Bernoulli **25**(4A), 2019, 3041–3068
https://doi.org/10.3150/18-BEJ1079

Signal detection via Phi-divergences for general mixtures

MARC DITZHAUS

Mathematical Institute, Heinrich-Heine University Düsseldorf, Universitätststraße 1, 40225 Düsseldorf, Germany. E-mail: marc.ditzhaus@hhu.de

The family of goodness-of-fit tests based on Φ-divergences is known to be optimal for detecting signals hidden in high-dimensional noise data when the heterogeneous normal mixture model is underlying. This test family includes Tukey's popular higher criticism test and the famous Berk–Jones test. In this paper we address the open question whether the tests' optimality is still present beyond the prime normal mixture model. On the one hand, we transfer the known optimality of the higher criticism test for different models, for example, for the heteroscedastic normal, general Gaussian and exponential-χ^2-mixture models, to the whole test family. On the other hand, we discuss the optimality for new model classes based on exponential families including the scale exponential, the scale Fréchet and the location Gumbel models. For all these examples we apply a general machinery which might be used to show the tests' optimality for further models/model classes in future.

Keywords: Berk and Jones test; detection boundary; Φ-divergences; sparse and dense signal detection; Tukey's higher criticism

1. Introduction

In several research areas it is of interest to detect rare and weak signals hidden in a huge noisy background. Examples for such areas are genomics [14,22,28], disease surveillance [36,38], local anomaly detection [39] as well as cosmology and astronomy [10,33]. Let us, exemplary, discuss an application in genomics. An important aim is to determine as early as possible whether a patient is healthy or affected by a common disease like cancer or leukemia. Many researchers assume (see the references above) that the majority of affected patients' genes behaves like genes of non-affected patients (noisy background) and only a small amount of genes displays a slightly different behavior (signals). In other words, if there are any signals at all then they are represented rarely and weakly. This combination makes it very difficult to detect the signals. In this paper, we study tests for this kind of signal detection problems. After introducing the mathematical model, we give more details about tests which were already suggested in the literature and explain our new insights into these.

Let P_n be a known continuous noise distribution and μ_n be an unknown signal distribution on $(\mathbb{R}, \mathcal{B})$. Popular examples are shift models, i.e., $\mu_n(-\infty, x] = P_n(-\infty, x - \vartheta_n]$ with signal strength $\vartheta_n \in \mathbb{R}$. Now, let $X_{n,1}, \dots, X_{n,n}$ be an *i.i.d.* (independent and identical distributed) sample with

$$X_{n,1} \sim Q_n = (1 - \varepsilon_n) P_n + \varepsilon_n \mu_n \qquad \text{fo some } \varepsilon_n \in [0, 1].$$

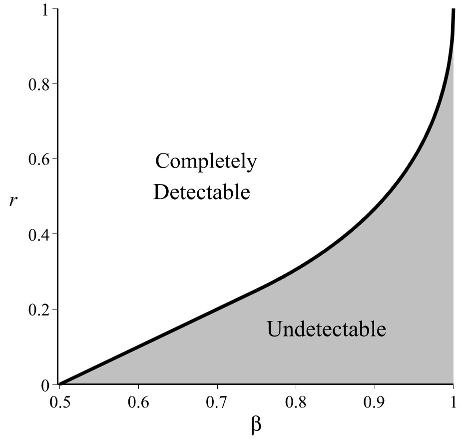

Figure 1. The detection boundary for the sparse heterogeneous normal mixture model $\beta \mapsto \rho(\beta)$.

The parameter ε_n can be seen as the parameter of the signal quantity. In fact, it is the probability that $X_{n,1}$ follows the signal distribution μ_n instead of the noise distribution P_n. The actual number of signals is random and approximately of the size $n\varepsilon_n$. When observing $X_{n,1}, \dots, X_{n,n}$ the question arises whether it is pure noise or some signals are hidden in it. This leads to the testing problem

$$\mathcal{H}_{0,n} : \varepsilon_n = 0 \quad \text{versus} \quad \mathcal{H}_{1,n} : \varepsilon_n > 0. \tag{1.1}$$

We focus on the challenging case of rare signals $\varepsilon_n \to 0$, where we differ between the sparse signal case ($n\varepsilon_n^2 \to 0$) and the dense signal case ($n\varepsilon_n^2 \to \infty$). Throughout this paper, if not stated otherwise, all limits are meant as $n \to \infty$. Clearly, the likelihood ratio test is the best test for (1.1). Its power behavior was studied by Ingster [27] for the heterogeneous normal mixture model $P_n = N(0, 1)$ and $\mu_n = N(\vartheta_n, 1)$. Using the parametrization $\varepsilon_n = n^{-\beta}$, $\beta \in (1/2, 1)$, and $\vartheta_n = \sqrt{2r \log(n)}$, $r > 0$, he showed that the detection boundary $\rho(\beta)$ given by

$$\rho(\beta) = \begin{cases} \beta - \dfrac{1}{2} & \text{if } \beta \in \left(\dfrac{1}{2}, \dfrac{3}{4}\right], \\ (1 - \sqrt{1-\beta})^2 & \text{if } \beta \in \left(\dfrac{3}{4}, 1\right), \end{cases} \tag{1.2}$$

splits the r-β-parametrization plane, see Figure 1, into the completely detectable and the undetectable area. If $r > \rho(\beta)$ then the likelihood ratio test can completely separate $\mathcal{H}_{0,n}$ and $\mathcal{H}_{1,n}$ asymptotically (completely detectable case), that is, the sum of type 1 and 2 error probabilities tends to 0. Otherwise, if $r < \rho(\beta)$, the likelihood ratio test and, thus, any other test cannot distinguish between $\mathcal{H}_{0,n}$ and $\mathcal{H}_{1,n}$ asymptotically (undetectable case). Later, Donoho and Jin [17] showed the optimality of a modified version of Tukey's higher criticism [42–44] in the sense that it can completely separate $\mathcal{H}_{0,n}$ and $\mathcal{H}_{1,n}$ asymptotically if $r > \rho(\beta)$. In contrast to the likelihood ratio test, the higher criticism test does not need the knowledge of the unknown signal probability ε_n and signal strength ϑ_n. Jager and Wellner [31] introduced a family of test statistics based on Φ-divergences including the higher criticism test statistic and the test statistic of Berk and Jones [6]. They extended the optimality result of [17] to their whole family. But in contrast to the higher criticism test [2–4,8,9,17,26,37], it is less known if this optimality also holds under more general model assumptions for the whole family.

As already mentioned, we differ between dense and sparse signals, where the main focus in the literature lies on the latter one. There are only a few positive results about the higher criticism test for dense signals. For instance, Cai et al. [8] proved the higher criticism test's optimality in the dense signal case for the normal location model introduced above with $\vartheta_n \to 0$.

When we explained the results of Ingster [27] we omitted the case $r = \rho(\beta)$, the behavior on the detection boundary. Ingster [27] determined the limit distribution of the likelihood ratio test on the boundary under $\mathcal{H}_{0,n}$ as well as under $\mathcal{H}_{1,n}$. An interesting observation is that non-Gaussian limits do also occur. In other words, he showed that there is a third area in the r-β-parametrization plane, namely the nontrivial power area on the boundary. Ditzhaus and Janssen [16] studied the asymptotic behavior of the likelihood ratio test and the higher criticism test on the detection boundary for general mixtures rigorously. In particular, they showed that the higher criticism test has no power on the boundary for various models, whereas the likelihood ratio test has nontrivial power there.

In short, these are our new insights:

(i) We enlarge the family of Jager and Wellner [30].

(ii) We give a positive answer to the uncertainty whether the Φ-divergence tests' optimality still hold beyond the normality assumption.

- We verify the optimality of the whole family for a model class recently suggested by Cai and Wu [9]. Among others, this (extended) class include the heterogeneous normal, the exponential-χ^2 and different exponential family mixture models, like the scale exponential or the scale Fréchet model.
- In contrast to the main literature we pay also attention to the dense case ($n\varepsilon_n^2 \to \infty$) and prove the Φ-divergence tests' optimality for general exponential family mixture models.

(iii) The negative result of the higher criticism test corresponding to the signal detectability on the detection boundary can be transferred to the whole test family.

The paper is organized as follows. In Section 2, we introduce the (enlarged) family of Φ-divergence test statistics and present the limit distribution under the null $\mathcal{H}_{0,n}$, which is the same for the whole family. Section 3 contains our tools to discuss the asymptotic power of the whole family under the alternative $\mathcal{H}_{1,n}$. These tools are applied in Section 4 to verify the optimality postulated in (ii) for the examples mentioned therein. Section 5 consists of a discussion.

2. The test family and its limit under the null

2.1. The test statistics

This papers' focus lies on continuous noise distributions. That is why we can assume without loss of generality that $P_n = P_0$ for all n, having, for example, a transformation to p-values in mind. Denote the distribution function of P_0 by F_0. The basic idea is to compare the empirical distribution function $\mathbb{F}_n(u) = n^{-1}\sum_{i=1}^n \mathbf{1}\{X_{n,i} \le u\}$ with the noise distribution function $F_0(u)$ for $u \in \mathbb{R}$ by using one of the Φ-divergence tests proposed by Csiszár [13] based on a convex function Φ, see also [1,12]. To be more specific, we introduce a family $(\phi_s)_{s\in\mathbb{R}}$ of convex functions mapping $(0,\infty)$ to $(0,\infty]$:

$$\phi_s(x) = \begin{cases} x - \log(x) - 1 & \text{for } s = 0, \\ x\big(\log(x) - 1\big) + 1 & \text{for } s = 1, \\ \big(1 - s + sx - x^s\big)/\big(s(1-s)\big) & \text{for } s \neq 0, 1. \end{cases}$$

Based on these the family of Φ-divergence statistics $(K_s)_{s\in\mathbb{R}}$ is given by

$$K_s(u,v) = v\phi_s\left(\frac{u}{v}\right) + (1-v)\phi_s\left(\frac{1-u}{1-v}\right)$$

for $u, v \in (0,1)$. It is easy to see that $\mathbb{R} \ni s \mapsto \phi_s(x)$ is continuous for every fixed $x \in (0,\infty)$ and so is $\mathbb{R} \ni s \mapsto K_s(u,v)$ for all fixed $u, v \in (0,1)$. Now, we consider the following family $\{S_n(s) : s \in \mathbb{R}\}$ of test statistics for (1.1) given by

$$S_n(s) = \sup_{X_{1:n} \le x < X_{n:n}} K_s\big(\mathbb{F}_n(x), F_0(x)\big), \tag{2.1}$$

where $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ denote the order statistics of the observation vector $(X_{n,1}, \ldots, X_{n,n})$. We want to point out that Jager and Wellner [31] restricted their family to $s \in [-1, 2]$, but as we will see there is no need for this constraint. As explained in [31] Tukey's higher criticism test ($s = 2$), the test of Berk and Jones [6] ($s = 1$), the "reversed Berk–Jones" statistic introduced by Jager and Wellner [30] ($s = 0$) and a studentized version of the higher criticism statistic studied by Eicker [20] ($s = -1$) are included in this family. Note that $S_n(s)$ does not always coincide with the corresponding known test statistic but is equivalent to them for s given in the parentheses. For all other s, the test statistic $S_n(s)$ was new. Jager and Wellner [31] give a special emphasis to $S_n(1/2)$, which is equivalent to the supremum of the pointwise Hellinger divergence between two Bernoulli distributions with parameters $F_0(u)$ and $\mathbb{F}_n(u)$.

2.2. Limit distribution under the null

The limit distribution of the higher criticism statistic is already known [29,40], Section 16.1, and so is the asymptotic behavior of $S_n(2)$. By the following two observations, we can motivate, at least heuristically, that the resulting asymptotic behavior under the null can be directly transferred to $S_n(s)$: 1. $\mathbb{F}_n(x)/F_0(x) \approx 1 \approx (1-\mathbb{F}_n(x))/(1-F_0(x))$ under the null. 2. $\phi_s(x)/\phi_2(x) \to 1$ for $x \to 1$. Clearly, for a mathematical correct proof there is a little bit more to do. For $s \in [-1, 2]$ this was already done by Jager and Wellner [31] and we extend their proof idea to all $s \in \mathbb{R}$. Here and subsequently, we denote by $\xrightarrow{\mathrm{d}}$, $\xrightarrow{\mathrm{P}_0^n}$, $\xrightarrow{\mathrm{Q}_n^n}$ convergence in distribution, in P_0^n-probability and in Q_n^n-probability, respectively.

Theorem 2.1. *Define*

$$r_n = \log\log(n) + \frac{1}{2}\log\log\log(n) - \frac{1}{2}\log(4\pi).$$

Then we have for all $s \in \mathbb{R}$ that under the null $\mathcal{H}_{0,n}$

$$nS_n(s) - r_n \xrightarrow{\mathrm{d}} Y, \tag{2.2}$$

where $Y - \log(4)$ is standard Gumbel distributed, i.e., $x \mapsto \exp(-4\exp(-x))$ is the distribution function of Y.

At least for $S_n(2)$ it is known that the convergence rate is really slow [34]. Since the basic proof idea of Theorem 2.1 is to approximate $nS_n(s)$ by $nS_n(2)$ the same bad rate can be expected for all $s \in \mathbb{R}$ or an even worse rate due to an additional approximation error. Consequently, we cannot recommend using a critical value based on the convergence result in Theorem 2.1 unless the sample size n is really huge. Since the noise distribution is assumed to be known, we suggest to use a standard Monte-Carlo simulation to estimate the α-quantile of $S_n(s)$. Alternatively, the reader can find finite recursion formulas for the exact finite distribution in the literature, see Jager and Wellner [30] (for $s = 0$, up to $n = 10^3$) and Khmaladze and Shinjikashvili [34] (for $s = 2$, up to $n = 10^4$).

3. Asymptotic power under the alternative

The same tool, already used by Ditzhaus and Janssen [16] to study the power behavior of the higher criticism test ($s = 2$), can be applied to the whole family. Its applicability is illustrated in Section 4. Let us become more concrete. We introduce $H_n : (0, 1/2) \to (0, \infty)$ given by

$$H_n(v) = \frac{\sqrt{n}\varepsilon_n}{\sqrt{v}}\left(\left|\mu_n^{F_0}(0, v] - v\right| + \left|\mu_n^{F_0}(1-v, 1) - v\right|\right), \qquad v \in \left(0, \frac{1}{2}\right), \tag{3.1}$$

where $\mu_n^{F_0}$ is the distribution of $F_0(X_{n,1})$ if $X_{n,1} \sim \mu_n$. Basically, $H_n(v)$ compares the tails near to 0 and near to 1 of the p-value $F_0(X_{n,1})$ for $X_{n,1}$ following the signal distribution μ_n and the noise distribution P_0, respectively.

Theorem 3.1 (Complete detection). *Suppose that there is a sequence $(v_n)_{n\in\mathbb{N}}$ in $(0, 1/2)$ such that $v_n n \to \infty$ and $(\log\log(n))^{-1} H_n(v_n) \to \infty$. Then for all $s \in \mathbb{R}$*

$$nS_n(s) - r_n \longrightarrow \infty \qquad \text{in } Q_n^n\text{-probability}. \tag{3.2}$$

By Theorems 2.1 and 3.1, we can conclude that under (3.2) there exists a sequence of critical values $c_n(s)$ such that the type 1 and 2 error probabilities of $\varphi_n(s) = \mathbf{1}\{S_n(s) > c_n(s)\}$ tend to 0. In other words, by using $S_n(s)$ we can completely separate $\mathcal{H}_{0,n}$ and $\mathcal{H}_{1,n}$ asymptotically. Ditzhaus and Janssen [16] showed that the assumptions of Theorem 3.1 are fulfilled, for example, for the heterogeneous normal mixture model from the Introduction. While Theorem 3.1 can be used to verify the optimality of the whole family, the next theorem is for a finer analysis of the asymptotic power, namely, for the power analysis on the detection boundary. For various model assumptions [8,16,27] it is known that the log-likelihood ratio has real-valued, nontrivial limits under $\mathcal{H}_{0,n}$ as well as under $H_{1,n}$ on the boundary, where also non-Gaussian limits occur. In particular, the asymptotic power of the log-likelihood ratio test is nontrivial in the sense that the sum of error probabilities tends neither to 0 nor to 1 but to a value in between. By the first lemma of Le Cam real-valued limits of the log-likelihood ratio imply mutual contiguity of P_0^n and Q_n^n, i.e., for all sequences $(A_n)_{n\in\mathbb{N}}$ of sets the following equivalence is true: $P_0^n(A_n) \to 0$ if and only if $Q_n^n(A_n) \to 0$.

Theorem 3.2 (No power). *Suppose that P_0^n and Q_n^n are mutually contiguous and that there are constants $\kappa, c_{1,n}, c_{2,n}, c_{3,n}, c_{4,n} \in (0,1)$ such that*

$$\sqrt{\log\log(n)}\sup\bigl\{H_n(v) : v \in [c_{1,n}, c_{2,n}] \cup [c_{3,n}, c_{4,n}]\bigr\} \to 0, \qquad \textit{where} \tag{3.3}$$

$$\frac{\log(c_{1,n})}{\log(n)} \to -1, \qquad \frac{\log(c_{4,n})}{\log(n)} \to 0 \quad \textit{and} \quad \frac{\log(c_{2,n})}{\log(n)}, \frac{\log(c_{3,n})}{\log(n)} \to -\kappa. \tag{3.4}$$

Then the distributional convergence in (2.2) holds under the alternative $\mathcal{H}_{n,1}$.

By Theorems 2.1 and 3.2 all tests of the form $\varphi_n(s) = \mathbf{1}\{S_n(s) > c_n(s)\}$ cannot distinguish between $\mathcal{H}_{0,n}$ and $\mathcal{H}_{1,n}$ asymptotically if (3.3) and (3.4) are fulfilled. We explain briefly why the supremum can be taken over $C_n = [c_{1,n}, c_{2,n}] \cup [c_{3,n}, c_{4,n}]$ instead of over the whole interval $[0, 1/2]$. The supremum in (2.1) is taken neither by x from the extreme tails nor from the middle. In fact, the supremum is take by the so-called intermediate values [23,24], i.e., by x with $F_0(x) \in C_n = [c_{1,n}, c_{2,n}] \cup [c_{3,n}, c_{4,n}]$ or $1 - F_0(x) \in C_n$. To be more concrete, we verify that under $\mathcal{H}_{0,n}$ (and so under $\mathcal{H}_{1,n}$ due to mutually contiguity)

$$n\Bigl(\sup_{X_{1:n}\le x < X_{n:n}: F_0(x), 1-F_0(x) \notin C_n} K_s\bigl(\mathbb{F}_n(x), F_0(x)\bigr)\Bigr) - r_n \to -\infty$$

in probability. We want to point out that the restriction to intermediate values in (3.3) is necessary for applying Theorem 3.2 to the majority of the examples.

Last, we give a simplification of our tool for the sparse case ($n\varepsilon_n^2 \to 0$).

Remark 3.3 (Simplification for the sparse case). Typically, in the sparse case we even have $\sqrt{\log\log(n)}n\varepsilon_n^2 \to 0$. Then it is easy to see that the statements in Theorems 3.1 and 3.2 remain true if $H_n(v)$ is replaced by

$$\widetilde{H}_n(v) = \sqrt{n}\varepsilon_n v^{-1/2}\bigl(\mu_n^{F_0}(0, v] + \mu_n^{F_0}[1 - v, 1)\bigr).$$

4. Application

4.1. Extension of Cai and Wu

Throughout this section, we consider (only) the sparse case

$$\varepsilon_n = n^{-\beta}, \qquad \beta \in \left(\frac{1}{2}, 1\right].$$

Starting with a fixed noise distribution P_0 and a fixed sequence $(\mu_n)_{n\in\mathbb{N}}$ of signal distributions, Cai and Wu [9] developed a technique to calculate a detection threshold $\beta^{\#}$ for the parameter β.

(i) (Undetectable) If β exceeds $\beta^{\#}$, then there is no sequence of tests, which can distinguish between $\mathcal{H}_{0,n}$ and $\mathcal{H}_{1,n}$ asymptotically.
(ii) (Completely detectable) If β is smaller than $\beta^{\#}$, then there is a sequence of likelihood ratio tests, which can completely separate $\mathcal{H}_{0,n}$ and $\mathcal{H}_{1,n}$ asymptotically.

In their paper they discussed two different cases, standard normal distributed and general noise. For standard normal distributed noise, they showed the optimality of the higher criticism test, that is, it can completely separate $\mathcal{H}_{0,n}$ and $\mathcal{H}_{1,n}$ asymptotically if $\beta < \beta^{\#}$. At the section's end, we discuss this situation more closely. Up until now, it was an open problem whether the higher criticism test is optimal in the general noise setting. After introducing the calculation technique we give a positive answer to this open question, where the optimality is not restricted to the higher criticism test but holds even for the whole family $\{S_n(s) : s \in \mathbb{R}\}$.

Proposition 4.1 (Theorem 3 in [9]). *Define for all* $t > 0$

$$h_{n,1}(t) = \log\left(\frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}\big(F_0^{-1}\big(n^{-t}\big)\big)\right), \qquad h_{n,2}(t) = \log\left(\frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}\big(F_0^{-1}\big(1-n^{-t}\big)\big)\right) \quad \textit{and}$$
$$h_n(t) = \max\big\{h_{n,1}(t), h_{n,2}(t)\big\}.$$

Suppose that

$$\sup\left\{\left|\frac{h_n(t)}{\log(n)} - \gamma(t)\right| : t \geq \frac{\log(2)}{\log(n)}\right\} \to 0 \tag{4.1}$$

for a measurable $\gamma : [0, \infty) \to \mathbb{R}$. *Then the detection threshold for* β *is given by*

$$\beta^{\#} = \frac{1}{2} + \operatorname*{ess\,sup}_{t\geq 0}\left\{\gamma(t) - t + \frac{\min(t,1)}{2}\right\}. \tag{4.2}$$

The simple scale exponential distribution model $P_0 = \mathrm{Exp}(1)$ and $\mu_n = \mathrm{Exp}(1 + n^r)$ $(r > 0)$ cannot be treated by Proposition 4.1 since $h_n(t)/\log(n)$ tends to $-\infty$ for $t \leq r - \delta$ for every $\delta > 0$. We relax the uniform convergence condition (4.1) such that this example and further interesting models can be treated. Moreover, we prove the optimality of the whole Φ-divergence test family.

Theorem 4.2 (Extension of Proposition 4.1). *Let h_n be defined as in Proposition 4.1. Assume that there exists some $\beta^* \in \mathbb{R}$ such that for every $\delta > 0$*

$$\liminf_{n\to\infty} \lambda\!\!\lambda\left(t \geq \frac{\log(2)}{\log(n)} : \beta^* - \delta - \frac{1}{2} \leq \frac{h_n(t)}{\log(n)} - t + \frac{\min\{t,1\}}{2}\right) > 0 \quad \text{and} \tag{4.3}$$

$$\lambda\!\!\lambda\left(t \geq \frac{\log(2)}{\log(n)} : \beta^* + \delta - \frac{1}{2} \leq \frac{h_n(t)}{\log(n)} - t + \frac{\min\{t,1\}}{2}\right) = 0 \tag{4.4}$$

for all sufficiently large $n \geq N_{1,\delta}$, where $\lambda\!\!\lambda$ denotes the Lebesgue measure. Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $(0,\infty)$ such that $\lambda_n \to 0$ and $\lambda_n n^\kappa \to \infty$ for all $\kappa > 0$. Suppose that for some $M \geq 1$:

$$\lim_{n\to\infty} \sup_{t\geq M} \left| \frac{h_n(t)}{\log(n)} - \gamma(t) \right| = 0 \tag{4.5}$$

for some $\gamma : (0,\infty) \to \mathbb{R}$ or for every $\delta > 0$ there exists $N_{2,\delta} \in \mathbb{N}$ such that

$$\lambda\!\!\lambda\left(t \geq M : \beta^* + \delta - 1 \leq \frac{h_n(t)}{\log(n)} - \left(1 - \frac{\lambda_n}{\log(n)}\right)t\right) = 0 \tag{4.6}$$

for all $n \geq N_{2,\delta}$. Then $\beta^\# = \beta^*$. Moreover, if $\beta < \beta^\#$ then for every $s \in \mathbb{R}$ there is a sequence $(c_n(s))_{n\in\mathbb{N}}$ of critical values such that $\varphi_n = \mathbf{1}\{S_n(s) > c_n(s)\}$ can completely separate $\mathcal{H}_{0,n}$ and $\mathcal{H}_{1,n}$ asymptotically.

The conditions (4.3) and (4.4) together are mimicking the essential supremum in (4.2), where γ is replaced by $h_n/\log(n)$. The uniform convergence of $h_n/\log(n)$ is not necessarily needed anymore. It can easily be checked that the conditions of Proposition 4.1 imply the ones of our Theorem 4.2. As already explained by Cai and Wu [9] Proposition 4.1 can be used to the derive the detection boundary for

- the general Gaussian location mixture model, where

$$\frac{\mathrm{d}P_0}{\mathrm{d}\lambda\!\!\lambda}(x) = \frac{\tau}{2\Gamma(\tau)} \exp\bigl(-|x|^\tau\bigr) \quad \text{and} \quad \frac{\mathrm{d}\mu_n}{\mathrm{d}\lambda\!\!\lambda}(x) = \frac{\mathrm{d}P_0}{\mathrm{d}\lambda\!\!\lambda}(x - \vartheta_n)$$

  for some shape parameter $\tau > 0$ and a shift $\vartheta_n = (r\log(n))^{1/\tau}$.
- the exponential-χ^2-mixture model, where $P_0 = \mathrm{Exp}(2)$ is the exponential distribution with scale parameter 2 and $\mu_n = \chi_2^2(\vartheta_n)$ is the (non-central) χ^2-distribution with 2 degrees of freedom and non-centrality parameter $\vartheta_n = 2r\log(n)$.

For the concrete detection boundaries we refer to Donoho and Jin [17], who already postulated the optimality of the higher criticism test for both examples. From our Theorem 4.2, we obtain the optimality for the whole family $\{S_n(s) : s \in \mathbb{R}\}$.

As mentioned above, the simple scale exponential distribution model and even more general exponential family models can be treated now. Recall that a function L is *slowly varying* at infinity if $\lim_{x\to\infty} L(\lambda x)/L(x) = 1$ for all $\lambda > 0$.

Theorem 4.3. *Let $(P_{(\vartheta)})_{\vartheta\in[0,\infty)}$ be a family of continuous distributions on $[0,\infty)$ with $P_{(\vartheta)} \ll P_{(0)}$ and Radon–Nikodym density*

$$\frac{\mathrm{d}P_{(\vartheta)}}{\mathrm{d}P_{(0)}} = C(\vartheta)\exp(\vartheta T) \tag{4.7}$$

for appropriate functions $T : [0,\infty) \to \mathbb{R}$ and $C : [0,\infty) \to (0,\infty)$ with $C(0) = 1$. Suppose that T is strictly decreasing on $[0,\eta]$ for some $\eta > 0$, $T(\eta) \geq T(x)$ for all $x \geq \eta$ and

$$T\big(F_0^{-1}(0)\big) - T\big(F_0^{-1}(u)\big) = u^{\frac{1}{p}} L\left(\frac{1}{u}\right) \qquad \text{as } u \searrow 0$$

for a slowly varying function L at infinity, where F_0^{-1} is the left continuous quantile function of $P_{(0)}$. Let $P_0 = P_{(0)}$ be the noise distribution and $\mu_n = P_{(\vartheta_n)}$ the signal distribution with signal strength $\vartheta_n \sim_{n\to\infty} n^r$ for $r > 0$, that is, $\vartheta_n/n^r \to 1$. Then the conditions of Theorem 4.2 are fulfilled for

$$\beta^{\#} = \beta^{\#}(r,p) = \frac{\min\{rp,1\}+1}{2}.$$

We refer the reader to Corollary 5.7 and Theorem 8.19 of Ditzhaus [15] for a discussion about the tests' power behavior on the detection boundary. In short, the likelihood ratio test has nontrivial power on the detection boundary, whereas the higher criticism test has no power. By Theorem 3.2, the latter can be extended to the whole family $\{S_n(s) : s \in \mathbb{R}\}$. We want to point out that Theorem 4.3 can be used for exponential families $(P_{(\vartheta)})_{\vartheta\geq 0}$ on $\mathbb{R}$ by applying it to $(P_{(\vartheta)}^{F_0})_{\vartheta\geq 0}$ or $(P_{(\vartheta)}^{1-F_0})_{\vartheta\geq 0}$ instead, where F_0 is the distribution function of $P_0 = P_{(0)}$. An example for this situation is the location Gumbel model. In the following, we give the detection boundary for this and two other models as an immediate consequence of Theorem 4.3.

Corollary 4.4. *Fix $r > 0$. Consider one of the following three models* (a)–(c):

(a) *the Gumbel location model with parameter $\theta_n = r\log(n)$*

$$\frac{\mathrm{d}P_0}{\mathrm{d}\lambda}(x) = \exp\big(-x - e^{-x}\big) \quad \text{and} \quad \frac{\mathrm{d}\mu_n}{\mathrm{d}\lambda}(x) = \frac{\mathrm{d}P_0}{\mathrm{d}\lambda}(x-\theta_n) \qquad (x\in\mathbb{R}). \tag{4.8}$$

(b) *the scale Fréchet model with shape parameter $\alpha > 0$ and scale parameter $\theta_n = n^{r/\alpha}$*

$$\frac{\mathrm{d}P_0}{\mathrm{d}\lambda}(x) = \alpha x^{-\alpha-1}\exp\big(-x^{-\alpha}\big) \quad \text{and} \quad \frac{\mathrm{d}\mu_n}{\mathrm{d}\lambda}(x) = \frac{1}{\theta_n}\frac{\mathrm{d}P_0}{\mathrm{d}\lambda}\left(\frac{x}{\theta_n}\right) \qquad (x>0). \tag{4.9}$$

(c) *the scale exponential distribution model $P_0 = \mathrm{Exp}(1)$ and $\mu_n = \mathrm{Exp}(1+n^r)$.*

In all three cases the detection boundary is given by

$$\beta^{\#}_{\mathrm{GFE}}(r) := \frac{\min(r,1)+1}{2}.$$

The whole family $\{S_n(s) : s \in \mathbb{R}\}$ is optimal in the sense that the null and the alternative can be completely separated asymptotically if $\beta < \beta^{\#}(r)$.

Last, we give our generalization of Theorems 1 and 4 of Cai and Wu [9] concerning (only) normal distributed noise, which we mentioned at the section's beginning.

Theorem 4.5. *Let $P_0 = N(0,1)$. Define for all $x > 0$*

$$\widetilde{h}_n(x) = \log\left(\frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}\big(x\sqrt{2\log(n)}\big)\right).$$

Suppose that there is some $\beta^* \in \mathbb{R}$ such that for every $\delta > 0$

$$\liminf_{n\to\infty} \lambda\!\!\lambda\left(x \in \mathbb{R} : \beta^* - \delta - \frac{1}{2} \le \frac{\widetilde{h}_n(x)}{\log(n)} - x^2 + \frac{\min\{x^2, 1\}}{2}\right) > 0 \quad \text{and} \tag{4.10}$$

$$\lambda\!\!\lambda\left(x \in \mathbb{R} : \beta^* + \delta - \frac{1}{2} \le \frac{\widetilde{h}_n(x)}{\log(n)} - x^2 + \frac{\min\{x^2, 1\}}{2}\right) = 0 \tag{4.11}$$

for all sufficiently large $n \ge N_{1,\delta}$. Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $(0,\infty)$ such that $\lambda_n \to 0$ and $\lambda_n n^{\kappa} \to \infty$ for all $\kappa > 0$. Suppose that for some $M \ge 1$:

$$\lim_{n\to\infty} \sup_{|x|\ge M} \left| \frac{\widetilde{h}_n(x)}{\log(n)} - \alpha(x) \right| = 0 \tag{4.12}$$

for some $\alpha : (0,\infty) \to \mathbb{R}$ or for every $\delta > 0$ there exists $N_{2,\delta} \in \mathbb{N}$ such that

$$\lambda\!\!\lambda\left(|x| \ge M : \beta^* + \delta - 1 \le \frac{\widetilde{h}_n(x)}{\log(n)} - \left(1 - \frac{\lambda_n}{\log(n)}\right)x^2\right) = 0 \tag{4.13}$$

for all $n \ge N_{2,\delta}$. Then $\beta^{\#} = \beta^*$. Moreover, if $\beta < \beta^{\#}$ then for every $s \in \mathbb{R}$ there is a sequence $(c_n(s))_{n\in\mathbb{N}}$ of critical values such that $\varphi_n = \mathbf{1}\{S_n(s) > c_n(s)\}$ can completely separate $\mathcal{H}_{0,n}$ and $\mathcal{H}_{1,n}$ asymptotically.

The results concerning the (heterogeneous) normal location model mentioned in Section 1 follow immediately from the previous theorem. More generally, Theorem 4.5 can be applied to convolution normal models $P_0 = N(0,1)$ and $\mu_n = \widetilde{\mu}_n * N(0,1)$ (compare to Corollary 1 and Section V-B in [9]), where $*$ denotes the convolution operation. An example for this convolution idea is

- the heteroscedastic normal location model $P_0 = N(0,1)$ and $\mu_n = N(\vartheta_n, \sigma_0^2)$ with signal strength $\sqrt{2r\log(n)}$, where the signal variance $\sigma_0^2 \in [0,\infty)$ may differ from 1.

Cai et al. [8] already verified the optimality of the higher criticism test for this model. By Theorem 4.5, this optimality can now be extended to the whole family. Moreover, Ditzhaus and Janssen [16] showed that the higher criticism test has no asymptotic power on the detection

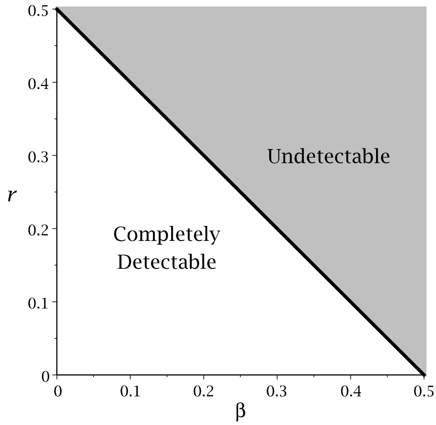

Figure 2. The detection boundary for dense exponential family mixtures is plotted, see Theorem 4.6.

boundary for the heterogeneous normal mixture model, whereas the likelihood ratio test has non-trivial power there. By Theorem 3.2 this negative result can be transferred to the whole family as well.

4.2. Dense exponential family

Beside the sparse heteroscedastic normal mixture model mentioned at the end of the previous section Cai et al. [8] also studied the dense case with vanishing signal strength $\vartheta_n \to 0$ and showed the optimality of the higher criticism test. Their result can be extended to general exponential families and to the whole Φ-divergence test family.

Theorem 4.6. *Let $(P_{(\vartheta)})_{\vartheta\in[0,\infty)}$ be a family of continuous distributions on $\mathbb{R}$ with Radon–Nikodym density given by (4.7) for $T:\mathbb{R}\to\mathbb{R}$ and $C:\mathbb{R}\to(0,\infty)$ with $C(0)=1$. Assume that $\mathrm{Var}_{P_{(0)}}(T)>0$. Consider the noise distribution $P_0=P_{(0)}$ and the signal distribution $\mu_n=P_{(\vartheta_n)}$ with the parametrization*

$$\varepsilon_n = n^{-\beta}, \qquad \beta\in\left(0,\frac{1}{2}\right) \quad \text{and} \quad \vartheta_n \sim_{n\to\infty} n^{-r}, \qquad r>0,$$

then the detection boundary for the parameter r is given by

$$\rho^*(\beta) = \frac{1}{2} - \beta, \tag{4.14}$$

see Figure 2. In particular, we have for all $s\in\mathbb{R}$:

(a) *If $r<\rho^*(\beta)$ then there is a sequence $(c_n(s))_{n\in\mathbb{N}}$ of critical values such that $\varphi_n = \mathbf{1}\{S_n(s)>c_n(s)\}$ can completely separate P_0^n and Q_n^n asymptotically.*

(b) *Suppose that* $r = \rho^*(\beta)$. *Then* $2N(0,1)((1/2)\mathrm{Var}_{P_{(0)}}(T)^{1/2}, \infty)$ *is the lower bound of the limit of the sum of type* 1 *and* 2 *error probabilities for all tests testing* P_0^n *versus* Q_n^n *and is attained by a likelihood ratio test sequence. But all tests of the form* $\varphi_n = \mathbf{1}\{S_n(s) > c_n(s)\}$ *cannot distinguish between* P_0^n *and* Q_n^n *asymptotically.*

(c) *If* $r > \rho^*(\beta)$ *then no test* φ_n *can distinguish between* P_0^n *and* Q_n^n *asymptotically.*

The detection boundary (4.14) and the statements (a)–(c) are valid, among others, for:

- the heterogeneous normal model $P_0 = N(0,1)$ and $\mu_n = N(n^{-r}, 1)$.
- the location Gumbel model (4.8) with $\theta_n = n^{-r}$.
- the scale Fréchet model (4.9) with shape parameter $\alpha > 0$ and $\theta_n = (1 + n^{-r})^{1/\alpha}$.
- the scale exponential distribution model $P_0 = \mathrm{Exp}(1)$ and $\mu_n = \mathrm{Exp}(1 + n^{-r})$.

4.3. Spike chimeric alternatives

The prime example of Ditzhaus and Janssen [16] was a p-value model inspired by the spike chimeric alternatives of Khmaladze [35]. If signals are present the corresponding p-values are usually small. Hence, it is reasonable to restrict the support of μ_n to the shrinking interval $(0, \kappa_n)$ with $\kappa_n \to 0$. Let $P_0 = \lambda\!\!\lambda_{|(0,1)}$ and let h be some Lebesgue-density on $(0,1)$ with $\int_0^1 h^2 \, \mathrm{d}\lambda\!\!\lambda < \infty$. Then we define the signal distribution by its rescaled density

$$\frac{\mathrm{d}\mu_n}{\mathrm{d}\lambda\!\!\lambda}(x) = \kappa_n^{-1} h\left(\frac{x}{\kappa_n}\right)\mathbf{1}\big\{x \in (0, \kappa_n)\big\} \qquad \big(x \in (0,1)\big).$$

Using the parametrization

$$\kappa_n = n^r \qquad (r > 0) \quad \text{and} \quad \varepsilon_n = n^{-\beta} \qquad \left(\beta \in \left(\frac{1}{2}, 1\right)\right)$$

Ditzhaus and Janssen [16] calculated the detection boundary

$$\rho(\beta) = 2\beta - 1 \qquad \left(\beta \in \left(\frac{1}{2}, 1\right)\right),$$

see Figure 3, for the parameter r and proved the optimality of the higher criticism test. Moreover, they verified that the higher criticism test has no asymptotic power on the detection boundary while the likelihood ratio test has nontrivial power there. Since our Theorems 3.1 and 3.2 are extensions of their Theorems 3.1 and 3.2 we can immediately transfer the optimality and the negative result concerning the power on the boundary to the whole family $\{S_n(s) : s \in \mathbb{R}\}$.

5. Discussion

The higher criticism test became quite popular recently. In this paper, we showed that the (enlarged) ϕ-divergence test family $\{S_n(s) : s \in \mathbb{R}\}$ of Cai and Wu [9] shares the higher criticism

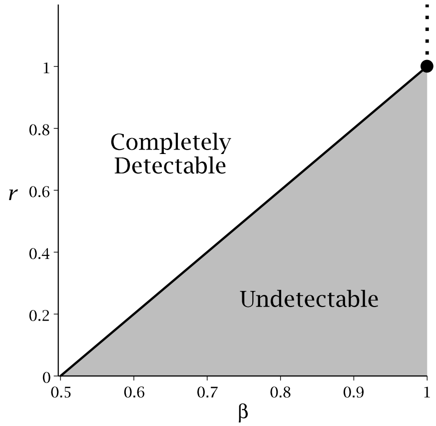

Figure 3. The detection boundary for the spike chimeric alternatives.

test's optimality under various model assumptions. The advantage of a whole test family is more flexibility in choosing a test statistic which suits the specific problem best. Jager and Wellner [31] already pointed out that $S_n(s)$ is more sensible for signal distributions with heavy or light tails if $s \geq 1$ or $s \leq 0$, respectively. As a good compromise they suggested their "new" $S_n(1/2)$. In future we wish to conduct a detailed simulation study in order to give a better advice for practitioners how to choose "the best" s.

Besides the detection problem, a more in-depth analysis of the data such as feature selection, classification and estimation of the signal amount is of great interest. The detection problem discussed in this paper is closely related to these problems [18,19,25,32] and the higher criticism statistic can be applied to them as well. Our results motivate a future investigation whether the whole class $\{S_n(s) : s \in \mathbb{R}\}$ can be used for these problems.

6. Proofs

6.1. Preliminaries

To prove Theorem 2.1, we use some results of Chang [11] and Wellner [45] about the asymptotic behavior of the empirical distribution function. We summarize them in the following lemma.

Lemma 6.1. *Let $X_{n,1}, \dots, X_{n,n}$ be independent and identical distributed random variables on the same probability space $(\Omega, \mathcal{A}, P)$ with continuous distribution function F_n. Let $\mathbb{F}_n$ be the corresponding empirical distribution function. Let $(d_n)_{n\in\mathbb{N}}$ be a decreasing sequence in $\mathbb{R}$, that is, $d_n > d_{n+1}$, such that $nF_n(d_n) \to \infty$. Then*

$$\sup_{d_n \leq x < \infty} \left| \frac{\mathbb{F}_n(x)}{F_n(x)} - 1 \right| \xrightarrow{P} 0. \tag{6.1}$$

If additionally $c_n = nF(d_n)/\log\log(n) \to \infty$, *then*

$$\sqrt{c_n} \sup_{d_n \le x < \infty} \left| \frac{\mathbb{F}_n(x)}{F_n(x)} - 1 \right| \xrightarrow{P} \sqrt{2}. \tag{6.2}$$

Moreover, for all $t \in \mathbb{R}$

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} P\big(B^c_{n,\lambda,t}\big) = 0 \qquad \textit{with } B_{n,\lambda,t} = \left\{ \sup_{x\in(X_{1:n},\infty)} \left(\frac{\mathbb{F}_n(x)}{x} \right)^t < \lambda \right\}. \tag{6.3}$$

Proof. First, suppose that $X_{n,1}, \dots, X_{n,n}$ are uniformly distributed on $(0, 1)$. Then (6.1) was stated by Chang [11], see also Theorem 0 of Wellner [45], and (6.3) follows by combining (i) and (ii) of Remark 1 of Wellner [45]. Moreover, (6.2) follows from Theorem 1S of Wellner [45]. For general continuous distribution, note that $F_n(X_{n,1}), \dots, F_n(X_{n,n})$ are independent and uniformly distributed random variables in $(0, 1)$. Consequently, it is easy to check that the statements for general distributions can be concluded from the ones for uniform distributions. □

In Section 4, we made several statements about the general detectability of signals. For this purpose, we have to study the best tests for the underlying testing problem, namely likelihood ratio tests. Let us first introduce the variational distance $\|P - Q\|$ between two probability measures P and Q on the same measure space:

$$\|P - Q\| = \frac{1}{2} \int \left| \frac{\mathrm{d}P}{\mathrm{d}(P+Q)} - \frac{\mathrm{d}Q}{\mathrm{d}(P+Q)} \right| \mathrm{d}(P+Q).$$

By Lemmas 2.2 and 2.3 of Strasser [41], the lower bound of the sum of type 1 and 2 error probabilities for all tests testing P_0^n versus Q_n^n equals $1 - \|P_0^n - Q_n^n\|$. Moreover, this bound is attained by the likelihood ratio test $\varphi_n = \mathbf{1}\{\mathrm{d}Q_n^n/\mathrm{d}P_0^n \ge 1\}$. It is well known and easy to show that weak convergence of binary experiments $\{P_0^n, Q_n^n\} \xrightarrow{w} \{P, Q\}$ implies convergence of the variational distance $\|P_0^n - Q_n^n\| \to \|P - Q\|$. Let us recall that $\{P_0^n, Q_n^n\} \xrightarrow{w} \{P, Q\}$ if and only if $\mathcal{L}(\mathrm{d}Q_n^n/\mathrm{d}P_0^n|P_0^n)$ tends weakly to $\mathcal{L}(\mathrm{d}Q/\mathrm{d}P|P)$, or equivalently $\mathcal{L}(\mathrm{d}Q_n^n/\mathrm{d}P_0^n|Q_n^n)$ converges weakly to $\mathcal{L}(\mathrm{d}Q/\mathrm{d}P|Q)$, where $\mathcal{L}(T|P) = P^T$ denotes the image measure. For more details about the convergence of binary or more general experiments, we refer the reader to Strasser [41]. Ditzhaus and Janssen [16] studied the asymptotic behavior of likelihood ratio tests for our testing problem rigorously. Below we present a simplification of their Theorem A.1.

Proposition 6.2 (Ditzhaus and Janssen [16]). *Let* $\sigma \ge 0$ *and* $\tau > 0$. *If*

$$I_{n,1,x} = n\varepsilon_n \mu_n\left(\varepsilon_n \frac{\mathrm{d}\mu_n}{\mathrm{d}P_0} > x \right) \to 0 \qquad \textit{for all } x > 0 \quad \textit{and} \tag{6.4}$$

$$I_{n,2,\tau} = n\varepsilon_n^2 E_{P_0}\left(\left(\frac{\mathrm{d}\mu_n}{\mathrm{d}P_0} \right)^2 \mathbf{1}\left\{ \varepsilon_n \frac{\mathrm{d}\mu_n}{\mathrm{d}P_n} \le \tau \right\} - 1 \right) \to \sigma^2 \tag{6.5}$$

then $\{P_0^n, Q_n^n\} \xrightarrow{w} \{N(-\sigma^2/2, \sigma^2), N(\sigma^2/2, \sigma^2)\}$, *where* $N(0, 0) = \epsilon_0$ *is the Dirac measure centred in* 0.

Under the conditions of Proposition 6.2 the lower bound of the error probabilities sum for all tests testing P_0^n versus Q_n^n tends to $1 - \|N(-\sigma^2/2, \sigma^2), N(\sigma^2/2, \sigma^2)\| = 2N(0,1)(\sigma/2, \infty)$. In particular, $\sigma = 0$ implies that no test sequence can asymptotically distinguish between P_0^n and Q_n^n since the error probabilities sum converges to 1 for all test sequences.

6.2. Proofs of our statements

Proof of Theorem 2.1. Having a transformation to p-values $p_{n,i} = F_0(X_{n,i})$ or $p_{n,i} = 1 - F_0(X_{n,i})$ in mind we can assume without loss of generality that $P_0 = \lambda\!\!\!\lambda_{|(0,1)}$ is the uniform distribution on the interval $(0,1)$. The proof is based, as the one of Theorem 3.1 of Jager and Wellner [31], on a Taylor expansion of $u \mapsto K_s(u,v)$ around $u = v$. It is easy to verify that

$$\frac{\partial}{\partial u} K_s(u,v)_{|u=v} = 0 = K_s(v,v), \qquad \frac{\partial^2}{\partial^2 u} K_s(u,v)_{|u=v} = \frac{1}{v(1-v)} \quad \text{and}$$
$$\frac{\partial^3}{\partial^3 u} K_s(u,v) = \frac{(s-2)}{v^2}\left(\frac{u}{v}\right)^{s-3} - \frac{(s-2)}{(1-v)^2}\left(\frac{1-u}{1-v}\right)^{s-3}.$$

Hence, by a careful calculation we obtain for all $x \in [X_{1:n}, X_{n:n})$

$$\begin{aligned} K_s\big(\mathbb{F}_n(x), x\big) &= K_2\big(\mathbb{F}_n(x), x\big)\left(1 + \frac{(s-2)}{3} R_{n,x,s}\right) \qquad \text{with} \\ R_{n,x,s} &= \frac{(\mathbb{F}_n(x) - x)}{x(1-x)}\left((1-x)^2\left(\frac{\mathbb{F}^*_{n,x}}{x}\right)^{s-3} - x^2\left(\frac{1-\mathbb{F}^*_{n,x}}{1-x}\right)^{s-3}\right), \end{aligned} \tag{6.6}$$

where $\mathbb{F}^*_{n,x}$ is a random variable satisfying $\min\{\mathbb{F}_n(x), x\} \le \mathbb{F}^*_{n,x} \le \max\{\mathbb{F}_n(x), x\}$. Clearly, $t \to t^{3-s}$ $(t > 0)$ is monotone. Thus, $|R_{n,x,s}| \le R^{(1)}_{n,x,s} + R^{(2)}_{n,x,s}$, where

$$R^{(1)}_{n,x,s} = \frac{|\mathbb{F}_n(x) - x|}{x} \max\left\{1, \left(\frac{\mathbb{F}_n(x)}{x}\right)^{s-3}\right\} \quad \text{and}$$
$$R^{(2)}_{n,x,s} = \frac{|\mathbb{F}_n(x) - x|}{1-x} \max\left\{1, \left(\frac{1-\mathbb{F}_n(x)}{1-x}\right)^{s-3}\right\}.$$

Let $d_n = n^{-1}(\log n)^5$. Obviously, $P_0^n(X_{1:n} > d_n) + P_0^n(X_{n:n} < 1 - d_n) = 2(1-d_n)^n \to 0$. By (6.6)

$$\begin{aligned} &\big|nK_s\big(\mathbb{F}_n(x), x\big) - r_n - \big(nK_2\big(\mathbb{F}_n(x), x\big) - r_n\big)\big| \\ &\quad \le \big(R^{(1)}_{n,x,s} + R^{(2)}_{n,x,s}\big)\big(\big|nK_2\big(\mathbb{F}_n(x), x\big) - r_n\big| + r_n\big). \end{aligned}$$

Consequently, for (2.2) it is sufficient to show that

$$n\Big(\sup_{x \in (d_n, 1-d_n)} K_2\big(\mathbb{F}_n(x), x\big)\Big) - r_n \xrightarrow{\text{d}} Y \qquad \text{under } P_0^n, \tag{6.7}$$

$$r_n \sup_{x\in(d_n,1-d_n)} R^{(j)}_{n,x,s} \xrightarrow{P_0^n} 0 \qquad \text{for } j=1,2 \quad \text{and} \tag{6.8}$$

$$I_{n,s} = n\Big(\sup_{x\in(X_{1:n},d_n)\cup(1-d_n,X_{n:n})} K_s\big(\mathbb{F}_n(x),x\big)\Big) - r_n \xrightarrow{P_0^n} -\infty. \tag{6.9}$$

Note that

$$I_{n,s} \le r_n\bigg(-1+\frac{n}{r_n}\sup_{x\in(X_{1:n},d_n)\cup(1-d_n,X_{n:n})} K_2\big(\mathbb{F}_n(x),x\big)\big(1+R^{(1)}_{n,x,s}+R^{(2)}_{n,x,s}\big)\bigg).$$

Hence, using the inequality

$$P(A_n) \le P(A_n\cap B_{n,\lambda}) + P\big(B^c_{n,\lambda}\big) \tag{6.10}$$

with appropriate sets we can deduce (6.9) from (6.6) if

$$\frac{n}{r_n}\sup_{x\in(X_{1:n},d_n)\cup(1-d_n,X_{n:n})} K_2\big(\mathbb{F}_n(x),x\big) \xrightarrow{P_0^n} 0 \quad \text{and} \tag{6.11}$$

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty} P_0^n\Big(\big(B^{(j)}_{n,\lambda}\big)^c\Big) = 0 \qquad \text{for both } j\in\{1,2\}, \tag{6.12}$$

$$\text{where } B^{(j)}_{n,\lambda} = \Big\{\sup_{x\in(X_{1:n},d_n)\cup(1-d_n,X_{n:n})} R^{(j)}_{n,x,s} < \lambda\Big\}.$$

Thus, it remains to verify (6.7), (6.8), (6.11) and (6.12). Due to symmetry it is sufficient to show (6.8) and (6.12) for $j=1$. Using again (6.10) we obtain (6.8) for $j=1$ from (6.2) and (6.3) setting $t=s-3$ since $c_n = nd_n/\log\log(n) = \log(n)^5/\log\log(n)\to\infty$ and $\sqrt{c_n}/r_n\to\infty$. Moreover, from the inequality

$$R^{(1)}_{n,x,s} \le 1+\frac{\mathbb{F}_n(x)}{x}+\Big(\frac{\mathbb{F}_n(x)}{x}\Big)^{s-2}+\Big(\frac{\mathbb{F}_n(x)}{x}\Big)^{s-3}$$

we can conclude (6.12) for $j=1$ by applying (6.3) for $t\in\{1,s-2,s-3\}$.

To prove the remaining (6.7) and (6.11) we introduce the supremum statistic of the normalized uniform empirical process

$$\mathbb{Z}_n(a,b) = \sup_{a<x<b}\sqrt{n}\frac{|\mathbb{F}_n(x)-x|}{\sqrt{x(1-x)}} \qquad \big(a,b\in(0,1)\big) \tag{6.13}$$

studied by Jaeschke [29], see also Chapter 16 of Shorack and Wellner [40]. In particular, by (19), (20), (25), (26) and (g) in Section 16.1 from Shorack and Wellner [40] and the symmetry $\mathbb{Z}_n(0,a) \stackrel{d}{=} \mathbb{Z}_n(1-a,1)$ we obtain

$$\frac{\mathbb{Z}_n(d_n,1-d_n)}{b_n} \xrightarrow{P_0^n} 1 \quad \text{and} \quad b_n\mathbb{Z}_n(d_n,1-d_n)-c_n \xrightarrow{\text{d}} Y \qquad \text{under } P_0^n, \tag{6.14}$$

$$b_n^{-1}\mathbb{Z}_n(0,d_n) \xrightarrow{P_0^n} 0 \quad \text{and} \quad b_n^{-1}\mathbb{Z}_n(1-d_n,1) \xrightarrow{P_0^n} 0, \qquad \text{where} \tag{6.15}$$

$$b_n = \sqrt{2\log\log(n)} \quad \text{and} \quad c_n = b_n^2 + \frac{1}{2}\log\log\log(n) - \frac{1}{2}\log(4\pi). \tag{6.16}$$

Observe that

$$n\Big(\sup_{x\in(a,b)} K_2\big(\mathbb{F}_n(x),x\big)\Big) - r_n$$

$$= \frac{1}{2}\mathbb{Z}_n(a,b)^2 - r_n \tag{6.17}$$

$$= \frac{1}{2}\big(b_n\mathbb{Z}_n(a,b) - c_n\big)\left(\frac{\mathbb{Z}_n(a,b)}{b_n} + \frac{c_n}{b_n^2}\right) + \left(\frac{1}{2}\frac{c_n^2}{b_n^2} - r_n\right) \qquad \big(a,b\in(0,1)\big). \tag{6.18}$$

Finally, (6.7) follows from (6.18) and (6.14). Furthermore, we conclude (6.11) from (6.15) and (6.17) since $b_n^2/r_n \to 2$. □

Proof of Theorem 3.1. Similar to the previous proof we can assume that $P_0 = \lambda\!\!\!\lambda_{|(0,1)}$ is the uniform distribution on $(0,1)$. Set $l_n = \log\log(n)$. Due to symmetry it is sufficient to give the proof under the assumption

$$n^{1/2}\varepsilon_n v_n^{-1/2} l_n^{-1}\big(\mu_n(0,v_n] - v_n\big) \to A \in \{-\infty,\infty\}. \tag{6.19}$$

Let G_n be the distribution function of Q_n defined by $G_n(v) = v + \varepsilon_n(\mu_n(0,v]-v)$ $(v\in(0,1))$. If $A=\infty$, then it is easy to see that

$$l_n^{-1} n G_n(v_n) \to \infty. \tag{6.20}$$

Moreover, $A=-\infty$ implies $l_n^{-1}\sqrt{nv_n}\to\infty$ and, thus, $l_n^{-1}nv_n(1-\varepsilon_n)\to\infty$. In all, (6.20) holds in both cases $A\in\{-\infty,\infty\}$. Due to (6.20), we obtain from Lemma 6.1 that

$$\frac{\mathbb{F}_n(v_n)}{G_n(v_n)} \to 1 \qquad \text{in } Q_n^n\text{-probability}. \tag{6.21}$$

From (6.20) we deduce $Q_n^n(X_{1:n} \le v_n) = 1-(1-G_n(v_n))^n \to 1$ and $v_n < 1/2$ implies $Q_n^n(X_{n:n} > v_n) \to 1$. Hence, $S_n(s) \ge K_s(\mathbb{F}_n(v_n), v_n) \ge v_n\phi_s(\mathbb{F}_n(v_n)/v_n)$ with probability tending to one. Combining this and $r_n l_n^{-1}\to 1$ we can conclude that (3.2) follows from

$$Q_n^n\left(\frac{nv_n}{l_n}\phi_s\left(\frac{\mathbb{F}_n(v_n)}{v_n}\right) > 2\right) \to 1. \tag{6.22}$$

Consequently, it is sufficient to verify (6.22). Since $v_n^{-1}G_n(v_n) \ge 1-\varepsilon_n \to 1$ we can assume without loss of generality that

$$v_n^{-1}G_n(v_n) \to C \in [1,\infty], \tag{6.23}$$

otherwise we use standard subsequence arguments.

First, consider $C < \infty$. Since $\phi_s(1) = \phi_s'(1) = 0$ and $\phi_s''(x) = x^{s-2}$, $x > 0$, we obtain from Taylor's Theorem that

$$v_n \phi_s\left(\frac{\mathbb{F}_n(v_n)}{v_n}\right) = \frac{(\mathbb{F}_n(v_n) - v_n)^2}{2v_n}\left(\frac{\mathbb{F}_n^*}{v_n}\right)^{s-2}, \tag{6.24}$$

where $\mathbb{F}_n^*$ is a random variable fulfilling $\min\{v_n, \mathbb{F}_n(v_n)\} \le \mathbb{F}_n^* \le \max\{v_n, \mathbb{F}_n(v_n)\}$. We can deduce from (6.21) and the monotonicity of $x \mapsto x^{s-2}$ $(x > 0)$ that for all $0 < \delta < \min\{1, C^{s-2}\}$

$$Q_n^n\left(\left(\frac{\mathbb{F}_n^*}{v_n}\right)^{s-2} > \delta\right) \ge Q_n^n\left(\min\left\{1, \left(\frac{\mathbb{F}_n(v_n)}{v_n}\right)^{s-2}\right\} > \delta\right) \to 1. \tag{6.25}$$

Ditzhaus and Janssen [16] showed in the proof of their Theorem 4.1 that under (6.19)

$$Q_n^n\left(\sqrt{n}\frac{|\mathbb{F}_n(v_n) - v_n|}{\sqrt{v_n(1-v_n)l_n}} > \gamma\right) \to 1 \qquad \text{for all } \gamma > 0. \tag{6.26}$$

The main idea of proving (6.26) is a simple application of Chebyschev's inequality. Combining (6.24) to (6.26) yields (6.22).

Now, consider $C = \infty$. By (6.20), (6.21) and (6.23) we have for all $\gamma > 0$:

$$Q_n^n\big(l_n^{-1} n \mathbb{F}_n(v_n) > \gamma\big) \to 1 \quad \text{and} \quad Q_n^n\big(v_n^{-1}\mathbb{F}_n(v_n) > \gamma\big) \to 1. \tag{6.27}$$

Due to the latter convergence statement we only need to analyse $\phi_s(x)$ for sufficiently large x more closely to prove (6.22). It is easy to verify that there exist some $c_{1,s}, c_{2,s} > 0$ and $c_{3,s} \in \mathbb{R}$ such that

$$\phi_s(x) \ge c_{2,s}x + c_{3,s} \qquad \text{for all } x \ge c_{1,s}.$$

For this purpose, consider the cases $s < 0, s = 0, s \in (0,1), s > 1$ separately. For example, if $s \in (0,1)$ then $\phi_s(x) \ge x(s - x^{s-1})/(s(1-s)) \ge x/(2(1-s))$ for all $x \ge (2^{-1}s)^{1/(s-1)}$. Finally,

$$\begin{aligned}
&Q_n^n\left(\frac{nv_n}{l_n}\phi_s\left(\frac{\mathbb{F}_n(v_n)}{v_n}\right) > 2\right) \\
&\ge Q_n^n\left(\frac{nv_n}{l_n}\left(\frac{\mathbb{F}_n(v_n)}{v_n} + \frac{c_{3,s}}{c_{2,s}}\right) > \frac{2}{c_{2,s}}, \frac{\mathbb{F}_n(v_n)}{v_n} > c_{1,s}, \frac{c_{3,s}}{c_{2,s}} \ge -\frac{1}{2}\frac{\mathbb{F}_n(v_n)}{v_n}\right) \\
&\ge Q_n^n\left(\frac{n\mathbb{F}_n(v_n)}{l_n} > \frac{4}{c_{2,s}}, \frac{\mathbb{F}_n(v_n)}{v_n} > \max\left\{c_{1,s}, -\frac{2c_{3,s}}{c_{2,s}}\right\}\right),
\end{aligned}$$

where the latter probability tend to 1 by (6.27). □

Proof of Theorem 3.2. Again, we can assume that $P_0 = \lambda\!\!\lambda_{|(0,1)}$. Let $d_n = n^{-1}(\log n)^5$. By Theorem 4.2 of Ditzhaus and Janssen [16] we have $b_n \mathbb{Z}_n(0,1) - c_n \to Y$ in distribution under Q_n^n, where b_n and c_n are defined by (6.16). Combining this and the mutual contiguity yields that all statements in (6.14) and (6.15) hold under Q_n^n. From these convergence statements and (6.18), we obtain the theorem's statement in the case of $s = 2$. Finally, the theorem's statement for general $s \in \mathbb{R}$ follows from the Taylor expansion (6.6), (6.8), (6.9) and, again, the mutual contiguity. □

Proof of Theorem 4.2. The proof is divided into two parts. First, we discuss the case $\beta < \beta^*$ by applying our Theorem 3.1 and, second, the case $\beta > \beta^*$ by make use of Proposition 6.2. Set $l_n = \log\log(n)$.

The case $\beta < \beta^*$. By Theorem 3.1 and Remark 3.3 it is sufficient to show that

$$l_n^{-1}\widetilde{H}_n(v_n) = n^{1/2-\beta}v_n^{-1/2}l_n^{-1}\left(\int_0^{v_n}\frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}\big(F_0^{-1}(x)\big) + \frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}\big(F_0^{-1}(1-x)\big)\,\mathrm{d}x\right)$$

converges to ∞ for some $\log(n)n^{-1} \le v_n \le (\log(n))^{-1}$. Using the parametrization $v_n = n^{-\tau_n}$ with $\tau_n \in [\widetilde{\tau}_n, 1-\widetilde{\tau}_n]$ and $\widetilde{\tau}_n = l_n(\log(n))^{-1}$ we obtain by substituting $n^{-t} = x$ that

$$l_n^{-1}\widetilde{H}_n(v_n) \ge n^{1/2-\beta+\tau_n/2}l_n^{-1}\log(n)\int_{\tau_n}^{\infty}\exp\big(h_n(t) - t\log(n)\big)\,\mathrm{d}t. \tag{6.28}$$

Fix $\delta \in (0,1)$ with $\delta^{-1} \in \mathbb{N}$ and $2\delta \le \beta^* - \beta$. By (4.3) there exists some $\kappa \in (0, 1/2)$ such that for every sufficiently large $n \in \mathbb{N}$

$$\lambda\!\!\lambda\big(t \in (1,\infty) : \big(\beta^* - \delta - 1 + t\big)\log(n) \le h_n(t)\big) \ge \kappa \quad \text{or} \tag{6.29}$$

$$\lambda\!\!\lambda\left(t \in (0,1) : \left(\beta^* - \delta + \frac{t}{2} - \frac{1}{2}\right)\log(n) \le h_n(t)\right) \ge \kappa. \tag{6.30}$$

If (6.29) holds, then we set $\tau_n = 1 - \widetilde{\tau}_n$ and get from (6.28)

$$l_n^{-1}\widetilde{H}_n(v_n) \ge \kappa\sqrt{\log(n)}\,l_n^{-1}n^{\delta}.$$

Otherwise, if (6.30) holds then

$$\lambda\!\!\lambda\left(t \in \big(\delta(j_n - 1) + \widetilde{\tau}_n, j_n\delta\big) : \left(\beta^* - \delta + \frac{t}{2} - \frac{1}{2}\right)\log(n) \le h_n(t)\right) \ge \frac{\delta\kappa}{2}$$

for some $j_n \in \{1, \dots, \delta^{-1}\}$. In this case, we set $\tau_n = \delta(j_n - 1) + \widetilde{\tau}_n$ and obtain

$$l_n^{-1}\widetilde{H}_n(v_n) \ge \frac{\kappa\delta}{2l_n}\log(n)n^{\beta^*-\beta-\delta+\tau_n/2-j_n\delta/2} \ge \frac{\kappa\delta}{2l_n}\big(\log(n)\big)^{3/2}n^{\delta/2}.$$

To sum up, we can conclude $l_n^{-1}\widetilde{H}_n(v_n) \to \infty$.

The case $\beta > \beta^*$. Fix $x > 0$. Note for a uniformly distributed U on $(0,1)$ that $F_0^{-1}(U)$ and $F_0^{-1}(1-U)$ are P_0-distributed. By Proposition 6.2 and the substitution $u = n^{-t}$ it remains to show that

$$n^{1-\beta}\mu_n\left(n^{-\beta}\frac{\mathrm{d}\mu_n}{\mathrm{d}P_0} > x\right)$$

$$
\begin{aligned}
&= n^{1-\beta} \int \mathbf{1}\Big\{u \in \Big(0, \frac{1}{2}\Big) : n^{-\beta} \frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}\big(F_0^{-1}(u)\big) > x\Big\} \frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}\big(F_0^{-1}(u)\big)\,\mathrm{d}u \\
&\quad + n^{1-\beta} \int \mathbf{1}\Big\{u \in \Big(0, \frac{1}{2}\Big) : n^{-\beta} \frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}\big(F_0^{-1}(1-u)\big) > x\Big\} \frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}\big(F_0^{-1}(1-u)\big)\,\mathrm{d}u \\
&\le 2n^{1-\beta} \log(n) \int \mathbf{1}\Big\{t \ge \frac{\log(2)}{\log(n)} : n^{-\beta} \exp\big(h_n(t)\big) > x\Big\} \\
&\quad \times \exp\big(h_n(t) - t\log(n)\big)\,\mathrm{d}t
\end{aligned}
\tag{6.31}
$$

converges to 0 and

$$
\begin{aligned}
&n^{1-2\beta} \int \mathbf{1}\Big\{n^{-\beta} \frac{\mathrm{d}\mu_n}{\mathrm{d}P_0} \le x\Big\} \Big(\frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}\Big)^2 \mathrm{d}P_0 \\
&\quad \le 2n^{1-2\beta} \log(n) \int \mathbf{1}\Big\{t \ge \frac{\log(2)}{\log(n)} : n^{-\beta} \exp\big(h_n(t)\big) \le x\Big\} \exp\big(2h_n(t) - t\log(n)\big)\,\mathrm{d}t
\end{aligned}
$$

does so as well. To verify both convergences, set

$$
\begin{aligned}
I_{n,1} &= \log(n) n^{1-2\beta} \int_{(\log(2)/\log(n))}^{1} \exp\Big(\log(n)\Big(\frac{2h_n(t)}{\log(n)} - t\Big)\Big)\,\mathrm{d}t, \\
I_{n,2} &= \log(n) n^{1-\beta} \int_{1}^{M} \exp\Big(\log(n)\Big(\frac{h_n(t)}{\log(n)} - t\Big)\Big)\,\mathrm{d}t \quad \text{and} \\
I_{n,3} &= \log(n) n^{1-\beta} \int_{M}^{\infty} \exp\Big(\log(n)\Big(\frac{h_n(t)}{\log(n)} - t\Big)\Big)\,\mathrm{d}t.
\end{aligned}
$$

Observe that

$$
\begin{aligned}
n^{1-\beta} \mu_n\Big(n^{-\beta} \frac{\mathrm{d}\mu_n}{\mathrm{d}P_0} > x\Big) &\le 2x^{-1} I_{n,1} + 2(I_{n,2} + I_{n,3}) \quad \text{and} \\
n^{1-2\beta} \int \mathbf{1}\Big\{n^{-\beta} \frac{\mathrm{d}\mu_n}{\mathrm{d}P_0} \le x\Big\} \Big(\frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}\Big)^2 \mathrm{d}P_0 &\le 2I_{n,1} + 2x(I_{n,2} + I_{n,3}).
\end{aligned}
$$

From (4.4) with $\delta = (\beta - \beta^*)/2 > 0$ we deduce for all $n \ge N_{1,\delta}$ that $I_{n,1} \le \log(n) n^{-2\delta} \to 0$ and $I_{n,2} \le \log(n) n^{-\delta} M \to 0$. Consequently, it remains to prove $I_{n,3} \to 0$.

First, assume that (4.6) holds for $\delta = (\beta - \beta^*)/2 > 0$. Then for all $n \ge N_{2,\delta}$

$$
I_{n,3} \le n^{-\delta} \log(n) \int_1^\infty \exp(-\lambda_n t)\,\mathrm{d}t = \log(n) n^{-\delta} \lambda_n^{-1} \to 0.
$$

Second, suppose that (4.5) is fulfilled. Similar to the calculation in (6.31), we obtain $\log(n) \int \exp(h_{n,j}(t)) n^{-t}\,\mathrm{d}t = 1$ and, thus, $\int \exp(h_n(t)) n^{-t}\,\mathrm{d}t \le 2$. For all $\kappa > 0$ there exists some $N_{3,\kappa} \in \mathbb{N}$ such that for all $n \geq N_{3,\kappa}$

$$\sup_{t \geq M} \left| \gamma(t) - \frac{h_n(t)}{\log(n)} \right| \leq \kappa \quad \text{and, hence,} \quad \int_M^\infty n^{\gamma(t)-t} \, \mathrm{d}t \leq 2n^\kappa. \tag{6.32}$$

From the latter and a simple proof by contradiction, we conclude that

$$\lambda\!\!\lambda\big(t \geq M : \gamma(t) - t > 0\big) = 0.$$

We want to point out that Cai and Wu [9] already showed the two previous statements for $M = 1$ under the assumption (4.1). Let $\lfloor x \rfloor$ be the integer part of $x \in \mathbb{R}$ and $\tau_n = l_n / \log(n)$. To show $I_{n,3} \to 0$, we divide $h_n(t)$ as follows: $h_n(t) = (1 - \tau_n)h_n(t) + \tau_n h_n(t)$. To get an upper bound, we use (4.4) with $\delta = (\beta - \beta^*)/2$ for the first summand and the first statement in (6.32) with $\kappa = 1$ for the second summand. Consequently, there is some $N_4 \geq N_{1,\delta} + \exp(N_{3,1})$ such that for all $n \geq N_4$

$$\begin{aligned}
I_{n,3} &= \log(n) n^{1-\beta} \int_M^\infty n^{(1-\tau_n)(h_n(t)/\log(n) - t)} n^{\tau_n h_n(t)/\log(n) - \tau_n t} \, \mathrm{d}t \\
&\leq \log(n) n^{1-\beta+(1-\tau_n)(\beta^*+\delta-1)} \int_M^\infty n^{\tau_n(\gamma(t)+1-t)} \, \mathrm{d}t \\
&\leq \log(n)^{3-\beta^*-\delta} n^{-\delta} \int_M^\infty \lfloor \log(n) \rfloor^{\gamma(t)-t} \, \mathrm{d}t \leq n^{-\delta/2} 2 \lfloor \log(n) \rfloor \to 0.
\end{aligned}$$

□

Proof of Theorem 4.3. Without loss of generality, we can assume that $P_0 = \lambda\!\!\lambda_{|(0,1)}$ and $T(0) = 0$. By assumption T restricted on $[0, \eta]$ is invertible. Denote by T^{-1} its inverse. We deduce from Theorem 1.5.12 of Bingham et al. [7] that for all $x \in [0, -T(\eta)]$ we have

$$\lambda\!\!\lambda_{|(0,1)}^{-T}(0, x) = T^{-1}(x) = x^p L_1\left(\frac{1}{x}\right)$$

for a slowly varying function L_1 at infinity. Hence, from Theorems XIII.5.2 and XIII.5.3 of Feller [21] we obtain $C(\vartheta_n) = n^{rp} L_2(n^r)$ for a slowly varying function L_2 at infinity. Moreover, it is well known, see Proposition 1.3.6 of Bingham et al. [7], that $\log(L_2(x)) = o(\log(x))$ and $L(x)x^\kappa \to \infty$ as $x \to \infty$ for all $\kappa > 0$. Let $h_{n,1}$, $h_{n,2}$ and h_n be defined as in Proposition 4.1. Fix $\delta \in (0, rp)$ and set $\lambda_n = \log\log(n)$. By the monotonicity of T

$$h_n(t) \leq \log\big(C(\vartheta_n)\big) + \vartheta_n T\big(n^{-rp+\delta/2}\big) = \big(rp + o(1)\big)\log(n) - n^{\delta/(2p)} L\big(n^{rp}\big)\big(1 + o(1)\big)$$

for all $t \leq rp - \delta/2$. Consequently, there is some constant $K > 0$ such that

$$\sup_{t \in (\log(2)/\log(n), rp - \delta/2)} \left\{ \frac{h_n(t)}{\log(n)} - \left(1 - \frac{\lambda_n}{\log(n)}\right)t + \frac{\min\{1, t\}}{2} \right\} \leq K - n^{\delta/(4p)} \to -\infty.$$

Since $T \le 0$ we have for sufficiently large n

$$\sup_{t \ge rp - \frac{\delta}{2}} \left\{ \frac{h_n(t)}{\log(n)} - \left(1 - \frac{\lambda_n}{\log(n)}\right) t + \frac{\min\{1, t\}}{2} \right\}$$
$$\le rp + o(1) - \left(rp - \frac{\delta}{2}\right) + \frac{\min\{rp - \delta/2, 1\}}{2} < \beta^{\#}(r, p) - \frac{1}{2} + \delta.$$

To sum up, (4.4) and (4.6) hold for $\beta^* = \beta^{\#}(r, p)$. Similarly to the previous calculations,

$$\sup_{t \in (rp + \delta/4, rp + \delta/2)} \left\{ \frac{h_n(t)}{\log(n)} - t + \frac{\min\{1, t\}}{2} \right\} > \beta^{\#}(r, p) - \frac{1}{2} - \delta$$

for sufficiently large n and, consequently, (4.3) is fulfilled for $\beta^* = \beta^{\#}(r, p)$. □

Proof of Theorem 4.5. Due to the analogy to the proof of Theorem 4.2, we skip some details here. Set $l_n = \log\log(n)$. By Φ, Φ^{-1}, ϕ we denote the distribution function, the left continuous quantile function and the density of $N(0, 1)$, respectively.

The case $\beta < \beta^*$. By $\Phi^{-1}(1 - x) = -\Phi^{-1}(x)$, Theorem 3.1 and Remark 3.3 it remains to show for some $v_n = n^{-\tau_n}$ with $\tau_n \in [\widetilde{\tau}_n, 1 - \widetilde{\tau}_n]$ and $\widetilde{\tau}_n = l_n / \log(n)$ that

$$l_n^{-1} \widetilde{H}_n(v_n) = n^{1/2 - \beta + \tau_n/2} \frac{1}{l_n \sqrt{2\pi}} \int \frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}(x) \exp\left(-\frac{x^2}{2}\right) \mathbf{1}\{|x| \ge \Phi^{-1}(1 - v_n)\} \,\mathrm{d}x$$

converges to ∞. A simple consequence of integration by parts is $\phi(x)(x/(1+x^2)) \le 1 - \Phi(x) \le \phi(x)/x$ for all $x > 0$. By this it is easy to obtain $\Phi^{-1}(1 - u) \le \sqrt{-2\log(u)}$ for all sufficiently small $u > 0$. Hence, $\Phi^{-1}(1 - v_n) \le \sqrt{2\tau_n \log(n)}$ for all sufficiently large n. Combining this and the substitution $x = y\sqrt{2\log(n)}$ yields

$$l_n^{-1} \widetilde{H}_n(v_n) \ge \int n^{1/2 - \beta + \tau_n/2 + \widetilde{h}_n(y)/\log(n) - y^2} \mathbf{1}\{|y| \ge \sqrt{\tau_n}\} \,\mathrm{d}y$$

for sufficiently large n. Fix $\delta \in (0, 1)$ with $\delta^{-1} \in \mathbb{N}$ and $2\delta \le \beta^* - \beta$. By (4.10) there exists some $\kappa \in (0, 1/2)$ such that for every sufficiently large n

$$\lambda\!\!\!\lambda\left(|y| > 1 : \beta^* - \delta - 1 + y^2 \le \frac{\widetilde{h}_n(y)}{\log(n)}\right) \ge \kappa \quad \text{or} \tag{6.33}$$
$$\lambda\!\!\!\lambda\left(|y| \in \left(\sqrt{\delta(j_n - 1) + \widetilde{\tau}_n}, \sqrt{j_n \delta}\right) : \beta^* - \delta + \frac{y^2 - 1}{2} \le \frac{\widetilde{h}_n(y)}{\log(n)}\right) \ge \kappa$$

for some appropriate $j_n \in \{1, \ldots, \delta^{-1}\}$. If (6.33) holds then set $\tau_n = 1 - \widetilde{\tau}_n$ and otherwise set $\tau_n = \delta(j_n - 1) + \widetilde{\tau}_n$. Consequently, we obtain analogously to the proof of Theorem 4.2 that $l_n^{-1} \widetilde{H}_n(v_n) \to \infty$.

The case $\beta > \beta^*$. Set

$$I_{n,1} = n^{1-2\beta}\sqrt{\frac{\log(n)}{\pi}}\int_{-1}^{1} n^{2\widetilde{h}_n(x)/\log(n)-x^2}\,\mathrm{d}x,$$

$$I_{n,2} = n^{1-\beta}\sqrt{\frac{\log(n)}{\pi}}\int n^{\widetilde{h}_n(x)/\log(n)-x^2}\mathbf{1}\big\{|x|\in(1,M)\big\}\,\mathrm{d}x \quad \text{and}$$

$$I_{n,3} = n^{1-\beta}\sqrt{\frac{\log(n)}{\pi}}\int n^{\widetilde{h}_n(x)/\log(n)-x^2}\mathbf{1}\big\{|x|\geq M\big\}\,\mathrm{d}x.$$

Fix $y > 0$. It remains to verify that

$$n^{1-\beta}\mu_n\left(n^{-\beta}\frac{\mathrm{d}\mu_n}{\mathrm{d}P_0} > y\right) \leq \frac{1}{y}I_{n,1} + I_{n,2} + I_{n,3} \quad \text{and}$$

$$n^{1-2\beta}\int \mathbf{1}\left\{n^{-\beta}\frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}\leq y\right\}\left(\frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}\right)^2\mathrm{d}P_0 \leq I_{n,1} + y(I_{n,2}+I_{n,3})$$

converges to 0. By (4.11) setting $\delta = (\beta-\beta^*)/2 > 0$ we have $I_{n,1} \leq 2n^{-2\delta}\sqrt{\log(n)/\pi} \to 0$ and $I_{n,2} \leq 2Mn^{-\delta}\sqrt{\log(n)/\pi} \to 0$. It remains to discuss $I_{n,3}$.

First, assume that (4.13) holds for $\delta = (\beta-\beta^*)/2 > 0$. Then

$$I_{n,3} \leq n^{-\delta}\sqrt{\frac{\log(n)}{\pi}}\int_{-\infty}^{\infty}\exp\big(-\lambda_n x^2\big)\,\mathrm{d}x = n^{-\delta}\sqrt{\frac{\log(n)}{\lambda_n}} \to 0.$$

Second, suppose that (4.12) is fulfilled. Analogously to the proof of Theorem 4.2, there is some $\widetilde{N}_\kappa \in \mathbb{N}$ such that $\int \mathbf{1}\{|x|\geq M\}n^{\alpha(x)-x^2}\,\mathrm{d}x \leq 2n^{\kappa}$ for every $\kappa > 0$ and all $n \geq \widetilde{N}_\kappa$. Moreover, $\lambda\!\!\lambda(|x|\geq M : \alpha(x)-x^2 > 0) = 0$. Let $\tau_n = l_n/\log(n)$. Finally, from (4.11) with $\delta = (\beta-\beta^*)/2$ and (4.12) we get for sufficiently large n

$$\begin{aligned} I_{n,3} &\leq \sqrt{\frac{\log(n)}{\pi}}n^{1-\beta+(1-\tau_n)(\beta^*+\delta-1)}\int \mathbf{1}\big\{|x|\geq M\big\}n^{\tau_n(\alpha(x)+1-x^2)}\,\mathrm{d}x \\ &\leq n^{-\delta/2}\int \mathbf{1}\big\{|x|\geq M\big\}\big\lfloor\log(n)\big\rfloor^{\alpha(x)-x^2}\,\mathrm{d}x \leq n^{-\delta/2}2\big\lfloor\log(n)\big\rfloor \to 0. \end{aligned}$$

□

Proof of Theorem 4.6. We split the proof into two steps:

1. *Likelihood ratio test sequences*: Let $r \geq \rho^*(\beta)$. Here, we give the proof for (c) and for the part of (b) regarding the lower bound of the error probabilities sum attained by a likelihood ratio test sequence. By the explanations in the preliminary Section 6.1 and, in particular, Proposition 6.2 therein it is sufficient to show for all $x > 0$ that

$$I_{n,1,x} \to 0 \quad \text{and} \quad I_{n,2,x} \to \sigma^2\mathbf{1}\big\{r = \rho^*(\beta)\big\} \qquad \text{with } \sigma^2 = \mathrm{Var}_{P_{(0)}}(T),$$

where in the situation here the quantities $I_{n,1,x}$ and $I_{n,2,x}$ can be simplified to

$$I_{n,1,x} = n^{1-\beta} P_{(\vartheta_n)}\left(\varepsilon_n \frac{\mathrm{d}P_{(\vartheta_n)}}{\mathrm{d}P_{(0)}} > x\right),$$

$$I_{n,2,x} = n^{1-2\beta}\left(\frac{C(\vartheta_n)^2}{C(2\vartheta_n)} P_{(2\vartheta_n)}\left(\varepsilon_n \frac{\mathrm{d}P_{(\vartheta_n)}}{\mathrm{d}P_{(0)}} \le x\right) - 1\right).$$

For this purpose, we introduce the Laplace transform ω defined by

$$\omega(\vartheta) = C(\vartheta)^{-1} = \int \exp(\vartheta T)\,\mathrm{d}P_{(0)}, \qquad \vartheta \in \Theta. \tag{6.34}$$

By Corollary 7.1 of Barndorff–Nielsen [5] the Laplace transform ω is analytic in a neighborhood around 0 and the derivatives can be determined by differentiation under the integral sign. Hence, there is $M \in (1,\infty)$ such that for all $x > 0$ and $n \ge N_x$

$$C(2\vartheta_n) \le 2, \qquad \frac{\varepsilon_n C(\vartheta_n)}{x} \le e^{-1} \quad \text{and} \quad \omega^{(4)}(2\vartheta_n) = \int T^4 \exp(2\vartheta_n T)\,\mathrm{d}P_{(0)} < M,$$

where $f^{(k)}$ denotes the derivative of order $k \in \mathbb{N}$ of the function f. Thus, we obtain for all $x > 0$ and $n \ge N_x$

$$\begin{aligned}
P_{(2\vartheta_n)}\left(\varepsilon_n \frac{\mathrm{d}P_{(\vartheta_n)}}{\mathrm{d}P_{(0)}} > x\right) &\le 2\int \mathbf{1}\left\{\vartheta_n T > \log\left(\frac{x}{\varepsilon_n C(\vartheta_n)}\right)\right\} \exp(2\vartheta_n T)\,\mathrm{d}P_{(0)} \\
&\le 2\int (\vartheta_n T)^4 \mathbf{1}\{\vartheta_n T > 1\} \exp(2\vartheta_n T)\,\mathrm{d}P_{(0)} \\
&\le 2\vartheta_n^4 \omega^{(4)}(2\vartheta_n) = o\big(n^{-2r}\big).
\end{aligned}$$

By a Taylor expansion around 0, we can conclude that as $\vartheta \to 0$

$$\begin{aligned}
&\omega(2\vartheta) = 1 + 2\vartheta E_{P_{(0)}}(T) + 2\vartheta^2 E_{P_{(0)}}\big(T^2\big) + o\big(\vartheta^2\big) \quad \text{and, thus,} \\
&\omega(\vartheta)^2 = \omega(2\vartheta) - \vartheta^2\sigma^2 + o\big(\vartheta^2\big).
\end{aligned}$$

Consequently, for all $x > 0$

$$I_{n,2,x} = n^{1-2\beta}\left(\frac{\omega(2\vartheta_n)}{\omega(\vartheta_n)^2}\big(1 + o\big(\vartheta_n^2\big)\big) - 1\right) = n^{1-2\beta-2r}\sigma^2\big(1 + o(1)\big),$$

which proves $I_{n,2,x} \to \sigma^2 \mathbf{1}\{r = \rho^*(\beta)\}$. Furthermore, for all $x > 0$

$$I_{n,1,x} \le \frac{1}{x} n^{1-2\beta} \frac{C(\vartheta_n)^2}{C(2\vartheta_n)} P_{(2\vartheta_n)}\left(\varepsilon_n \frac{\mathrm{d}P_{(\vartheta_n)}}{\mathrm{d}P_{(0)}} > x\right) = o\big(n^{1-2\beta-2r}\big).$$

2. *Test sequences based on* $S_n(s)$: Here, we verify (a) and the statement of (b) about $S_n(s)$. To apply Theorems 3.1 and 3.2, we analyse the asymptotic behavior of $H_n(v)$ given by (3.1) more

closely. Again, we assume $P_{(0)} = \lambda\!\!\!\lambda_{|(0,1)}$. As already stated, the Laplace transform ω introduced in (6.34) is analytic in $(-\delta, \delta)$ for sufficiently small $\delta \in (0, 1)$ and so is C. From now on, let n be sufficiently large such that $\vartheta_n \leq \delta/2$. Observe that for all $v \in (0, 1/2)$

$$\mu_n(0, v) - v = \int_{B_{1,v}} \chi_x(\vartheta_n)\,\mathrm{d}x \quad \text{and} \quad \mu_n(1-v, 1) - v = \int_{B_{2,v}} \chi_x(\vartheta_n)\,\mathrm{d}x,$$

$$\text{where } \chi_x(\vartheta) = C(\vartheta)\exp\bigl(\vartheta T(x)\bigr) - 1,\ B_{1,v} = (0, v) \text{ and } B_{2,v} = (1-v, 1).$$

For every (fixed) $x \in (0, 1)$

$$\begin{aligned} \chi_x^{(1)}(\vartheta) &= C^{(1)}(\vartheta)\exp\bigl(\vartheta T(x)\bigr) + C(\vartheta)T(x)\exp\bigl(\vartheta T(x)\bigr) \quad \text{and} \\ \chi_x^{(2)}(\vartheta) &= \bigl[C^{(2)}(\vartheta) + 2C^{(1)}(\vartheta)T(x) + C(\vartheta)T^2(x)\bigr]\exp\bigl(\vartheta T(x)\bigr). \end{aligned}$$

From $C^{(1)}(0) = -\omega^{(1)}(0) = -\int_0^1 T\,\mathrm{d}\lambda\!\!\!\lambda$ and a Taylor expansion around 0 we deduce that

$$\chi_x(\vartheta_n) = \vartheta_n\left(-\int_0^1 T\,\mathrm{d}\lambda\!\!\!\lambda + T(x)\right) + \frac{\vartheta_n^2}{2}\chi_x^{(2)}\bigl(r_n(x)\bigr) \qquad \text{with } r_n(x) \in [0, \vartheta_n].$$

Since C and ω are analytic there exists $M > 1$ independent of ϑ such that

$$\bigl|C(\vartheta)\bigr| + \bigl|C^{(1)}(\vartheta)\bigr| + \bigl|C^{(2)}(\vartheta)\bigr| + \sum_{k=0}^{2}\int_0^1 |T|^k \exp(\vartheta T)\,\mathrm{d}\lambda\!\!\!\lambda \leq M.$$

By Hölder's inequality it holds for all $f : (0, 1) \to \mathbb{R}$ with $\int_0^1 f^4\,\mathrm{d}\lambda\!\!\!\lambda \leq M$ that

$$\int_{B_{j,v}} |f|\,\mathrm{d}\lambda\!\!\!\lambda \leq v^{3/4}M^{1/4} \leq v^{3/4}M \qquad \text{for } j = 1, 2.$$

Hence, for all $j \in \{1, 2\}$, $v \in (0, 1/2)$

$$\begin{aligned} \int_{B_{j,v}} \bigl|\chi_x^{(2)}\bigl(r_n(x)\bigr)\bigr|\,\mathrm{d}x &\leq M\int_{B_{j,v}} \bigl(1 + 2|T| + T^2\bigr)\bigl(\exp(\vartheta_n T) + \exp(-\vartheta_n T)\bigr)\,\mathrm{d}\lambda\!\!\!\lambda \\ &\leq 8v^{3/4}M^2. \end{aligned}$$

Consequently,

$$\begin{aligned} H_n(v) &\geq n^{1/2-\beta}v^{-1/2}\left|\int_{B_{j,v}} \chi_x(\vartheta_n)\,\mathrm{d}x\right| \\ &= n^{1/2-\beta-r}v^{-1/2}\left|\int_{B_{j,v}}\left(T - \int_0^1 T\,\mathrm{d}\lambda\!\!\!\lambda\right)\mathrm{d}\lambda\!\!\!\lambda + o(1)\right|\bigl(1 + o(1)\bigr). \end{aligned} \tag{6.35}$$

If $\int_0^v (T - \int_0^1 T\,\mathrm{d}\lambda\!\!\!\lambda)\,\mathrm{d}\lambda\!\!\!\lambda = 0 = \int_{1-v}^1 (T - \int_0^1 T\,\mathrm{d}\lambda\!\!\!\lambda)\,\mathrm{d}\lambda\!\!\!\lambda$ would hold for every $v \in (0, 1/2)$, then $T \equiv \int_0^1 T\,\mathrm{d}\lambda\!\!\!\lambda$ would be true $\lambda\!\!\!\lambda_{|(0,1)}$-almost surely. But the latter contradicts the assumption

$\mathrm{Var}_{P_{(0)}}(T) > 0$. Thus, we can deduce from (6.35) that $(\log\log(n))^{-1} H_n(v^*) \to \infty$ for some $v^* \in (0, 1/2)$ if $r < \rho^*(\beta)$. Applying Theorem 3.1 with $v_n = v^*$ gives us (a).

It remains to discuss the case $r = \rho^*(\beta)$. Set $c_{n,1} = 1/n$, $c_{n,2} = c_{n,3} = 1/\sqrt{n}$ and $c_{n,4} = (\log(n))^{-4}$. Clearly, (3.4) holds with $\kappa = 1/2$. Moreover, we can conclude from our previous considerations

$$\sup_{v \in [c_{n,1}, c_{n,4}]} H_n(v) \le 2n^{1/2-\beta-r} \sup_{v \in [c_{n,1}, c_{n,4}]} v^{-1/2}\left(vM + v^{3/4}M + 8\frac{\vartheta_n}{2}v^{3/4}M^2\right)$$
$$\le 2 \sup_{v \in [c_{n,1}, c_{n,4}]} v^{1/4} 10M^2 = 20M^2\big(\log(n)\big)^{-1} \to 0.$$

Hence, (3.3) is fulfilled. Recall that in the first proof step we applied Proposition 6.2 to verify $\{P_0^n, Q_n^n\} \xrightarrow{w} \{N(-\sigma^2/2, \sigma^2), N(\sigma^2/2, \sigma^2)\}$ for some $\sigma^2 > 0$. In particular, by the first Lemma of Le Cam P_0^n and Q_n^n are mutually contiguous. Finally, the desired statement in (b) follows from Theorem 3.2. □

Acknowledgments

The author thanks the *Deutsche Forschungsgemeinschaft* (DFG) for financial support (DFG Grant no. 618886). Moreover, the author thanks the reviewer, whose suggestions improved the paper's clarity.

References

[1] Ali, S.M. and Silvey, S.D. (1966). A general class of coefficients of divergence of one distribution from another. *J. Roy. Statist. Soc. Ser. B* **28** 131–142. MR0196777

[2] Arias-Castro, E., Candès, E.J. and Plan, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.* **39** 2533–2556. MR2906877

[3] Arias-Castro, E. and Wang, M. (2015). The sparse Poisson means model. *Electron. J. Stat.* **9** 2170–2201. MR3406276

[4] Arias-Castro, E. and Wang, M. (2017). Distribution-free tests for sparse heterogeneous mixtures. *TEST* **26** 71–94. MR3613606

[5] Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. Chichester: Wiley. MR0489333

[6] Berk, R.H. and Jones, D.H. (1979). Goodness-of-fit test statistics that dominate the Kolmogorov statistics. *Z. Wahrsch. Verw. Gebiete* **47** 47–59. MR0521531

[7] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation. Encyclopedia of Mathematics and Its Applications* **27**. Cambridge: Cambridge Univ. Press. MR0898871

[8] Cai, T.T., Jeng, X.J. and Jin, J. (2011). Optimal detection of heterogeneous and heteroscedastic mixtures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 629–662. MR2867452

[9] Cai, T.T. and Wu, Y. (2014). Optimal detection of sparse mixtures against a given null distribution. *IEEE Trans. Inform. Theory* **60** 2217–2232. MR3181520

[10] Cayon, L., Jin, J. and Treaster, A. (2004). Higher criticism statistic: Detecting and identifying non-Gaussianity in the WMAP first year data. *Mon. Not. R. Astron. Soc.* **362** 826–832.

[11] Chang, L.-C. (1955). On the ratio of an empirical distribution function to the theoretical distribution function. *Acta Math. Sinica* **5** 347–368. MR0076213
[12] Cressie, N. and Read, T.R.C. (1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. B* **46** 440–464. MR0790631
[13] Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.* **2** 299–318. MR0219345
[14] Dai, H., Charnigo, R., Srivastava, T., Talebizadeh, Z. and Qing, S. (2012). Integrating P-values for genetic and genomic data analysis. *J. Biom. Biostat.* 3–7.
[15] Ditzhaus, M. (2017). The power of tests for signal detection under high-dimensional data. Ph.D. thesis, Heinrich-Heine-Univ. Duesseldorf. Available at https://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=42808.
[16] Ditzhaus, M. and Janssen, A. (2017). Detectability of nonparametric signals: Higher criticism versus likelihood ratio. Available at 1709.07264v2.
[17] Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195
[18] Donoho, D. and Jin, J. (2009). Feature selection by higher criticism thresholding achieves the optimal phase diagram. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **367** 4449–4470. MR2546396
[19] Donoho, D. and Jin, J. (2015). Higher criticism for large-scale inference, especially for rare and weak effects. *Statist. Sci.* **30** 1–25. MR3317751
[20] Eicker, F. (1979). The asymptotic distribution of the suprema of the standardized empirical processes. *Ann. Statist.* **7** 116–138. MR0515688
[21] Feller, W. (1966). *An Introduction to Probability Theory and Its Applications. Vol. II*. New York: Wiley. MR0210154
[22] Goldstein, D.B. (2009). Common genetic variation and human traits. *N. Engl. J. Med.* **360** 1696–1698.
[23] Gontscharuk, V., Landwehr, S. and Finner, H. (2015). The intermediates take it all: Asymptotics of higher criticism statistics and a powerful alternative based on equal local levels. *Biom. J.* **57** 159–180. MR3298224
[24] Gontscharuk, V., Landwehr, S. and Finner, H. (2016). Goodness of fit tests in terms of local levels with special emphasis on higher criticism tests. *Bernoulli* **22** 1331–1363. MR3474818
[25] Hall, P., Pittelkow, Y. and Ghosh, M. (2008). Theoretical measures of relative performance of classifiers for high dimensional data with small sample sizes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 159–173. MR2412636
[26] Ingster, Y.I., Tsybakov, A.B. and Verzelen, N. (2010). Detection boundary in sparse regression. *Electron. J. Stat.* **4** 1476–1526. MR2747131
[27] Ingster, Yu.I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–69. MR1456646
[28] Iyengar, S.K. and Elston, R.C. (2007). The genetic basis of complex traits: Rare variants or "common gene, common disease"? *Methods Mol. Biol.* **376** 71–84.
[29] Jaeschke, D. (1979). The asymptotic distribution of the supremum of the standardized empirical distribution function on subintervals. *Ann. Statist.* **7** 108–115. MR0515687
[30] Jager, L. and Wellner, J. (2004). A new goodness of fit test: The reversed Berk–Jones statistic. Technical report, Dept. Statistics, Univ. Washington, DC.
[31] Jager, L. and Wellner, J.A. (2007). Goodness-of-fit tests via phi-divergences. *Ann. Statist.* **35** 2018–2053. MR2363962
[32] Jin, J. (2009). Impossibility of successful classification when useful features are rare and weak. *Proc. Natl. Acad. Sci. USA* **106** 8859–8864. MR2520682
[33] Jin, J., Starck, J.-L., Donoho, D.L., Aghanim, N. and Forni, O. (2005). Cosmological non-Gaussian signature detection: Comparing performance of different statistical tests. *EURASIP J. Appl. Signal Process.* **15** 2470–2485. MR2210857

[34] Khmaladze, E. and Shinjikashvili, E. (2001). Calculation of noncrossing probabilities for Poisson processes and its corollaries. *Adv. in Appl. Probab.* **33** 702–716. MR1860097
[35] Khmaladze, E.V. (1998). Goodness of fit tests for "chimeric" alternatives. *Stat. Neerl.* **52** 90–111. MR1615550
[36] Kulldorff, M., Heffernan, R., Hartman, J., Assunção, R. and Mostashari, F. (2005). A space–time permutation scan statistic for disease outbreak detection. *PLoS Med.* **2** e59.
[37] Mukherjee, R., Pillai, N.S. and Lin, X. (2015). Hypothesis testing for high-dimensional sparse binary regression. *Ann. Statist.* **43** 352–381. MR3311863
[38] Neill, D. and Lingwall, J. (2007). A nonparametric scan statistic for multivariate disease surveillance. *Advances in Disease Surveillance* **4** 106–116.
[39] Saligrama, V. and Zhao, M. (2012). Local anomaly detection. *JMLR W&CP* **22** 969–983.
[40] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*. New York: Wiley. MR0838963
[41] Strasser, H. (1985). *Mathematical Theory of Statistics*. *De Gruyter Studies in Mathematics* **7**. Berlin: de Gruyter. MR0812467
[42] Tukey, J.W. (1976). *T13 N: The Higher Criticism*. *Coures Notes*. *Stat* **411**. Princeton: Princeton University Press.
[43] Tukey, J.W. (1989). Higher Criticism for individual significances in several tables or parts of tables. Internal working paper, Princeton Univ.
[44] Tukey, J.W. (1994). *The Collected Works of John W. Tukey. Vol. VIII*. London: Chapman and Hall. MR1263027
[45] Wellner, J.A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrsch. Verw. Gebiete* **45** 73–88. MR0651392

Received March 2018 and revised August 2018