1} is well-defined. Thus we end up with a finite or countable set $\mathcal{T}^* := \{t_1, t_2, t_3, \ldots\}$, and its construction entails that $\sigma(t_1) \ge \sigma(t_2) \ge \sigma(t_3) \ge \cdots$. For $0 < \delta \le 1$ the set

$$\mathcal{T}^*(\delta) \;:=\; \Big\{ t \in \mathcal{T}^* : \delta/2 < \sigma(t) \le \delta \Big\}$$

is contained in $\Big\{ t \in \mathcal{T} : \sigma(t) \le \delta \Big\}$ with $\rho(s,t) \ge u(\delta/2)\delta/2$ for different $s, t \in \mathcal{T}^*(\delta)$. Consequently,

$$\#\mathcal{T}^*(\delta) \;\le\; A 2^B u(\delta/2)^{-B} \delta^{-V}.$$

In particular, if $\mathcal{T}^*$ is infinite, then $\lim_{m\to\infty} \sigma(t_m) = 0$. An important property of this set $\mathcal{T}^*$ is that for any $s \in \mathcal{T}$ there exists a point $t \in \mathcal{T}^*$ such that

$$\sigma(s) \ \le\ \sigma(t) \quad \text{and} \quad \rho(s,t) \ <\ u(\sigma(s))\sigma(s). \tag{7.4}$$

For let m be a maximal index such that $\sigma(t_m) \ge \sigma(s)$. If $\rho(s, t_i) \ge u(\sigma(s))\sigma(s)$ for all $i \le m$, then s would belong to the set (7.3), whence $\sigma(t_{m+1}) \ge \sigma(s)$. But this would be a contradiction to the definition of m.

In order to bound $|Z(t)|/\sigma(t)$ for all $t \in \mathcal{T}^*$ we define

$$H_1(t) \ :=\ G\Big(V\log(1/\sigma(t)) + B\log(1/u(\sigma(t))) + 2\log\log(e/\sigma(t)), \sigma(t)\Big).$$

Then for $0 < \delta \le 1$,

$$\begin{aligned}
&\mathbb{P}\left\{\sup_{t\in\mathcal{T}^*:\,\sigma(t)\le\delta}\Big(\frac{|Z(t)|}{\sigma(t)} - H_1(t)\Big) > 0\right\} \\
&\le \sum_{t\in\mathcal{T}^*:\,\sigma(t)\le\delta} \mathbb{P}\Big\{\frac{|Z(t)|}{\sigma(t)} \ge H_1(t)\Big\} \\
&\le 2\sum_{t\in\mathcal{T}^*:\,\sigma(t)\le\delta} \exp\Big(-V\log(1/\sigma(t)) - B\log(1/u(\sigma(t))) - 2\log\log(e/\sigma(t))\Big) \\
&= 2\sum_{k=0}^\infty \sum_{t\in\mathcal{T}^*(2^{-k}\delta)} \sigma(t)^V u(\sigma(t))^B \log(e/\sigma(t))^{-2} \\
&\le 2\sum_{k=0}^\infty \sum_{t\in\mathcal{T}^*(2^{-k}\delta)} (2^{-k}\delta)^V u(2^{-k}\delta)^B (\log(e/\delta) + \log(2)k)^{-2} \\
&\le C_1\sum_{k=0}^\infty (\log(e/\delta) + \log(2)k)^{-2} \\
&\le C_2\log(e/\delta)^{-1},
\end{aligned}$$

where

$$C_1 \ :=\ A2^{B+1}\sup_{0<x\le 1}\frac{u(x)}{u(x/2)} \quad \text{and} \quad C_2 \ :=\ (1 + (\log 2)^{-1})C_1.$$

Considering the function H_1 closely, an elegant choice for $u(\delta)$ might be

$$u(\delta) \ :=\ \log(e/\delta)^{-\gamma}$$

for some $\gamma > 0$. For then $u(x)/u(x/2) \le \log(2e)^\gamma$, and

$$H_1(t) \ =\ G\Big(V\log(1/\sigma(t)) + (B\gamma + 2)\log\log(e/\sigma(t)), \sigma(t)\Big).$$

Now let s be an arbitrary point in $\mathcal{T}$, and let $t \in \mathcal{T}^*$ satisfy (7.4). Then

$$\frac{\sigma(t)}{\sigma(s)} - 1 \ \le\ \frac{\rho(s,t)}{\sigma(s)} \ <\ u(\sigma(s)),$$

so that on the event $\mathcal{A}(2\sigma(s))$,

$$\begin{aligned}
\frac{|Z(s)|}{\sigma(s)} - \frac{|Z(t)|}{\sigma(t)} &\le \frac{|Z(s)-Z(t)|}{\sigma(s)} + \frac{|Z(t)|}{\sigma(t)}\Big(\frac{\sigma(t)}{\sigma(s)} - 1\Big) \\
&\le \frac{Q\rho(s,t)\log(e/\rho(s,t))^q}{\sigma(s)} + \frac{|Z(t)|}{\sigma(t)}\frac{\rho(s,t)}{\sigma(s)} \\
&\le Qu(\sigma(s))\log\Big(e/(u(\sigma(s))\sigma(s))\Big)^q + \frac{|Z(t)|}{\sigma(t)}u(\sigma(s)) \\
&\le C_3\log(e/\sigma(s))^{q-\gamma} + \frac{|Z(t)|}{\sigma(t)}\log(e/\sigma(s))^{-\gamma}
\end{aligned}$$

for some constant $C_3 = C_3(q, Q, \gamma)$. Consequently, if in addition $|Z(t)|/\sigma(t) \le H_1(t)$, then

$$\begin{aligned}
\frac{|Z(s)|}{\sigma(s)} &\le H_1(t) + C_3\log(e/\sigma(s))^{q-\gamma} + H_1(t)\log(e/\sigma(s))^{-\gamma} \\
&\le H_1(s) + C_3\log(e/\sigma(s))^{q-\gamma} + H_1(s)\log(e/\sigma(s))^{-\gamma} \\
&\le H_1(s) + C_3\log(e/\sigma(s))^{q-\gamma} \\
&\quad + \Big(1 + V\log(1/\sigma(t)) + (B\gamma+2)\log\log(e/\sigma(t))\Big)\log(e/\sigma(s))^{-\gamma} \\
&\le H_1(s) + C_4\log(e/\sigma(s))^{\max(1,q)-\gamma}
\end{aligned}$$

for some constant $C_4 = C_4(G_o, B, V, q, Q, \gamma)$. Finally note that $\sigma(s) \le \delta$ implies that $\sigma(t) \le 2\delta$. Consequently, with probability at least $\mathbb{P}(\mathcal{A}(2\delta)) - C_2\log(e/(2\delta))^{-1}$, the ratio $|Z(s)|/\sigma(s)$ is not greater than

$$G\Big(V\log(1/\sigma(s)) + (B\gamma+2)\log\log(e/\sigma(s)), \sigma(s)\Big) + C_4\log(e/\sigma(s))^{\max(1,q)-\gamma}$$

for all $s \in \{\sigma \le \delta\}$. This yields the assertion if we take $\gamma = \max(1,q)+1$ and a suitable $C = C(G_o, A, B, V, q, Q)$. $\square$

8 Proofs

8.1 Proofs of Propositions 1, 5 and 6

The proofs rely on an elementary inequality which we state without proof:

Lemma 10 *Let G_o and G be distribution functions on an interval (a,b) with densities g_o and g, respectively. Suppose that $g - g_o \le 0$ on (a,c) and $g - g_o \ge 0$ on (c,b), where $a < c < b$. Then $G^{-1} \ge G_o^{-1}$.* $\square$

Note that the conditions in Lemma 10 are satisfied if, for instance, g_o and g are differentiable with derivatives satisfying $g' \ge g_o'$.

Proof of Proposition 1. It is well-known that $U_{(1)}, \ldots, U_{(n)}$ are distributed as the order statistics of n independent random variables having uniform distribution on $[0,1]$. Suppose that f and thus f_{jk} is non-decreasing on $\mathcal{I}_{jk}$, where $k-j>1$. Then the assumptions of Lemma 10 are satisfied with $g = f_{jk}$ and $g_o(x) := 1\{x \in \mathcal{I}_{jk}\}/|\mathcal{I}_{jk}|$. This implies that for $j<i<k$,

$$X_{(i)} \;=\; G^{-1}(U_{(i;j,k)}) \;\geq\; G_o^{-1}(U_{(i;j,k)}) \;=\; X_{(j)} + (X_{(k)} - X_{(j)})U_{(i;j,k)},$$

whence $T_{jk}(\boldsymbol{X}) \geq T_{jk}(\boldsymbol{U})$. In case of f being non-increasing on $\mathcal{I}_{jk}$ the reverse inequality $T_{jk}(\boldsymbol{X}) \leq T_{jk}(\boldsymbol{U})$ follows from Lemma 10 with $g(x) = 1\{x \in \mathcal{I}_{jk}\}/|\mathcal{I}_{jk}|$ and $g_o := f_{jk}$. □

Proof of Proposition 5. Again we apply Lemma 10, this time with the densities

$$g(u) \;:=\; |\mathcal{I}_{jk}| f_{jk}\big(X_{(j)} + |\mathcal{I}_{jk}|u\big)$$

and $g_o := g_S$ on $(0,1)$. Note that

$$\inf_{(0,1)} g' \;=\; |\mathcal{I}_{jk}|^2 \inf_{\mathcal{I}_{jk}} f'_{jk} \;=\; S \;\equiv\; g'_S\,.$$

Thus it follows from Lemma 10 that

$$T_{jk}(\boldsymbol{X}) \;=\; \sum_{i=j+1}^{k-1} \beta\big(G^{-1}(U_{(i;j,k)})\big) \;\geq\; \sum_{i=j+1}^{k-1} \beta\big(G_S^{-1}(U_{(i;j,k)})\big).$$

As for the moments of $\beta(G_c^{-1}(U))$, note first that generally

$$\mathbb{E}h\big(\beta(G_c^{-1}(U))\big) \;=\; \int_0^1 h(\beta(u))(1 + c(u-1/2))\,du \;=\; \frac{1}{2}\int_{-1}^1 h(v)\Big(1 + \frac{c}{2}\,v\Big)\,dv$$

for $h : [-1,1] \to \mathbb{R}$. Letting $h(v) := v^j$ with $j = 1,2$ shows that the first and second moment of $\beta(G_c^{-1}(U))$ are given by $c/6$ and $1/3$, respectively. Moreover, letting $h(v) := \exp(tv)$ yields

$$M_c(t) \;:=\; \log \mathbb{E}\exp\big(t\beta(G_c^{-1}(U))\big) - ct/6 \;=\; \log(A(t) + cB(t)) - ct/6\,,$$

where

$$\begin{aligned} A(t) \;&:=\; \frac{1}{2}\int_{-1}^1 e^{tv}\,dv \;=\; \sinh(t)/t \;=\; \sum_{k=0}^\infty \frac{t^{2k}}{(2k+1)!}\,, \\ B(t) \;&:=\; \frac{1}{4}\int_{-1}^1 e^{tv} v\,dv \;=\; \big(\cosh(t)/t - \sinh(t)/t^2\big)/2 \;=\; \frac{t}{6}\sum_{k=0}^\infty \frac{3}{2k+3}\frac{t^{2k}}{(2k+1)!}\,. \end{aligned}$$

We have to show that $M_c(t) \leq t^2/6$ for any $t \neq 0$. To this end, note that $\partial M_c(t)/\partial c$ equals $B(t)/(A(t) + cB(t)) - t/6$ and $\partial^2 M_c(t)/\partial c^2 < 0$. Thus $M_c(t)$ is strictly concave in $c \in \{c :$

$A(t)+cB(t)>0\}$. The equation $\partial M_c(t)/\partial c=0$ is equivalent to $A(t)+cB(t)$ being equal to $6B(t)/t>0$, and this means $ct/6=1-tA(t)/(6B(t))$. Hence elementary manipulations of the series expansions yield

$$\begin{aligned}
M_c(t) &\leq \log\Big(\frac{6B(t)}{t}\Big)+\frac{tA(t)}{6B(t)}-1 \\
&= \log\Big(\sum_{k=0}^{\infty}\frac{3}{2k+3}\frac{t^{2k}}{(2k+1)!}\Big) \\
&\quad + \frac{t^2}{15}\sum_{k=0}^{\infty}\frac{5\cdot 3}{(2k+5)(2k+3)}\frac{t^{2k}}{(2k+1)!}\Big/\sum_{k=0}^{\infty}\frac{3}{2k+3}\frac{t^{2k}}{(2k+1)!} \\
&\leq \log\Big(\sum_{k=0}^{\infty}\frac{(t^2/10)^k}{k!}\Big)+\frac{t^2}{15} \\
&= \frac{t^2}{6}. \qquad \square
\end{aligned}$$

Proof of Proposition 6. By construction, the vector $(X'_{(i)}-X'_{(0)})_{i=1}^{n+1}$ is distributed as the vector of order statistics of $n+1$ independent random variables with standard exponential distribution. Well-known facts imply that the variables D'_i are independent with standard exponential distribution. Hence $(W'_1,\ldots,W'_n)=_{\mathcal{L}}(U_{(1)},\ldots,U_{(n)})$, while $W'_0=0$ and $W'_{n+1}=1$.

Now we assume that the failure rate is non-decreasing on $\mathcal{I}_{jk}$; the non-increasing case is treated analogously. Then the function $G(x):=-\log(1-F(x))$ is convex on $\mathcal{I}_{jk}$. Hence $\alpha_s:=D'_s/D_s$ is non-decreasing in $s\in\{j+1,\ldots,k\}$. Consequently for $j<i<k$,

$$\begin{aligned}
W_{(i;j,k)}-W'_{(i;j,k)} &= \frac{\sum_{s=j+1}^{i}D_s}{\sum_{s=j+1}^{k}D_s}-\frac{\sum_{s=j+1}^{i}\alpha_s D_s}{\sum_{s=j+1}^{k}\alpha_s D_s} \\
&= \frac{\sum_{s=j+1}^{i}\sum_{t=i+1}^{k}(\alpha_t-\alpha_s)D_sD_t}{\sum_{s=j+1}^{k}D_s\sum_{t=j+1}^{k}\alpha_t D_t} \\
&\geq 0.
\end{aligned}$$

Hence $T_{jk}(\boldsymbol{W})\geq T_{jk}(\boldsymbol{W}')$. $\square$

8.2 Proof of Theorem 2

We embed our test statistics T_{jk} into a stochastic process Z_n on

$$\mathcal{T}_n := \Big\{(\tau_{jn},\tau_{kn}) : 0\leq j<k\leq n+1\Big\},$$

where $\tau_{in}:=i/(n+1)$, equipped with the distance

$$\rho((u,v),(u',v')) := \Big(|u-u'|+|v-v'|\Big)^{1/2}$$

on $\mathcal{T} := \{(u,v) : 0 \le u < v \le 1\}$. Namely, let

$$Z_n(\tau_{jn}, \tau_{kn}) \ := \ 3^{1/2}(n+1)^{-1/2} T_{jk}(\boldsymbol{U}).$$

Moreover, for $(u,v) \in \mathcal{T} \setminus \mathcal{T}_n$ let

$$Z_n(u,v) \ := \ Z_n(\tau_n(u), \tau_n(v)) \quad \text{with } \tau_n(c) \ := \ \frac{\lfloor (n+1)c \rfloor}{n+1}.$$

Note that

$$\mathbb{E}(Z_n(u,v)) \ = \ 0 \quad \text{and} \quad \mathrm{Var}(Z_n(u,v)) \ \le \ \sigma(u,v)^2,$$

where $\sigma(u,v) := (v-u)^{1/2}$. In fact, these functions ρ and σ satisfy (7.1). For

$$\begin{aligned}
\big|\sigma(u,v) - \sigma(u',v')\big| \ &\le \ \frac{\big|(v-u) - (v'-u')\big|}{\sqrt{v-u} + \sqrt{v'-u'}} \\
&\le \ \frac{\sqrt{(v-u)+(v'-u')}\sqrt{|u-u'| + |v-v'|}}{\sqrt{v-u} + \sqrt{v'-u'}} \\
&\le \ \sqrt{|u-u'| + |v-v'|} \\
&= \ \rho((u,v),(u',v')).
\end{aligned}$$

Later on we shall prove the following two results about these processes Z_n and the limiting processs Z defined in Theorem 2:

Lemma 11 *The processes Z on $\mathcal{T}$ and Z_n on $\mathcal{T}_n$ ($n \in \mathbb{N}$) satisfy conditions (i–iii) of Corollary 9 with $A = 12$, $B = 4$, $V = 2$ and some universal constant K.*

Lemma 12 *For any finite subset $\mathcal{T}_o$ of $\mathcal{T}$, the random variable $(Z_n(t))_{t \in \mathcal{T}_o}$ converges in distribution to $(Z(t))_{t \in \mathcal{T}_o}$.*

Now we consider the preliminary test statistic

$$\begin{aligned}
\tilde{T}_n \ &:= \ \max_{0 \le j < k \le n+1} \Big(3^{1/2}(k-j)^{-1/2} T_{jk}(\boldsymbol{U}) - \Gamma\Big(\frac{k-j}{n+1}\Big)\Big) \\
&= \ \max_{t \in \mathcal{T}_n} \Big(\frac{|Z_n(t)|}{\sigma(t)} - \Gamma(\sigma(t)^2)\Big),
\end{aligned}$$

where $T_{jk}(\boldsymbol{U}) := 0$ if $k - j = 1$. We define

$$\tilde{T}_n(\delta, \delta') \ := \ \max_{t \in \mathcal{T}_n \,:\, \delta < \sigma(t) \le \delta'} \Big(\frac{|Z_n(t)|}{\sigma(t)} - \Gamma(\sigma(t)^2)\Big)$$

for $0 \le \delta < \delta' \le 1$ and $n \in \mathbb{N} \cup \{\infty\}$, where $(Z_\infty, \mathcal{T}_\infty) := (Z, \mathcal{T})$. Then it follows from Corollary 9 and Lemma 11 that for any fixed $\epsilon > 0$,

$$\lim_{\delta \searrow 0} \ \sup_{n \in \mathbb{N} \cup \{\infty\}} \ \mathbb{P}\{\tilde{T}_n(0,\delta) \ge \epsilon\} \ = \ 0 \tag{8.1}$$

and

$$\text{(8.2)} \qquad \lim_{\delta \searrow 0} \sup_{n \in \mathbb{N} \cup \{\infty\}} \mathbb{P}\Big\{ \sup_{s,t \in \mathcal{T}_n : \rho(s,t) \le \delta} \big| Z_n(s) - Z_n(t) \big| \ge \epsilon \} \ = \ 0.$$

The latter asymptotic continuity condition (8.2) and Lemma 12 imply that for any fixed $\delta \in (0,1]$,

$$\text{(8.3)} \qquad \tilde{T}_n(\delta, 1) \ \to_{\mathcal{L}} \ \tilde{T}_\infty(\delta, 1).$$

Finally, as in Dümbgen (2002) one can show that

$$\text{(8.4)} \qquad \lim_{\delta \searrow 0} \mathbb{P}\{\tilde{T}_\infty(\delta, 1) \le -\epsilon\} \ = \ 0$$

for any fixed $\epsilon > 0$. Combining the three facts (8.1), (8.3) and (8.4) yields that

$$\tilde{T}_n \ \to_{\mathcal{L}} \ T(W).$$

Finally we have to show that $T_n(\boldsymbol{U}) = \tilde{T}_n + o_p(1)$. Note that

$$T_n(\boldsymbol{U}) \ = \ \max_{t \in \mathcal{T}_n} \Big(\frac{|Z_n(t)|}{\sigma_n(t)} - \Gamma(\sigma(t)^2) \Big)$$

with

$$\sigma_n(t) \ := \ \Big(\sigma(t)^2 - (n+1)^{-1} \Big)^{1/2},$$

where we use the convention that $0/0 := 0$. The inequality $|Z_n(t)| \le (n+1)^{1/2} \sigma_n(t)^2$ entails that for $t \in \mathcal{T}_n$ with $\sigma(t) \le \delta_n := (\log(n+1)/(n+1))^{1/2}$,

$$\begin{aligned} \frac{|Z_n(t)|}{\sigma_n(t)} - \Gamma(\sigma(t)^2) \ &\le \ (n+1)^{1/2} \sigma_n(t) - \Gamma(\sigma(t)^2) \\ &\le \ (n+1)^{1/2} \delta_n - \Gamma(\delta_n^2) \\ &= \ \log(n+1)^{1/2} - (2 \log(n+1))^{1/2} + o(1) \\ &\to \ -\infty, \end{aligned}$$

and for $t \in \mathcal{T}_n$ with $\sigma(t) \ge \delta_n$,

$$\begin{aligned} \frac{|Z_n(t)|}{\sigma_n(t)} - \frac{|Z_n(t)|}{\sigma(t)} \ &= \ \frac{(\sigma(t) - \sigma_n(t))|Z_n(t)|}{\sigma_n(t)\sigma(t)} \\ &\le \ (n+1)^{1/2} (\sigma(t) - \sigma_n(t)) \\ &= \ (n+1)^{-1/2} (\sigma(t) + \sigma_n(t))^{-1} \\ &\le \ (n+1)^{-1/2} \delta_n^{-1} \\ &\to \ 0. \end{aligned}$$

Consequently,

$$T_n(\boldsymbol{U}) \ = \ \tilde{T}_n(\delta_n, 1) + o_p(1) \ = \ \tilde{T}_n + o_p(1). \qquad \square$$

Proof of Lemma 11. A proof of condition (i) is given by Dümbgen and Spokoiny (2001, proof of Theorem 2.1) in a slightly different setting. For the reader's convenience we repeat the argument here: For fixed $u, \delta \in (0,1]$ let $\epsilon := u^2\delta^2/2$ and define $I_j := [(j-1)\epsilon, j\epsilon] \cap [0,1]$ for $1 \leq j \leq m+1$, where $m := \lfloor \epsilon^{-1} \rfloor$. If $(a,b),(a',b') \in \mathcal{T}$ with $a, a' \in I_j$ and $b, b' \in I_k$ for some $j, k \in \{1, \ldots, m+1\}$, then $\rho((a,b),(a',b')) \leq u\delta$. On the other hand, $\sigma(a,b) \leq \delta$ implies that $(k-j-1)\epsilon \leq \delta^2$, whence $0 \leq k-j \leq 1 + 2/u^2$. These considerations show that

$$D\Big(u\delta, \{t \in \mathcal{T} : \sigma(t) \leq \delta\}, \rho\Big) \;\leq\; \#\Big\{(j,k) : 1 \leq j \leq m+1, j \leq k \leq j+1+2/u^2\Big\},$$

and the latter number is not greater than $(m+1)(2+2/u^2) \leq 12u^{-4}\delta^{-2}$.

Next we verify condition (ii). In order to bound the increment $Z_n(s) - Z_n(t)$ in terms of $\rho(s,t)$ we consider first the special case that $s = (0,1)$ and $t = (\tau, 1)$, where $\tau = \tau_{kn}$ for some $k \in \{1, \ldots, n\}$. Note that

$$\begin{aligned}
\sum_{i=1}^{n}(2U_{(i)}-1) &= \sum_{i=1}^{k-1}(2U_{(i)}-1) + 2U_{(k)} - 1 + \sum_{i=k+1}^{n}(2U_{(i)}-1), \\
\sum_{i=1}^{k-1}(2U_{(i)}-1) &= \sum_{i=1}^{k-1}\Big(2\frac{U_{(i)}}{U_{(k)}} - 1\Big)U_{(k)} + (k-1)U_{(k)}, \\
\sum_{i=k+1}^{n}(2U_{(i)}-1) &= \sum_{i=k+1}^{n}(2(U_{(i)} - U_{(k)}) - 1) + 2(n-k)U_{(k)} \\
&= \sum_{i=k+1}^{n}\Big(2\frac{U_{(i)} - U_{(k)}}{1 - U_{(k)}} - 1\Big)(1 - U_{(k)}) + (n-k)U_{(k)},
\end{aligned}$$

whence

$$Z_n(0,1) \;=\; Z_n(0,\tau)U_{(k)} + Z_n(\tau,1)(1-U_{(k)}) + 3^{1/2}(n+1)^{1/2}(U_{(k)} - \tau).$$

Consequently,

$$\begin{aligned}
&Z_n(0,1) - Z_n(\tau,1) \\
&\quad= \Big(Z_n(0,\tau) - Z_n(\tau,1)\Big)U_{(k)} + 3^{1/2}(n+1)^{1/2}(U_{(k)} - \tau) \\
&\quad= 3^{1/2}(n+1)^{-1/2}\left(\sum_{i=1}^{k-1}\beta\Big(\frac{U_{(i)}}{U_{(k)}}\Big) - \sum_{i=k+1}^{n}\beta\Big(\frac{U_{(i)} - U_{(k)}}{1 - U_{(k)}}\Big)\right)U_{(k)} \\
&\qquad + 3^{1/2}(n+1)^{1/2}(U_{(k)} - \tau) \\
&\quad=_{\mathcal{L}} 3^{1/2}(n+1)^{-1/2}\sum_{i=1}^{n-1}\beta(U_i')\,U_{(k)} + 3^{1/2}(n+1)^{1/2}(U_{(k)} - \tau),
\end{aligned}$$

where $U_1, \ldots, U_n, U_1', \ldots, U_{n-1}'$ are independent and identically distributed. Note that $U_{(k)}$ has a Beta-distribution with parameters k and $n+1-k$. This entails that

$$\mathbb{P}\Big\{\pm(U_{(k)} - \tau) \geq c\Big\} \;\leq\; \exp\Big(-(n+1)\Psi(\tau \pm c, \tau)\Big) \quad \text{for all } c \geq 0,$$

where $\Psi(x,\tau) := \tau\log(\tau/x) + (1-\tau)\log((1-\tau)/(1-x))$ if $x \in (0,1)$, and $\Psi(x,\tau) := \infty$ otherwise; see Proposition 2.1 of Dümbgen (1998). Elementary calculations show that $\Psi(\tau \pm c, \tau)$ is not smaller than $c^2/(2\tau(1-\tau)+2c)$, whence

$$\mathbb{P}\{\pm(U_{(k)} - \tau) \ge c\} \;\le\; \exp\Bigl(-\frac{(n+1)c^2}{2\tau(1-\tau)+2c}\Bigr) \tag{8.5}$$

for all $c \ge 0$. Consequently, for any $r \ge 0$,

$$\begin{aligned}
&\mathbb{P}\Bigl\{\Bigl|3^{1/2}(n+1)^{1/2}(U_{(k)} - \tau)\Bigr| \ge r\rho((0,1),(\tau,1))\Bigr\} \\
&= \;\mathbb{P}\Bigl\{\Bigl|3^{1/2}(n+1)^{1/2}(U_{(k)} - \tau)\Bigr| \ge r\tau^{1/2}\Bigr\} \\
&= \;\mathbb{P}\Bigl\{|U_{(k)} - \tau| \ge \frac{r\tau^{1/2}}{3^{1/2}(n+1)^{1/2}}\Bigr\} \\
&\le \;2\exp\Bigl(-\frac{r^2\tau}{6\tau(1-\tau) + 12^{1/2}r(n+1)^{-1/2}\tau^{1/2}}\Bigr) \\
&\le \;2\exp\Bigl(-\frac{r^2}{6 + 12^{1/2}r((n+1)\tau)^{-1/2}}\Bigr) \\
&\le \;2\exp\Bigl(-\frac{r^2}{6+4r}\Bigr) \\
&\le \;4\exp(-r/4). \qquad (8.6)
\end{aligned}$$

Here we used the fact that $(n+1)\tau \ge 1$. Moreover, for any $r \ge 1$,

$$\begin{aligned}
&\mathbb{P}\Bigl\{\Bigl|3^{1/2}(n+1)^{-1/2}\sum_{i=1}^{n-1}\beta(U_i')U_{(k)}\Bigr| \ge r\tau^{1/2}\Bigr\} \\
&\le \;\mathbb{P}\Bigl\{\Bigl|(3/n)^{1/2}\sum_{i=1}^{n-1}\beta(U_i')\Bigr| \ge r^{1/2}\Bigr\} + \mathbb{P}\Bigl\{U_{(k)} \ge r^{1/2}\tau^{1/2}\Bigr\} \\
&\le \;2\exp(-r/2) + \mathbb{P}\Bigl\{U_{(k)} - \tau \ge r^{1/2}\tau^{1/2} - \tau\Bigr\} \\
&\le \;2\exp(-r/2) + \exp\Bigl(-\frac{(n+1)(r^{1/2}-1)^2\tau}{2\tau(1-\tau) + 2(r^{1/2}-1)\tau^{1/2}}\Bigr) \\
&\le \;2\exp(-r/2) + \exp\Bigl(-\frac{(n+1)(r^{1/2}-1)^2\tau^{1/2}}{2 + 2(r^{1/2}-1)}\Bigr) \\
&\le \;2\exp(-r/2) + \exp\Bigl(-\frac{(n+1)^{1/2}(r^{1/2}-1)^2}{2r^{1/2}}\Bigr).
\end{aligned}$$

Note that the probability in question is zero if r is greater than $3^{1/2}(n+1)^{-1/2}(n-1)\tau^{-1/2}$, and the latter number is smaller than $3^{1/2}n$. Thus suppose that $r \le 3^{1/2}n$. Then

$$\frac{(n+1)^{1/2}(r^{1/2}-1)^2}{2r^{1/2}} \;\ge\; \frac{(3^{-1/2}r+1)^{1/2}(r^{1/2}-1)^2}{2r^{1/2}} \;\ge\; 3^{-1}(r^{1/2}-1)^2.$$

Consequently, for all $r \ge 0$ and some positive constant C_1,

$$\mathbb{P}\Bigl\{\Bigl|3^{1/2}(n+1)^{-1/2}\sum_{i=1}^{n-1}\beta(U_i')\,U_{(k)}\Bigr| \ge r\tau^{1/2}\Bigr\} \;\le\; C_1\exp(-r/C_1). \tag{8.7}$$

Combining (8.6) and (8.7) yields

$$\text{(8.8)} \qquad \mathbb{P}\Big\{ \Big| Z_n(0,1) - Z_n(\tau_{kn},1) \Big| \ge r\rho((0,1),(\tau_{kn},1)) \Big\} \ \le\ C_2 \exp(-r/C_2)$$

for some positive constant C_2. Symmetry considerations show that the same bound applies to $s = (0,1)$ and $t = (0,\tau)$, i.e.

$$\text{(8.9)} \qquad \mathbb{P}\Big\{ \Big| Z_n(0,1) - Z_n(0,\tau) \Big| \ge r\rho((\tau_{kn},1),(0,1)) \Big\} \ \le\ C_2 \exp(-r/C_2).$$

In order to treat the general case, note that the processes Z_n rescale as follows: For $0 \le J < K \le n+1$,

$$\begin{aligned} &\Big(Z_n(\tau_{J+j,n}, \tau_{J+k,n}) \Big)_{0 \le j < k \le K-J} \\ &=_{\mathcal{L}} \quad \sigma(\tau_{Jn}, \tau_{Kn}) \Big(Z_{K-J}(\tau_{j,K-J}, \tau_{k,K-J}) \Big)_{0 \le j < k \le K-J}, \end{aligned}$$

while for $0 \le j < k \le K-J$ and $0 \le j' < k' \le K-J$,

$$\begin{aligned} &\rho\Big((\tau_{J+j,n}, \tau_{J+k,n}), (\tau_{J+j',n}, \tau_{J+k',n}) \Big) \\ &= \quad \sigma(\tau_{Jn}, \tau_{Kn}) \rho\Big((\tau_{j,K-J}, \tau_{k,K-J}), (\tau_{j',K-J}, \tau_{k',K-J}) \Big). \end{aligned}$$

This entails that the bounds (8.8) and (8.9) can be extended as follows:

$$\text{(8.10)} \qquad \left. \begin{array}{c} \mathbb{P}\Big\{ \Big| Z_n(u,v) - Z_n(u,v') \Big| \ge r\rho((u,v),(u,v')) \Big\} \\[2ex] \mathbb{P}\Big\{ \Big| Z_n(u,v') - Z_n(u',v') \Big| \ge r\rho((u,v'),(u',v')) \Big\} \end{array} \right\} \ \le\ C_2 \exp(-r/C_2)$$

for arbitrary $(u,v), (u',v') \in \mathcal{T}$, where $u \le u'$. But note that

$$Z_n(u,v) - Z_n(u',v') \ =\ \big(Z_n(u,v) - Z_n(u,v') \big) + \big(Z_n(u,v') - Z_n(u',v') \big) \ =:\ \Delta_1 + \Delta_2$$

and

$$\rho((u,v),(u',v'))^2 \ =\ \rho((u,v),(u,v'))^2 + \rho((u,v'),(u',v'))^2 \ =:\ \rho_1^2 + \rho_2^2.$$

Consequently,

$$\begin{aligned} \mathbb{P}\Big\{ |\Delta_1 + \Delta_2| \ge r(\rho_1^2 + \rho_2^2)^{1/2} \Big\} &\le \mathbb{P}\Big\{ |\Delta_1 + \Delta_2| \ge r2^{-1/2}(\rho_1 + \rho_2) \Big\} \\ &\le \sum_{i=1}^{2} \mathbb{P}\{ |\Delta_i| \ge r2^{-1/2}\rho_i \} \\ &\le 2C_2 \exp(-r/(2C_2)). \end{aligned}$$

Hence condition (ii) is satisfied with $K = 2C_2$.

Finally, according to Proposition 5, $\mathbb{E}\exp(r\beta(U_i)) \le \exp(r^2/6)$ for all $r \in \mathbb{R}$, whence

$$\mathbb{E}\exp\Big(r\sigma(t)^{-1} Z_n(t) \Big) \ \le\ \exp(r^2/2) \quad \text{for } r \in \mathbb{R}, t \in \mathcal{T}_n.$$

Then a standard argument involving Markov's inequality yields condition (iii). □

Proof of Lemma 12. Recall the representation $U_{(i)} - U_{(i-1)} = E_i/S_n$ with independent, standard exponential variables E_i and $S_n = \sum_{j=1}^{n+1} E_j$. Starting from (2.1) one can write

$$\begin{aligned} Z_n(\tau_{jn}, \tau_{kn}) &= -3^{1/2} \frac{k-j}{(U_{(k)} - U_{(j)})S_n} (n+1)^{-1/2} \sum_{i=1}^{n+1} \beta\Big(\frac{i-j-1/2}{k-j}\Big) E_i \\ &= \frac{\tau_{kn} - \tau_{jn}}{U_{(k)} - U_{(j)}} \frac{n+1}{S_n} \times \tilde{Z}_n(\tau_{jn}, \tau_{kn}), \end{aligned}$$

where

$$\tilde{Z}_n(\tau_{jn}, \tau_{kn}) := 3^{1/2}(n+1)^{-1/2} \sum_{i=1}^{n+1} \beta\Big(\frac{\tau_{in} - \tau_{jn} - \delta_n}{\tau_{kn} - \tau_{jn}}\Big)(1 - E_i)$$

and $\delta_n := (2(n+1))^{-1}$. The centering of the variables E_i is possible because the sum of the coefficients $\beta((i-j-1/2)/(k-j))$, $j < i \le k$, is zero. Since $S_n/(n+1) \to_p 1$ and $\max_{1 \le i \le n} |U_{(i)} - \tau_{in}| \to_p 0$, it suffices to consider the stochastic process $\tilde{Z}_n$ in place of Z_n. But then the assertion follows from the multivariate version of Lindeberg's Central Limit Theorem and elementary covariance calculations. □

8.3 Proofs for Section 4

At first we prove the lower bounds comprising Theorem 3 (b–c). The following lemma is a surrogate for Lemma 6.2 of Dümbgen and Spokoiny (2001) in order to treat likelihood ratios and i.i.d. data.

Lemma 13 *Let $X_1, X_2, \ldots, X_n$ be i.i.d. with distribution P on some measurable space $\mathcal{X}$. Let $f_1, \ldots, f_m$ be probability densities with respect to P such that the sets $B_j := \{f_j \ne 1\}$ are pairwise disjoint, and define $L_j := \prod_{i=1}^n f_j(X_i)$. Then*

$$\mathbb{E}\Big| m^{-1} \sum_{j=1}^m L_j - 1 \Big| \to 0$$

provided that $m \to \infty$, $\Delta_\infty \le C(\log m)^{-1/2}$ for some fixed constant C and

$$\sqrt{\log m}\Big(1 - \frac{n\Delta_2^2}{2\log m}\Big) \to \infty,$$

where $\Delta_\infty := \max_j \sup_x |f_j(x) - 1|$ and $\Delta_2 := \max_j \Big(\int (f_j - 1)^2 dP\Big)^{1/2}$.

Proof of Lemma 13. The likelihood ratio statistics L_j are not stochastically independent, but conditional on $\boldsymbol{\nu} = (\nu_j)_{j=1}^m$ with $\nu_j := \#\{i : X_i \in B_j\}$ they are. Furthermore, $\mathbb{E}(L_j) = 1 =$

$\mathbb{E}(L_j \,|\, \boldsymbol{\nu})$. Thus a standard truncation argument shows that for any $\epsilon > 0$ and $0 < \gamma \le 1$,

$$\begin{aligned}
&\mathbb{E}\Big(\Big|m^{-1}\sum_j L_j - 1\Big| \,\Big|\, \boldsymbol{\nu}\Big) \\
&\le \; m^{-1}\mathrm{Var}\Big(\sum_j 1\{L_j \le \epsilon m\}L_j \,\Big|\, \boldsymbol{\nu}\Big)^{1/2} + 2m^{-1}\sum_j \mathbb{E}\Big(1\{L_j > \epsilon m\}L_j \,\Big|\, \boldsymbol{\nu}\Big) \\
&\le \; m^{-1}\Big(\sum_j \mathbb{E}\Big(1\{L_j \le \epsilon m\}L_j^2 \,\Big|\, \boldsymbol{\nu}\Big)\Big)^{1/2} + 2m^{-1}\sum_j \mathbb{E}\Big(1\{L_j > \epsilon m\}L_j \,\Big|\, \boldsymbol{\nu}\Big) \\
&\le \; m^{-1}\Big(\sum_j \mathbb{E}\Big(\epsilon m L_j \,\Big|\, \boldsymbol{\nu}\Big)\Big)^{1/2} + 2\epsilon^{-\gamma}m^{-(1+\gamma)}\sum_j \mathbb{E}\Big(L_j^{1+\gamma} \,\Big|\, \boldsymbol{\nu}\Big) \\
&= \; \epsilon^{1/2} + 2\epsilon^{-\gamma}m^{-(1+\gamma)}\sum_j \mathbb{E}\Big(L_j^{1+\gamma} \,\Big|\, \boldsymbol{\nu}\Big).
\end{aligned}$$

Thus it suffices to show that

$$\inf_{\gamma\in(0,1]} \max_j m^{-\gamma}\mathbb{E}(L_j^{1+\gamma}) \;\to\; 0$$

under the stated conditions on m, Δ_∞ and Δ_2. Note that $\mathbb{E}(L_j^{1+\gamma})$ equals $\mathbb{E}(f_j(X_1)^{1+\gamma})^n$, and elementary calculus reveals that

$$(1+y)^{1+\gamma} \;\le\; 1 + (1+\gamma)y + \gamma(1+\gamma)y^2/2 + 3\gamma|y|^3 \quad \text{for } |y| \le 1.$$

Hence $\mathbb{E}(f_j(X_1)^{1+\gamma}) \le 1 + \gamma(1+\gamma)\Delta_2^2/2 + 3\gamma\Delta_\infty\Delta_2^2$ and

$$\begin{aligned}
\max_j m^{-\gamma}\mathbb{E}(L_j^{1+\gamma}) \;&\le\; m^{-\gamma}\Big(1 + \gamma(1+\gamma)\Delta_2^2/2 + 3\gamma\Delta_\infty\Delta_2^2\Big)^n \\
&\le\; \exp\Big(-\gamma\log m + \gamma(1+\gamma)n\Delta_2^2/2 + 3\gamma\Delta_\infty n\Delta_2^2\Big). \qquad (8.11)
\end{aligned}$$

Suppose that $n\Delta_2^2 \le 2(1-b_m)\log m$, where $(0,1) \ni b_m \to 0$ and $b_m^2\log m \to \infty$ as $m \to \infty$. Then the right hand side of (8.11) does not exceed

$$\begin{aligned}
&\exp\Big(-\gamma(1-(1+\gamma)(1-b_m))\log m + 6\gamma\Delta_\infty\log m\Big) \\
&\le\; \exp\Big(-\frac{b_m^2\log m}{4(1-b_m)} + 3Cb_m(\log m)^{1/2}\Big) \quad \text{if } \gamma = \frac{b_m}{2(1-b_m)} \\
&\to\; 0 \quad \text{as } m \to \infty. \qquad \Box
\end{aligned}$$

Proof of Theorem 3 (b). Let $\tilde{c}_n := c_n\sqrt{\log(e/\delta_n)/n}$, and set $f_0 := 1_{[0,1)}$ and

$$f_{nj}(x) \;:=\; f_0(x) + 1\{x \in I_{nj}\}\tilde{c}_n\delta_n^{-3/2}(x - (j-1/2)\delta_n)$$

for $j = 1, \ldots, m_n := \lfloor 1/\delta_n \rfloor$ and $I_{nj} := [(j-1)\delta_n, j\delta_n)$. Each f_{nj} is a probability density with respect to the uniform distribution on $[0,1)$ such that the corresponding distribution F_{nj} satisfies

$F_{nj}(I_{nj}) = \delta_n$ and $\inf_{I_{nj}} f'_{nj} \cdot |I_{nj}|^2/\sqrt{F_{nj}(I_{nj})} = \tilde{c}_n$, i.e. $f_{nj} \in \mathcal{F}(\delta_n, \tilde{c}_n)$. Thus, for any test $\phi_n(\boldsymbol{X})$ with $\mathbb{E}_{f_0}\phi_n(\boldsymbol{X}) \le \alpha + o(1)$,

$$\begin{aligned}\inf_{f\in\mathcal{F}(\delta_n,\tilde{c}_n)} \mathbb{E}_f\phi_n(\boldsymbol{X}) - \alpha &\le m_n^{-1}\sum_{j=1}^{m_n} \mathbb{E}_{f_{nj}}\phi_n(\boldsymbol{X}) - \alpha \\ &= \mathbb{E}_{f_0}\left(\Big(m_n^{-1}\sum_{j=1}^{m_n} L_{nj} - 1\Big)\phi_n(\boldsymbol{X})\right) + o(1) \\ &\le \mathbb{E}_{f_0}\Big|m_n^{-1}\sum_{j=1}^{m_n} L_{nj} - 1\Big| + o(1),\end{aligned}$$

where $L_{nj} := \prod_{i=1}^n f_{nj}(X_i)$. The latter expectation tends to zero by Lemma 13. For $\Delta_2^2 = \tilde{c}_n^2/12$, and $\Delta_\infty = \tilde{c}_n\delta_n^{-1/2}/2$ is less than $\sqrt{6\log(e/\delta_n)/(n\delta_n)} = O(\log(n)^{-1/2}) = O(\log(m_n)^{-1/2})$, because $n\delta_n \ge \log(n)^2$ and hence $m_n = \delta_n^{-1} + O(1) = o(n)$. Finally,

$$\begin{aligned}\sqrt{\log m_n}\Big(1 - \frac{n\Delta_2^2}{2\log m_n}\Big) &= \frac{24\log m_n - c_n^2\log(e/\delta_n)}{24\sqrt{\log m_n}} \\ &\ge \sqrt{24}(\sqrt{24} - c_n)\sqrt{\log(e/\delta_n)}(1 + o(1)) + o(1)\end{aligned}$$

tends to infinity by assumption on δ_n and c_n. □

Proof of Theorem 3 (c). We may assume w.l.o.g. that the left endpoint of I_n is 0. Now we define probability densities f_n and g_n via

$$\begin{aligned}f_n(x) &:= \frac{\delta_n}{|I_n|}1\{x \in [0, |I_n|/\delta_n]\}, \\ g_n(x) &:= f_n(x) + \frac{\sqrt{\delta_n}b_n}{\sqrt{n}|I_n|^2}(x - |I_n|/2)1\{x \in I_n\}.\end{aligned}$$

Note that $g_n \ge 0$ because $b_n \le 2\sqrt{n\delta_n}$. Furthermore, f_n is non-increasing on I_n while g_n belongs to $\mathcal{F}(I_n, \delta_n, b_n/\sqrt{n})$.

Now we apply LeCam's notion of contiguity (cf. LeCam and Yang, 1990, Chapter 3): If a test $\phi_n(\boldsymbol{X})$ satisfies $\mathbb{E}_{f_n}\phi_n(\boldsymbol{X}) \le \alpha$, then $\limsup \mathbb{E}_{g_n}\phi_n(\boldsymbol{X}) < 1$, provided that

$$\mathcal{L}_{f_n}\Big(\sum_{i=1}^n \log(g_n/f_n)(X_i)\Big) \to_w Q \tag{8.12}$$

for some probability measure Q on the real line such that $\int e^x Q(dx) = 1$.

Note that $\mathcal{L}_{f_n}\Big(\sum_{i=1}^n \log(g_n/f_n)(X_i)\Big)$ equals the distribution of $\sum_{i=1}^{N_n} \log(1 + c_nV_i)$ with $c_n := b_n/(2\sqrt{n\delta_n}) \in [0, 1]$ and independent random variables N_n, V_1, V_2, V_3, …such that $N_n \sim \mathrm{Bin}(n, \delta_n)$ and $V_i \sim \mathrm{Unif}[-1, 1]$.

Suppose first that $n\delta_n \not\to \infty$. By extracting a subsequence, if necessary, we may assume that $n\delta_n \to \lambda \in [0,\infty)$ and $c_n \to c \in [0,1]$. Then (8.12) holds for the distribution $Q := \sum_{k=0}^\infty p_\lambda(k)\mathcal{L}\Big(\sum_{i=1}^k \log(1+cV_i)\Big)$ with the Poisson weights $p_\lambda(k) := e^{-\lambda}\lambda^k/k!$. But this measure Q satisfies $\int e^x Q(dx) = 1$, whence $\limsup \mathbb{E}_{g_n}\phi_n(\boldsymbol{X}) < 1$. This contradiction shows that $n\delta_n \to \infty$.

Secondly suppose that $n\delta_n \to \infty$ but $b_n \not\to \infty$. We assume w.l.o.g. that $b_n \to b \in [0,\infty)$. Lindeberg's Central Limit Theorem and elementary calculations yield (8.12) with gaussian distribution $Q = \mathcal{N}(-b^2/24, b^2/12)$. Again the limit distribution satisfies $\int e^x Q(dx) = 1$. Hence $b_n \to \infty$. $\Box$

Theorem 4 concerns our specific multiscale procedure. It will be derived from the following basic result.

Lemma 14 *For a bounded open interval I and $\delta \in (0,1]$ let f be a density in $\mathcal{F}\big(I,\delta,D\sqrt{\log(e/\delta)/n}\big)$ with $D \ge \sqrt{24}$. Then*

$$n\delta \ \ge\ \tilde{D}\max\big(\log(e/\delta), K\log(en)\big)$$

with $\tilde{D} := D^2/4$ and $K \ge 1 - \big(\log\tilde{D} + \log\log(en)\big)/\log(en)$. Suppose that

$$D \ \ge\ \frac{\sqrt{24}}{(1-\epsilon)^2\sqrt{1-\gamma-2/(n\delta)}}\Big(1 + \frac{\kappa_n(\alpha)+\eta}{\Gamma(\delta)} + \frac{\gamma+2/(n\delta)}{\Gamma(\delta)^2}\Big) \tag{8.13}$$

for certain numbers $\epsilon \in (0,1)$, $\gamma \in (0,1/2]$ and $\eta > 0$. Then

$$\begin{aligned}&\mathbb{P}\Big(\mathcal{D}^+(\alpha) \text{ contains no interval } J \subset I\Big)\\ &\le\ \exp(-n\delta\gamma^2/2) + 2\exp\big(-D\sqrt{n\delta\log(e/\delta)}\,\epsilon^2/8\big) + \exp(-\eta^2/2)\,.\end{aligned}$$

Proof of Lemma 14. The inequalities $2\sqrt{\delta} \ge H(f,I) \ge D\sqrt{\log(e/\delta)/n}$ entail that $n\delta \ge \tilde{D}\log(e/\delta)$. Now write $n\delta = \tilde{D}K\log(en)$ for some $K > 0$. In case of $K \le 1$,

$$\begin{aligned}\tilde{D}K\log(en) \ &\ge\ \tilde{D}\log(e/\delta) \ =\ \tilde{D}\big(\log(en) - \log(\tilde{D}K\log(en))\big)\\ &\ge\ \tilde{D}\log(en)\Big(1 - \frac{\log\tilde{D} + \log\log(en)}{\log(en)}\Big),\end{aligned}$$

and dividing both sides by $\tilde{D}\log(en)$ yields the asserted lower bound for K.

The number $N := \#\{i : X_{(i)} \in I\}$ has distribution $\mathrm{Bin}(m,\delta)$ with $m \in \{n, n+1, n+2\}$. Consequently it follows from Chernov's exponential inequality for binomial distributions (cf. van der Vaart and Wellner 1996, A.6.1) that

$$\mathbb{P}(N \le (1-\gamma)n\delta) \ \le\ \exp(-n\delta\gamma^2/2)\,.$$

Since $D \geq \sqrt{24}$ by assumption, we can conclude that $n\delta \geq D^2/4 > 6$, so that $(1-\gamma)n\delta \geq 3$. In case of $N \geq 3$ let $j := \min\{i : X_{(i)} \in I\}$ and $k := \max\{i : X_{(i)} \in I\}$, i.e. $N = k-j+1$. In order to bound the probability of $|\mathcal{I}_{jk}|/|I| < 1-\epsilon$, we write $I = (a,b)$ and define $I_{(\ell)} := (a, a+\epsilon|I|/2]$, $I_{(r)} := [b-\epsilon|I|/2, b)$. Then

$$\begin{aligned} nF(I_{(r)}) &\geq nF(I_{(\ell)}) \geq n \inf_I f' \cdot |I_{(\ell)}|^2/2 = nH(f,I)\sqrt{\delta}\,\epsilon^2/8 \\ &\geq D\sqrt{n\delta\log(e/\delta)}\,\epsilon^2/8\,, \end{aligned}$$

whence

$$\begin{aligned} &\mathbb{P}(N \leq 1 \text{ or } |\mathcal{I}_{jk}|/|I| \leq 1-\epsilon) \\ &\quad\leq \mathbb{P}(\text{no observations in } I_{(\ell)}) + \mathbb{P}(\text{no observations in } I_{(r)}) \\ &\quad\leq 2\exp\bigl(-D\sqrt{n\delta\log(e/\delta)}\,\epsilon^2/8\bigr)\,. \end{aligned}$$

From now on we always assume that $N \geq (1-\gamma)n\delta$ and $|\mathcal{I}_{jk}|/|I| \geq 1-\epsilon$. With $\mathbb{P}^*(\cdot)$ we denote conditional probabilities given these two inequalities. The definition of $\mathcal{D}^+(\alpha)$ implies that $\mathbb{P}^*\bigl(\mathcal{D}^+(\alpha) \text{ contains no } J \subset I\bigr)$ is not greater than $\mathbb{P}^*\bigl(T_{jk}(\boldsymbol{X}) \leq c_{jk}(\alpha)\bigr)$. On the other hand, it follows from Proposition 5 that

$$\mathbb{P}^*\Bigl(T_{jk}(\boldsymbol{X}) \leq \frac{\tilde{D}(N-2)}{6} - \eta\sqrt{\frac{N-2}{3}}\Bigr) \leq \exp(-\eta^2/2) \quad \text{for any } \eta \geq 0\,,$$

where $\tilde{C} := H(f,\mathcal{I}_{jk})/\sqrt{F(\mathcal{I}_{jk})}$. Thus it suffices to show that

$$\frac{\tilde{C}(N-2)}{6} - \eta\sqrt{\frac{N-2}{3}} \geq c_{jk}(\alpha)\,.$$

By definition of $c_{jk}(\alpha)$ this is equivalent to

$$\tilde{C}\sqrt{\frac{N-2}{12}} \geq \Gamma\Bigl(\frac{N-1}{n+1}\Bigr) + \kappa_n(\alpha) + \eta\,.$$

But the left hand side is not smaller than

$$\begin{aligned} \frac{(1-\epsilon)^2 H(f,I)}{\sqrt{\delta}}\sqrt{\frac{N-2}{12}} &\geq \frac{(1-\epsilon)^2 H(f,I)\sqrt{(1-\gamma)n\delta-2}}{\sqrt{12\delta}} \\ &\geq D\,\frac{(1-\epsilon)^2\sqrt{1-\tilde{\gamma}}}{\sqrt{24}}\,\Gamma(\delta) \\ &\geq \Gamma(\delta) + \kappa_n(\alpha) + \eta + \frac{\tilde{\gamma}}{\Gamma(\delta)} \end{aligned}$$

with $\tilde{\gamma} := \gamma + 2/(n\delta)$, whereas $\Gamma((N-1)/(n+1)) \leq \Gamma((N-2)/n)$ is not greater than

$$\Gamma(\delta(1-\tilde{\gamma})) \leq \Gamma(\delta) - \log(1-\tilde{\gamma})/\Gamma(\delta) \leq \Gamma(\delta) + \tilde{\gamma}/\Gamma(\delta)\,. \qquad \Box$$

Proof of Theorem 4. Note first that (4.2) and the first part of Lemma 14 entail that

$$n\delta_n \ \geq \ C^2/4 \ \geq \ 6 \quad \text{and} \quad n\delta_n \ \geq \ (C^2/4 + o(1)) \log n\,.$$

In particular, $\#\mathcal{I}_n \leq \delta_n^{-1} = o(n)$.

We apply Lemma 14 to $f = f_n$ and all intervals $I \in \mathcal{I}_n$. Precisely, we shall introduce suitable numbers $\gamma_n \in (0, 1/2]$, $\epsilon_n \in (0,1)$ and $\eta_{n,I} > 0$. According to Lemma 14, the probability that some $I \in \mathcal{I}_n$ does not cover an interval from $\mathcal{D}^+(\alpha)$ is bounded by

$$\#\mathcal{I}_n \Big(\exp(-n\delta_n\gamma_n^2/2) + 2\exp\big(-C\sqrt{n\delta_n \log(e/\delta_n)}\,\epsilon_n^2/8\big)\Big) + \sum_{I\in\mathcal{I}_n} \exp(-\eta_{n,I}^2/2), \tag{8.14}$$

provided that

$$C\Big(1 + \frac{\sqrt{2}b_n}{\Gamma(F_n(I))}\Big) \ \geq \ \frac{\sqrt{24}}{(1-\epsilon_n)^2\sqrt{1-\tilde\gamma_n}} \Big(1 + \frac{\kappa_n(\alpha) + \eta_{n,I}}{\Gamma(F_n(I))} + \frac{\tilde\gamma_n}{\Gamma(F_n(I))^2}\Big)$$

for all $I \in \mathcal{I}_n$, where $\tilde\gamma_n := \gamma_n + 2/(n\delta_n) = O(1)$. Note also that $\kappa_n(\alpha) = O(1)$ by virtue of Theorem 2. Hence the preceding requirement is met if for every constant $A > 0$ and sufficiently large n,

$$C\Big(1 + \frac{\sqrt{2}b_n}{\Gamma(F_n(I))}\Big) \ \geq \ \frac{\sqrt{24}}{(1-\epsilon_n)^2\sqrt{1-\tilde\gamma_n}} \Big(1 + \frac{A + \eta_{n,I}}{\Gamma(F_n(I))}\Big) \quad \text{for all } I \in \mathcal{I}_n. \tag{8.15}$$

In setting (i) we use constants $\gamma_n = \gamma \in (0, 1/2]$, $\epsilon_n = \epsilon \in (0,1)$ to be specified later and define

$$\eta_{n,I} \ := \ \sqrt{2\log(1/F_n(I)) + b_n} \ \leq \ \Gamma(F_n(I)) + \sqrt{b_n}.$$

Since $\delta\log(e/\delta)$ is nondecreasing in $\delta \in (0,1]$, it follows from $n\delta_n \geq (C^2/4 + o(1))\log n$ that

$$\sqrt{n\delta_n \log(e/\delta_n)} \ \geq \ (C/2 + o(1)) \ \log n\,.$$

Hence the bound in (8.14) equals

$$\begin{aligned} &o(1)\cdot\Big(\exp\big(-(C^2\gamma^2/8 - 1 + o(1))\log n\big) + \exp\big(-(C^2\epsilon^2/16 - 1 + o(1))\log n\big)\Big) \\ &\quad + \sum_{I\in\mathcal{I}_n} F_n(I)\exp(-b_n/2) \end{aligned}$$

and tends to zero, provided that $\gamma > \sqrt{8}/C$ and $\epsilon > 4/C$. Moreover, the right hand side of (8.15) is not greater than

$$\frac{\sqrt{24}}{(1-\epsilon)^2\sqrt{1-\gamma-o(1)}}\Big(2 + \frac{A + \sqrt{b_n}}{\Gamma(F_n(I))}\Big) \ = \ \frac{2\sqrt{24} + o(1)}{(1-\epsilon^2)\sqrt{1-\gamma}}\Big(1 + \frac{o(b_n)}{\Gamma(F_n(I))}\Big).$$

Hence the conclusion for setting (i) is correct if, say, $\epsilon = 4/(2\sqrt{24}) = \sqrt{1/6}$, $\gamma = \sqrt{8}/(2\sqrt{24}) = \sqrt{1/12}$, while C is strictly larger than

$$\frac{2\sqrt{24}}{(1-\epsilon)^2\sqrt{1-\gamma}} \; < \; 34 \, .$$

In settings (ii–iii) we define

$$\begin{aligned} \gamma_n &:= \big(2\log(D\#\mathcal{I}_n)/(n\delta_n)\big)^{1/2}, \\ \epsilon_n &:= \Big((8/C)\log(D\#\mathcal{I}_n)/\sqrt{n\delta_n\log(e/\delta_n)}\Big)^{1/2}, \\ \eta_{n,I} &:= \begin{cases} \sqrt{2\log(1/F_n(I)) + b_n} & \text{in setting (ii)}, \\ b_n/D & \text{in setting (iii)}, \end{cases} \end{aligned}$$

for some (large) constant $D > 1$. Then the bound in (8.14) is not greater than

$$3/D + \left\{ \begin{array}{cc} \exp(-b_n/2) & \text{in setting (ii)} \\ \exp\big(\log\#\mathcal{I}_n - b_n^2/(2D^2)\big) & \text{in setting (iii)} \end{array} \right\} \; = \; 3/D + o(1).$$

Thus it remains to verify (8.15).

Note that $\gamma_n \to 0$ by assumption. Moreover, since $\#\mathcal{I}_n \le \delta_n^{-1}$, the term $\log(D\#\mathcal{I}_n)$ is not greater than $\log(D/\delta_n)^{1/2}\log(D\#\mathcal{I}_n)^{1/2}$, whence

$$\epsilon_n \; \le \; \sqrt{8/C}(\log D)^{1/4}\big(\log(D\#\mathcal{I}_n)/(n\delta_n)\big)^{1/4} \; \to \; 0.$$

Hence in setting (ii) the right hand side of (8.15) is not greater than

$$\big(2\sqrt{24} + o(1)\big)\Big(1 + \frac{o(b_n)}{\Gamma(F_n(I))}\Big),$$

so that (8.15) is satisfied for sufficiently large n, if $C > 2\sqrt{24}$. In setting (iii), the right hand side of (8.15) is not greater than

$$\sqrt{24}\Big(1 + O(\tilde{\gamma}_n + \epsilon_n) + (1 + o(1))\frac{A + b_n/D}{\Gamma(F_n(I))}\Big).$$

According to the first part of Lemma 14, $n\delta_n \ge (C^2/4)\log(e/\delta_n) \ge (C^2/8)\Gamma(F_n(I))^2$ for all $I \in \mathcal{I}_n$. Thus

$$\tilde{\gamma}_n + \epsilon_n \; \le \; \frac{O\big(\log(D\#\mathcal{I}_n)^{1/2}\big)}{\Gamma(F_n(I))} \; = \; \frac{o(b_n)}{\Gamma(F_n(I))} \quad \text{for all } I \in \mathcal{I}_n.$$

Consequently, (8.15) is satisfied if $C \ge \sqrt{24}$. □

References

[1] BARLOW, R.E. and DOKSUM, K.A. (1972). Isotonic tests for convex orderings. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* ***1***, 293–323.

[2] BICKEL, P.J. and DOKSUM, K.A. (1969). Tests for monotone failure rates based on normalized spacings. *Ann. Math. Statist.* ***40***, 1216–1235.

[3] CHENG, M.Y. and HALL, P. (1998). Calibrating the excess mass and dip tests of modality. *J. Roy. Statist. Soc. Ser. B* ***60***, 579–589.

[4] CHENG, M.Y. and HALL, P. (1999). Mode testing in difficult cases. *Ann. Statist.* ***27***, 1294–1315.

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

[6] CHAUDHURI, P. and MARRON, J.S. (2000). Scale space view of curve estimation. *Ann. Statist.* ***28***, 408–428.

[7] DAVIES, P.L. and KOVAC, A. (2004). Densities, spectral densities and modality. *Ann. Statist.* ***32***, 1093–1136.

[8] DÜMBGEN, L. (1998). New goodness-of-fit tests and their application to nonparametric confidence sets. *Ann. Statist.* ***26***, 288–314.

[9] DÜMBGEN, L. (2002). Application of local rank tests to nonparametric regression. *J. Nonpar. Statist.* ***14***, 511–537.

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

[11] FISHER, N.I., MAMMEN, E. and MARRON, J.S. (1994). Testing for multimodality. *Comput. Statist. Data Anal.* ***18***, 499–512.

[12] GIJBELS, I. and HECKMAN, N. (2004). Nonparametric testing for a monotone hazard function via normalized spacings. *J. Nonpar. Statist.* ***16***, 463–477.

[13] GOOD, I.J. and GASKINS, R.A. (1980). Density estimation and bump hunting by the penalized maximum likelihood method exemplified by scattering and meteorite data (with discussion). *J. Amer. Statist. Assoc.* ***75***, 42–73.

[14] HALL, P. and VAN KEILEGOM, I. (2005). Testing for monotone increasing hazard rate. *Ann. Statist.* ***33***, 1109–1137.

[15] HARTIGAN, J.A. and HARTIGAN, P.M. (1985). The DIP test of multimodality. *Ann. Statist.* ***13***, 70–84.

[16] HARTIGAN, J.A. (1987). Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.* ***82***, 267–270.

[17] HASMINSKII, R.Z. (1979). Lower bounds for the risk of nonparametric estimates of the mode. In: *Contributions to Statistics, Jaroslav Hàjek Memorial Volume (J. Jureckova, ed.)*, pp. 91–97. Academia, Prague

[18] HENGARTNER, N.W. and P.B. STARK (1995). Finite-sample confidence envelopes for shape-restricted densities. *Ann. Statist.* ***23***, 525–550.

[19] LECAM, L. and YANG, G.L. (1990). *Asymptotics in Statistics.* Springer, New York

[20] MAMMEN, E., MARRON, J.S. and FISHER, N.I. (1992). Some asymptotics for multimodality tests based on kernel density estimates. *Prob. Theory Rel. Fields* ***91***, 115–132.

[21] MINNOTTE, M.C. and SCOTT, D.W. (1993). The mode tree: a tool for visualization of nonparametric density features. *J. Comp. Graph. Statist.* ***2***, 51–68.

[22] MINNOTTE, M.C. (1997). Nonparametric testing of the existence of modes. *Ann. Statist.* ***25***, 1646–1660.

[23] MÜLLER, D.W. and SAWITZKI, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.* ***86***, 738–746.

[24] POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York

[25] POLONIK, W. (1995). Measuring mass concentrations and estimating density contour clusters - An excess mass approach. *Ann. Statist.* ***23***, 855–881.

[26] PYKE, R. (1995). Spacings (with discussion). *J. Roy. Statist. Soc. Ser. B* ***27***, 395–449.

[27] PROSCHAN, F. and PYKE, R. (1967). Tests for monotone failure rate. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* ***3***, 293–313.

[28] ROEDER, K. (1992). Semiparametric estimation of normal mixture densities. *Ann. Statist.* ***20***, 929–943.

[29] SHORACK, G.R. and WELLNER, J.A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York

[30] SILVERMAN, B.W. (1981). Using kernel density estimates to investigate multimodality. *J. Roy. Statist. Soc. Ser. B* ***43***, 97–99.

[31] VAN DER VAART, A.W. and WELLNER, J.A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer, New York

[32] WALTHER, G. (2001). Multiscale maximum likelihood analysis of a semiparametric model, with applications. *Ann. Statist.* ***29***, 1297–1319.