Bernoulli **22**(3), 2016, 1331–1363
DOI: 10.3150/14-BEJ694

Goodness of fit tests in terms of local levels with special emphasis on higher criticism tests

VERONIKA GONTSCHARUK[1,2,*], SANDRA LANDWEHR[1,2,**] and HELMUT FINNER[2,†]

[1]*Department of Statistics in Medicine, Faculty of Medicine, Heinrich-Heine-University, Düsseldorf, Germany.*
E-mail: [*]*veronika.gontscharuk@ddz.uni-duesseldorf.de;* [**]*sandra.landwehr@ddz.uni-duesseldorf.de*
[2]*Institute for Biometrics and Epidemiology, German Diabetes Center at the Heinrich-Heine-University, Düsseldorf, Germany. E-mail:* [†]*finner@ddz.uni-duesseldorf.de*

Instead of defining goodness of fit (GOF) tests in terms of their test statistics, we present an alternative method by introducing the concept of local levels, which indicate high or low local sensitivity of a test. Local levels can act as a starting point for the construction of new GOF tests. We study the behavior of local levels when applied to some well-known GOF tests such as Kolmogorov–Smirnov (KS) tests, higher criticism (HC) tests and tests based on phi-divergences. The main focus is on a rigorous characterization of the asymptotic behavior of local levels of the original HC tests which leads to several further asymptotic results for local levels of other GOF tests including GOF tests with equal local levels. While local levels of KS tests, which are related to the central range, are asymptotically strictly larger than zero, all local levels of HC tests converge to zero as the sample size increases. Consequently, there exists no asymptotic level α GOF test such that all local levels are asymptotically bounded away from zero. Finally, by means of numerical computations we compare classical KS and HC tests to a GOF test with equal local levels.

Keywords: higher criticism statistic; Kolmogorov–Smirnov test; local levels; minimum p-value test; Normal and Poisson approximation; order statistics

1. Introduction

Let $X_1, \dots, X_n$ be real-valued independently identically distributed (i.i.d.) random variables with continuous cumulative distribution function (c.d.f.) F. We are interested in testing the null hypothesis

$$H_0^{\leq}\colon\ F(x) \leq F_0(x) \quad \text{or} \quad H_0^{=}\colon\ F(x) = F_0(x) \qquad \text{for all } x \in \mathbb{R}, \tag{1.1}$$

for a prespecified continuous c.d.f. F_0. Since $F_0(X_i)$, $i = 1, \dots, n$, are i.i.d. uniformly distributed on $[0, 1]$ if $F = F_0$, we restrict our attention to the case where

$$F_0(x) = x \qquad \text{for all } x \in [0, 1].$$

Consequently, we assume that X_i, $i = 1, \dots, n$, take values in $[0, 1]$. We focus on the following class of goodness of fit (GOF) tests in terms of order statistics $X_{1:n}, \dots, X_{n:n}$ related to the underlying sample $X_1, \dots, X_n$. For testing $H_0^{\leq}$ we consider a one-sided test $\varphi : [0,1] \to \{0,1\}$ based on critical values $0 \leq c_{1,n} < \cdots < c_{n,n} < 1$ such that

$$\varphi = 1 \quad \text{iff} \quad X_{i:n} \leq c_{i,n} \qquad \text{for at least one } i = 1, \dots, n. \tag{1.2}$$

A two-sided test $\tilde{\varphi} : [0,1] \to \{0,1\}$ for testing $H_0^{=}$ is defined by

$$\tilde{\varphi} = 1 \quad \text{iff} \quad X_{i:n} \leq c_{i,n} \quad \text{or} \quad X_{i:n} \geq \tilde{c}_{i,n} \qquad \text{for at least one } i = 1, \dots, n, \tag{1.3}$$

where $0 \leq c_{1,n} < \cdots < c_{n,n} < 1$ and $0 < \tilde{c}_{1,n} < \cdots < \tilde{c}_{n,n} \leq 1$ are the corresponding critical values fulfilling $c_{i,n} < \tilde{c}_{i,n}$, $i = 1, \dots, n$. Thereby, $H_0^{\leq}$ is rejected if $\varphi = 1$, while $H_0^{=}$ is rejected if $\tilde{\varphi} = 1$. The global level of the test φ and/or $\tilde{\varphi}$ is given by $\mathbb{E}_0(\varphi) \equiv \mathbb{P}(\varphi = 1 | H_0^{=})$ and/or $\mathbb{E}_0(\tilde{\varphi}) \equiv \mathbb{P}(\tilde{\varphi} = 1 | H_0^{=})$, respectively.

We restrict attention to non-parametric tests only. Among the most famous non-parametric GOF tests we find the Kolmogorov–Smirnov (KS), Anderson–Darling (AD), Cramér–von Mises and Berk–Jones (BJ) tests, where KS and BJ tests and the supremum form of AD tests can be rewritten in the form (1.2) and/or (1.3). In addition, recently proposed GOF tests based on the so-called phi-divergences introduced in [19] are non-parametric tests and can also be represented in the desired form.

In Section 1.1, we briefly discuss the union-intersection principle in relation to GOF tests and local levels. Section 1.2 is concerned with the behavior of local levels of the Kolmogorov–Smirnov test. Some further brief remarks concerning GOF tests in terms of local levels are given in Section 1.3. In Section 1.4, we discuss the idea of GOF tests with equal local levels, related ideas and relations to recent work. In Section 1.5, we switch to higher criticism (HC) tests and some further tests based on phi-divergences and provide some figures which roughly illustrate the behavior of the local levels of these tests. An outline of the remaining part of the paper the focus of which is on the asymptotics of local levels of the original HC statistic is given in Section 1.6.

1.1. The union-intersection principle and local levels

In multiple hypotheses testing local levels appear in a natural way, especially in the case of multiple test procedures based on the union-intersection principle. Such tests accept the global null hypothesis, that is, the intersection of a suitable set of elementary hypotheses H_i, if and only if all elementary hypotheses are accepted. Roughly speaking, a local level α_i for H_i denotes the probability to reject H_i if it is true. Local levels tell us which amount of the overall level α is attributed to each H_i. Often multiple test procedures based on the union-intersection principle have equal local levels. Prominent examples are the classical Bonferroni test, Tukey's multiple range test for pairwise comparisons, Dunnett's test for multiple comparisons with a control or Scheffe's multiple contrast test. A further general example is the minimum p-value test which corresponds to the *minimum level attained* test studied in [3]. The weighted Bonferroni test may serve as an example with different local levels.

GOF tests of the form (1.2) and (1.3) are related to the union-intersection principle in the following way. Let $U_1, \dots, U_n$ be i.i.d. uniformly distributed random variables and $U_{1:n}, \dots, U_{n:n}$ be the corresponding order statistics. For each $i = 1, \dots, n$ consider null hypotheses $H_i^{\leq}$ and $H_i^{=}$

on the distribution of a single order statistic $X_{i:n}$ such that $H_i^{\le}$ is true if $\mathbb{P}(X_{i:n} \le x) \le \mathbb{P}(U_{i:n} \le x)$ for $x \in [0, 1]$, that is, if $X_{i:n}$ is stochastically larger than or equal to $U_{i:n}$ and $H_i^{=}$ is true if $X_{i:n}$ is equal to $U_{i:n}$ in distribution. Define tests for $H_i^{\le}$ by

$$\varphi_i = 1 \quad \text{iff} \quad X_{i:n} < c_{i,n}$$

and tests for $H_i^{=}$ by

$$\tilde{\varphi}_i = 1 \quad \text{iff} \quad X_{i:n} \le c_{i,n} \quad \text{or} \quad X_{i:n} \ge \tilde{c}_{i,n}.$$

Then $H_0^{\le} \subseteq \bigcap_{i=1}^n H_i^{\le}$, $H_0^{=} \subseteq \bigcap_{i=1}^n H_i^{=}$, $\{\varphi = 1\} = \bigcup_{i=1}^n \{\varphi_i = 1\}$ and $\{\tilde{\varphi} = 1\} = \bigcup_{i=1}^n \{\tilde{\varphi}_i = 1\}$ so that the GOF tests φ and $\tilde{\varphi}$ can be seen as union intersection tests. We define *local levels* of a GOF test by

$$\alpha_{i,n} = \mathbb{P}\big(\varphi_i = 1 | H_0^{=}\big) = \mathbb{P}(U_{i:n} \le c_{i,n}) \tag{1.4}$$

in the one-sided case and

$$\alpha_{i,n}^{=} = \mathbb{P}\big(\tilde{\varphi}_i = 1 | H_0^{=}\big) = \mathbb{P}(U_{i:n} \le c_{i,n}) + \mathbb{P}(U_{i:n} \ge \tilde{c}_{i,n}) \tag{1.5}$$

in the two-sided case. Noting that $U_{i:n}$ is beta-distributed with parameters i and $n - i + 1$ and denoting the related c.d.f. by $F_{i,n-i+1}$, we get $\mathbb{P}(U_{i:n} \le x) = F_{i,n-i+1}(x)$.

Local levels can be viewed as an interesting characteristic of a GOF test and may be interpreted as weights for testing the family of null hypotheses $H_i^{\le}$ or $H_i^{=}$, $i = 1, \dots, n$. The larger a local level $\alpha_{i,n}$ or $\alpha_{i,n}^{=}$, the higher the chance to reject the null hypothesis corresponding to the ith smallest order statistic $X_{i:n}$ at least under the null hypothesis. In other words, local levels can be regarded as a tool to signify areas of high/low sensitivity of a test. For example, if deviations from $H_0^{\le}$ and/or $H_0^{=}$ are expected in the tails, one would prefer a GOF test with larger local levels for indices i close to 1 and/or close to n. However, we have to take into account that order statistics are dependent, see, for example, [7] and [31]. This may influence the probability to reject null hypotheses corresponding to a set of ith order statistics with indices i in several ranges.

1.2. Local levels of the Kolmogorov–Smirnov test

One of the most widely-used GOF tests, which can be written in terms of (1.2) and/or (1.3), is the well-known Kolmogorov–Smirnov (KS) test. We consider a one-sided asymptotic level α KS test, which rejects $H_0^{\le}$ if the KS test statistic

$$\mathrm{KS}^+ = \max_{1 \le i \le n} \sqrt{n}(i/n - X_{i:n})$$

is larger than the asymptotic critical value $c_\alpha = \sqrt{-\log(\alpha)/2}$ with $\alpha \in (0, 1)$. It holds $\lim_{n\to\infty} \mathbb{P}(\mathrm{KS}^+ > c_\alpha | H_0^{=}) = \alpha$, cf., for example, [31], page 11. Even for $n \ge 40$, the probability $\mathbb{P}(\mathrm{KS}^+ > c_\alpha | H_0^{=})$ is approximately α. The one-sided KS test can be represented in the form (1.2) with critical values $c_{i,n}^{\mathrm{KS}} = \max(0, i/n - c_\alpha/\sqrt{n})$, $i = 1, \dots, n$. In accordance with (1.4), the corresponding local levels are given by $\alpha_{i,n}^{\mathrm{KS}} = F_{i,n-i+1}(c_{i,n}^{\mathrm{KS}})$, $i = 1, \dots, n$. Note that $\alpha_{i,n}^{\mathrm{KS}} = 0$ for

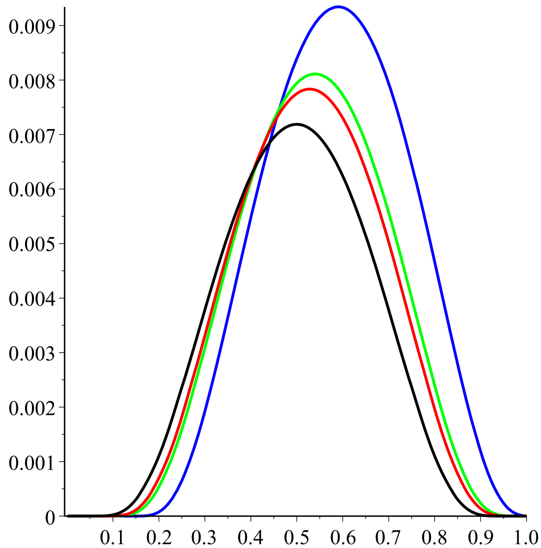

Figure 1. Local levels $\alpha_{i,n}^{\mathrm{KS}}$ as a function of i/n for one-sided KS tests with $\alpha = 0.05$ and $n = 100$, 500, 1000 together with the corresponding asymptotic local levels (from top to bottom in $i/n = 0.8$).

$i \leq c_\alpha \sqrt{n}$. For a finite n, the remaining $\alpha_{i,n}^{\mathrm{KS}}$ can be calculated numerically. Moreover, using the normal approximation, we get for $i \equiv i_n$ satisfying $i_n/n \to \zeta \in (0, 1)$ that

$$\lim_{n\to\infty} \alpha_{i_n,n}^{\mathrm{KS}} = 1 - \Phi\Bigl(\sqrt{-\log(\alpha)/\bigl(2\zeta(1-\zeta)\bigr)}\Bigr),$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. The largest asymptotic local level is attained at $\zeta = 1/2$ and equals $1 - \Phi(\sqrt{-2\log\alpha})$. Figure 1 shows asymptotic and exactly calculated local levels $\alpha_{i,n}^{\mathrm{KS}}$ as a function of i/n for various n-values. For i_n/n in a central range of $[0, 1]$, the limiting local levels are bounded away from zero, whereas for $i_n/n \to \zeta \in \{0, 1\}$ we get $\lim_{n\to\infty} \alpha_{i_n,n}^{\mathrm{KS}} = 0$. This coincides with the well-known fact that KS tests have higher power for alternatives that differ from the null distribution in the central range and low power against alternative distributions which mainly deviate from the null in the tails. Alternatives of this kind, however, are common in many applications, for example, in genome-wide association studies, in which we face a very large number of hypotheses to test with only a small number of them being non-null. For more practical applications see, for example, [6,15] and [16].

Various modifications of the KS test have been proposed in the past. For example, Révész [30] constructed a test based on a statistic which combines the advantages of the classical and normalized KS statistics with regard to their sensitivity ranges. Mason and Schuenemeyer [26] introduced a modified KS test by combining the classical KS with Rényi-type statistics and investigated the finite sample and asymptotic distribution of this modification. Test statistics that are determined by order statistics, in particular tail order statistics, are studied by Lockhart in [24] with respect to asymptotic relative efficiency against a certain class of alternatives. Bahadur efficiencies for a lot of non-parametric GOF tests are extensively studied by Nikitin in [28]. More recently, Jager and Wellner [19] proposed GOF tests based on phi-divergences. Their supremum- and integral-type statistics cover various forms of Anderson–Darling and Berk–Jones statistics.

1.3. GOF tests in terms of local levels

For many (non-parametric) GOF tests, there is a class of alternatives against which this test is the most powerful. Hence, if we have some information about the range, where the alternative distribution mainly deviates from the null distribution, it seems worthwhile to apply such an appropriately tailored GOF test. However, from the viewpoint of test statistics it is difficult to determine whether the corresponding GOF test is sensitive for a predefined range of deviations. Fortunately, the construction of tailored GOF tests is much easier by means of local levels. Thereby, the aim is to construct a GOF test with larger local levels in the crucial area. For example, assuming sparse signals, a GOF test with larger local levels for indices close to 1 and smaller local levels for the remaining indices seems to be a reasonable choice.

In general, for a given suitable set of local levels $\alpha_{i,n}$, $i = 1, \dots, n$, we are able to construct a corresponding GOF test of the form (1.2) and/or (1.3). In the one-sided case the related critical values are given by $c_{i,n} = F^{-1}_{i,n-i+1}(\alpha_{i,n})$, $i = 1, \dots, n$, where $F^{-1}_{i,n-i+1}$ denotes the inverse function of the c.d.f. $F_{i,n-i+1}$. For a two-sided GOF test $\tilde{\varphi}$ we have to decide how to split $\alpha^{=}_{i,n}$ into two non-negative terms $\alpha^{(1)}_{i,n}$ and $\alpha^{(2)}_{i,n}$ such that $\alpha^{(1)}_{i,n} + \alpha^{(2)}_{i,n} = \alpha^{=}_{i,n}$ and $\mathbb{P}(U_{i:n} \le c_{i,n}) = \alpha^{(1)}_{i,n}$, $\mathbb{P}(U_{i:n} \ge \tilde{c}_{i,n}) = \alpha^{(2)}_{i,n}$. One possibility may be $\alpha^{(1)}_{i,n} = \alpha^{(2)}_{i,n} = \alpha^{=}_{i,n}/2$, which leads to $c_{i,n} = F^{-1}_{i,n-i+1}(\alpha^{=}_{i,n}/2)$ and $\tilde{c}_{i,n} = F^{-1}_{i,n-i+1}(1 - \alpha^{=}_{i,n}/2)$. The latter can be calculated at least numerically.

1.4. GOF tests with equal local levels

If we do not have any idea on alternatives, it seems natural to choose a GOF test with equal local levels, that is,

$$\alpha_{1,n} = \cdots = \alpha_{n,n} = \alpha^{\mathrm{loc}}_n \quad \text{and/or} \quad \alpha^{=}_{1,n} = \cdots = \alpha^{=}_{n,n} = \alpha^{\mathrm{loc}}_n$$

for some suitable $\alpha^{\mathrm{loc}}_n \in (0, 1)$. The idea behind this proposal is similar to the idea behind the KS test, where the distance between the empirical c.d.f. $\hat{F}_n(x)$ and the underlying c.d.f. $F_0(x)$, that is, $\hat{F}_n(x) - F_0(x)$ for the one-sided test case and $|\hat{F}_n(x) - F_0(x)|$ for the two-sided case, is compared to the same critical value for each x. That is, the KS test can be seen as a GOF test with equal distances for all feasible x-values. Considering other measures of the distance between the theoretical and the corresponding empirical distributions, one may construct various GOF tests with some quantities being equal. For example, a family of GOF tests introduced in [19] can be seen as tests with equal phi-divergences. A prominent example here is the Berk–Jones test which corresponds to equal Kullback-Leibler divergences. Altogether, the idea of considering equal quantities such as equal distances, critical values, test statistics and also local levels, is a natural approach when constructing GOF tests.

GOF tests with local levels equal to some $\alpha_n^{\text{loc}} \in (0,1)$ are given as follows. The one-sided version of the test $\varphi(\alpha_n^{\text{loc}})$ (say) is defined by (1.2) with $c_{i,n} = F_{i,n-i+1}^{-1}(\alpha_n^{\text{loc}})$, $i = 1, \dots, n$. The two-sided test $\tilde{\varphi}(\alpha_n^{\text{loc}})$ is given by (1.3) with $c_{i,n} = F_{i,n-i+1}^{-1}(\alpha_n^{\text{loc}}/2)$ and $\tilde{c}_{i,n} = 1 - F_{i,n-i+1}^{-1}(\alpha_n^{\text{loc}}/2)$, $i = 1, \dots, n$. In order to get a level α test we have to choose α_n^{loc} such that $\mathbb{E}_0(\varphi(\alpha_n^{\text{loc}})) = \alpha$ and/or $\mathbb{E}_0(\tilde{\varphi}(\alpha_n^{\text{loc}})) = \alpha$. Unfortunately, it seems there does not exist any analytically manageable formula for α_n^{loc} as a function of n and α so that α_n^{loc} has to be calculated numerically. Nevertheless, we are able to provide some bounds for α_n^{loc}. For example, the Bonferroni inequality implies

$$\alpha/n < \alpha_n^{\text{loc}} < \alpha, \qquad n \in \mathbb{N}.$$

Moreover, it can easily be seen that α_n^{loc} lies between the smallest and largest local levels for any (exact) level α GOF test of type (1.2) and (1.3), respectively. Thus, knowledge of local levels corresponding to suitable GOF tests leads at least to upper and lower bounds for α_n^{loc}. For example, by means of the asymptotic KS local levels, we get for the one-sided case

$$0 < \alpha_n^{\text{loc}} \le \Phi\bigl(\sqrt{-2\log(\alpha)}\bigr) + \mathrm{o}(1), \qquad n \in \mathbb{N},$$

which is, unfortunately, a very wide range. Thus, we have to study local levels related to other level α GOF tests.

Once we have α_n^{loc}, one may redefine the corresponding GOF tests with equal local levels as minimum p-value (minP) tests based on the one-sided p-values $p_{i,n} = F_{i,n}(X_{i:n})$, $i = 1, \dots, n$. Setting $M_n^+ = \min_{1 \le i \le n} p_{i,n}$ and $M_n = \min_{1 \le i \le n}\{p_{i,n}, 1 - p_{i,n}\}$, we get $\varphi(\alpha_n^{\text{loc}}) = 1$ iff $M_n^+ \le \alpha_n^{\text{loc}}$ and $\tilde{\varphi}(\alpha_n^{\text{loc}}) = 1$ iff $M_n \le \alpha_n^{\text{loc}}/2$.

The minP statistics M_n^+ and M_n were already introduced by Berk and Jones in 1979 (cf. [4]) and they referred to these statistics as minimum level attained statistics. Implicitly, Berk and Jones were the first proposing the construction of equal local level GOF tests (even though they did not use the term local levels). Among others, they extensively studied M_n^+ and M_n with respect to optimality and Bahadur efficiency, see also [3]. A further representation of GOF tests with equal local levels was provided in the unpublished manuscript [5] in 2006. Moreover, such tests were recently provided by several authors. At the 7th International Conference on Multiple Comparison Procedures (MCP) 2011 we introduced the concept of local levels and proposed GOF tests with equal local levels as an improvement of the higher criticism (HC) tests. At the MCP 2013 we presented asymptotic as well as finite properties of GOF tests with equal local levels, cf. [14]. In contrast to the formulation via local levels, the GOF test in [5] is formulated in terms of bounding functions. This representation of the test is elaborated on in [1]. What is more, the same test is provided in [25] in the HC framework. Finally, the test is also considered in the preprints [20], [21] and [27].

1.5. Higher criticism and phi-divergence

In connection with high dimensional data and associated multiple testing issues, the so-called higher criticism (HC) tests generated considerable interest during the last decade, cf. for example, [8–10] and [16]. For example, Donoho and Jin proposed the use of HC tests when testing the global null hypothesis in high-dimensional models with sparse signals against some specific alternatives. Studying the HC test statistic they showed in [8] that, asymptotically, HC related tests are successful throughout the same region of amplitude sparsity where the corresponding oracle likelihood ratio test would succeed. This means that a further specification of an alternative is not necessary. Note that HC tests can also be seen as GOF tests of the type (1.2) and/or (1.3). What is more, it appears that studying local levels corresponding to HC tests is essential in order to construct new GOF tests, which have a high power against alternative distributions that mainly deviate from the null distribution in the considered range.

Alternatively, instead of HC tests one may consider other GOF tests which are based on the phi-divergences introduced in [19]. Thereby, the family of these tests is parametrized by $s \in [-1, 2]$ so that the HC test corresponds to $s = 2$, the Berk–Jones test to $s = 1$, the reversed Berk–Jones test to $s = 0$ and the studentized version of the HC test to $s = -1$. As suggested by a referee, we compare local levels for some selected s-values. Figure 2 shows two-sided local levels of the exact level α tests based on the phi-divergences for $\alpha = 0.05$, $n = 1000$ and $s = 2, 1.5, 1, 0.5, 0, -0.5, -1$. What these local levels have in common is that they are large in the tails and small and approximately equal in the central range. However, the range of the local level values is largest for the HC test and smallest for the Berk–Jones test. Therefore, it looks that the Berk–Jones test leads to the narrowest bounds for α_n^{loc} while the HC test to the widest ones. Due to the fact that under the null hypothesis statistics based on phi-divergences have the same asymptotic behavior in a specific range relevant for the asymptotics, any of these tests will lead to the same asymptotic results for most local levels. Therefore, it does not matter which test we consider. Since the tests with $s = 2$ (HC tests) and $s = -1$ (studentized HC tests) have the simplest representation of the form (1.2) and/or (1.3), we prefer to restrict attention to the original HC test, which has received a lot of attention during the past decade.

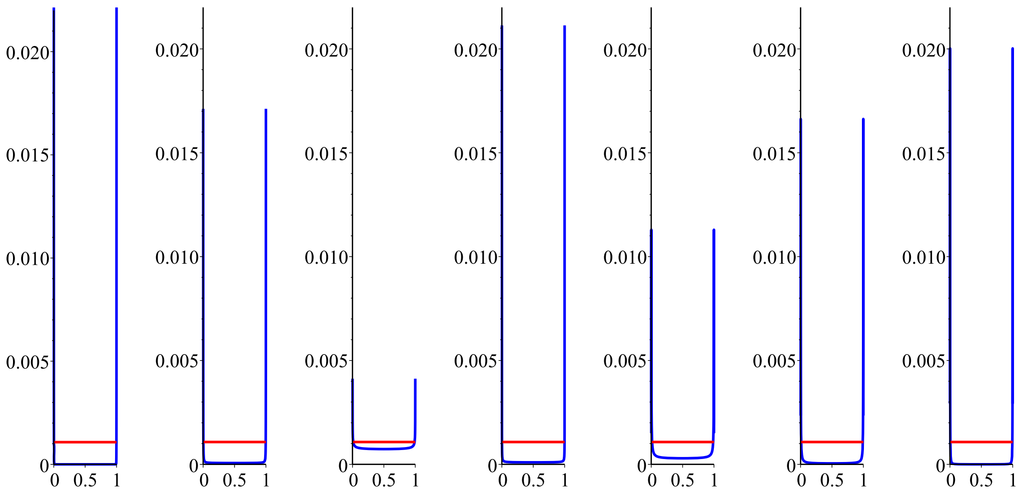

Figure 2. Two-sided local levels of level α GOF tests based on phi-divergences with $s = 2, 1.5, 1, 0.5, 0, -0.5, -1$ (from left to right) together with $\alpha_n^{\mathrm{loc}} = 0.001075$ related to the two-sided level α test $\tilde{\varphi}(\alpha_n^{\mathrm{loc}})$ for $\alpha = 0.05$ and $n = 1000$.

1.6. Outlook of the remaining part of the paper

In this paper, we calculate local levels of asymptotic level α HC tests and show that these local levels converge to zero as $n \to \infty$, which differs drastically from the KS case, cf. Figure 1. This implies for local levels of any asymptotic level α GOF test of the form (1.2) and/or (1.3) that

$$\lim_{n\to\infty} \min_{1\le i\le n} \alpha_{i,n} = 0,$$

that is, there are no level α tests for which the local levels are all asymptotically bounded away from zero. Finally, by a careful study of asymptotic HC local levels we get for $\varphi \equiv \varphi(\alpha_n^{\mathrm{loc}})$ and/or $\varphi \equiv \tilde{\varphi}(\alpha_n^{\mathrm{loc}})$ that

$$\lim_{n\to\infty} \mathbb{E}_0(\varphi) = \alpha \quad \text{iff} \quad \lim_{n\to\infty} \alpha_n^{\mathrm{loc}} \cdot \frac{2\log(\log(n))\log(n)}{-\log(1-\alpha)} = 1.$$

This result seems to be the most precise result concerning the asymptotics of the one- and two-sided GOF tests with equal local levels. In general, there are only few other works, in which asymptotics is investigated, cf. [20,21] and [27]. Due to a long revision process, some highlights of this paper have been already summarized in [14], where the focus lies on the sensitivity range of the HC tests statistic, extremely slow HC asymptotics, relations to the Ornstein–Uhlenbeck process, and power comparisons of the test with equal local levels and the original HC test. The remaining part of the paper is organized as follows. In Section 2, we study local levels of the HC test. We further derive the critical value and rejection curves corresponding to asymptotic level α HC tests and provide a result on the asymptotic behavior of the HC critical values. As zones of normal and Poisson convergence play a crucial role in the derivation of asymptotic results, we provide some basic results on these approximations for HC local levels in Section 3. Section 4 contains explicit asymptotic expressions of the local levels $\alpha_{i,n}$ of the one-sided HC test. They are derived for various growth rates of i utilizing the approximation results from Section 3. In Section 5, we investigate the asymptotic monotonicity of the local levels of one-sided HC tests and provide some results concerning the asymptotic behavior of local levels related to general level α GOF tests and tests with equal local levels. In Section 6, we compare classical KS and HC tests to GOF tests with equal local levels by means of numerical computations. Future investigations and open questions are discussed in Section 7. All proofs mostly of technical nature are deferred to Appendices A, B and C.

2. Higher criticism tests and local levels

First, we introduce the version of the higher criticism GOF tests that we are dealing with. Let

$$G_{i,n}(u) = \sqrt{n}\frac{i/n - u}{\sqrt{u(1-u)}} \quad \text{and} \quad \tilde{G}_{i,n}(u) = \sqrt{n}\frac{u - (i-1)/n}{\sqrt{u(1-u)}}, \qquad u \in (0,1).$$

A class of one-sided and two-sided HC test statistics can be expressed as

$$\mathrm{HC}^{+} = \max_{1\le i\le n} G_{i,n}(X_{i:n}) \quad \text{and} \quad \mathrm{HC}^{=} = \max_{1\le i\le n} \{G_{i,n}(X_{i:n}), \tilde{G}_{i,n}(X_{i:n})\},$$

respectively. A one-sided HC test based on a critical value d rejects $H_0^{\le}$ iff $\mathrm{HC}^+ > d$ and a two-sided HC test with the same critical value rejects $H_0^{=}$ iff $\mathrm{HC}^= > d$. In accordance with (1.4), local levels of a one-sided HC test based on a critical value $d > 0$ are given by

$$\alpha_{i,n} = \mathbb{P}\big(G_{i,n}(U_{i:n}) > d\big), \qquad i = 1, \dots, n. \tag{2.1}$$

Analogously, local levels of the corresponding two-sided HC test are given by

$$\alpha_{i,n}^{=} = \mathbb{P}\big(\big\{G_{i,n}(U_{i:n}) > d\big\} \cup \big\{\tilde{G}_{i,n}(U_{i:n}) > d\big\}\big), \qquad i = 1, \dots, n. \tag{2.2}$$

Setting $u_n = \log(\log(n))$ and

$$d_n(t) = \big(t + 2u_n + \big(\log(u_n) - \log(\pi)\big)/2\big)/\sqrt{2u_n}, \tag{2.3}$$

the asymptotic distributions of HC^+ and $\mathrm{HC}^=$ are given by

$$\lim_{n\to\infty} \mathbb{P}\big(\mathrm{HC}^+ \le d_n(t) | H_0^{=}\big) = \exp\big(-\exp(-t)\big) \tag{2.4}$$

and

$$\lim_{n\to\infty} \mathbb{P}\big(\mathrm{HC}^= \le d_n(t) | H_0^{=}\big) = \exp\big(-2\exp(-t)\big), \tag{2.5}$$

respectively, cf. [11] and [18]. For $t = t_\alpha$ or $t = t_\alpha^=$ with

$$t_\alpha = -\log\big(-\log(1-\alpha)\big) \quad \text{and} \quad t_\alpha^= = -\log\big(-\log(1-\alpha)/2\big)$$

we get a one-sided or two-sided asymptotic level α HC test, respectively. Note that

$$\big\{G_{i,n}(U_{i:n}) > d\big\} = \big\{U_{i:n} < h_{i,n}(d)\big\} \quad \text{and} \quad \big\{\tilde{G}_{i,n}(U_{i:n}) > d\big\} = \big\{U_{i:n} > \tilde{h}_{i,n}(d)\big\},$$

where

$$h_{i,n}(d) = \frac{d^2 + 2i - d\sqrt{d^2 + 4i - 4i^2/n}}{2(d^2+n)} \tag{2.6}$$

and

$$\tilde{h}_{i,n}(d) = \frac{d^2 + 2(i-1) + d\sqrt{d^2 + 4(i-1) - 4(i-1)^2/n}}{2(d^2+n)}. \tag{2.7}$$

Thereby, $\tilde{h}_{i,n}(d) > h_{i,n}(d)$, $i = 1, \dots, n$, for d large enough. Below, let $d \ge 1$, which guarantees $\tilde{h}_{i,n}(d) > h_{i,n}(d)$ for all $i = 1, \dots, n$ and $\{U_{i:n} < h_{i,n}(d)\} \cap \{U_{i:n} > \tilde{h}_{i,n}(d)\} = \varnothing$. Thus, local levels can be expressed as

$$\alpha_{i,n} = \mathbb{P}\big(U_{i:n} < h_{i,n}\big(d_n(t)\big)\big), \qquad i = 1, \dots, n,$$

for one-sided HC tests and

$$\alpha_{i,n}^{=} = \mathbb{P}\big(U_{i:n} < h_{i,n}\big(d_n(t)\big)\big) + \mathbb{P}\big(U_{i:n} > \tilde{h}_{i,n}\big(d_n(t)\big)\big), \qquad i = 1, \dots, n,$$

for two-sided HC tests. Assuming that Z_n ($\tilde{Z}_n$) is a binomially distributed random variable with parameters n and $h_{i,n}(d_n(t))$ ($\tilde{h}_{i,n}(d_n(t))$), that is, $Z_n \sim \mathcal{B}(n, h_{i,n}(d_n(t)))$, and $\tilde{Z}_n \sim \mathcal{B}(n, \tilde{h}_{i,n}(d_n(t)))$, we get

$$\alpha_{i,n} = \mathbb{P}(Z_n \geq i) \quad \text{and} \quad \alpha_{i,n}^{=} = \mathbb{P}(Z_n \geq i) + \mathbb{P}(\tilde{Z}_n < i). \tag{2.8}$$

Since $\tilde{h}_{i,n}(d) = 1 - h_{n,n-i+1}(d)$ for $i = 1, \dots, n$ and $n \in \mathbb{N}$, we obtain

$$\alpha_{i,n}^{=} = \alpha_{i,n} + \alpha_{n-i+1,n}, \qquad i = 1, \dots, n.$$

Hence, two-sided local levels are symmetric in the sense $\alpha_{i,n}^{=} = \alpha_{n-i+1,n}^{=}$ for $i = 1, \dots, n$, and can be easily calculated if one-sided local levels are known.

Note that the considered HC tests (and a lot of multiple tests) can be alternatively defined in terms of a rejection curve, which is a general inverse of the corresponding critical value curve, cf. [13]. Critical value curves related to (2.6) and (2.7) are given by

$$\rho_n(x, d) = \frac{d^2 + 2xn - d\sqrt{d^2 + 4xn - 4x^2 n}}{2(d^2 + n)} \tag{2.9}$$

and

$$\tilde{\rho}_n(x, d) = \frac{d^2 + 2(xn - 1) + d\sqrt{d^2 + 4(xn - 1) - 4(xn - 1)^2/n}}{2(d^2 + n)}, \tag{2.10}$$

respectively, that is, $h_{i,n}(d) = \rho_n(i/n, d)$ and $\tilde{h}_{i,n}(d) = \tilde{\rho}_n(i/n, d)$. The corresponding rejection curves are given by

$$r_n(x, d) = x + d\sqrt{\frac{x(1-x)}{n}} \quad \text{and} \quad \tilde{r}_n(x, d) = x + \frac{1}{n} - d\sqrt{\frac{x(1-x)}{n}},$$

respectively. It holds $\rho_n(x, d) = 1 - \tilde{\rho}_n(1 - x + 1/n, d)$ and $r_n(x, d) = 1 - \tilde{r}_n(1 - x, d) + 1/n$. Figure 3 shows critical value curves ρ_n, $\tilde{\rho}_n$ and the corresponding rejection curves r_n, $\tilde{r}_n$ for $n = 1000$ and $d = 10$. For increasing n and/or decreasing d, the corresponding curves tend to the diagonal.

The following lemma shows the asymptotic behavior of the critical values $h_{i_n,n} \equiv h_{i_n,n}(d_n(t))$ for different ranks i.

Lemma 2.1. *Let $n \in \mathbb{N}$ and $i_n \in \mathbb{N}$ with $i_n \leq n$. It holds*:

(i) *if $i_n = \mathrm{o}(u_n)$ as $n \to \infty$, then*

$$\begin{aligned} \frac{n h_{i_n,n}}{i_n} &= \frac{i_n}{2u_n}\left(1 - \frac{\log(u_n) + 2t - \log(\pi)}{2u_n} - \frac{i_n}{u_n}\right) \\ &\quad + \mathrm{O}\left(\frac{i_n \log(u_n)^2 + i_n^3}{u_n^3}\right); \end{aligned} \tag{2.11}$$

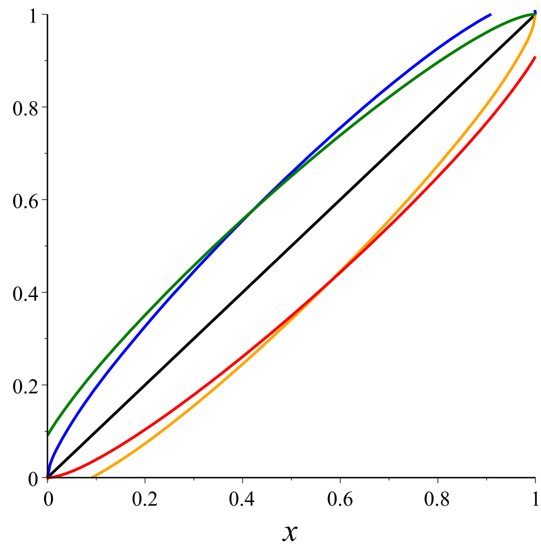

Figure 3. For $n = 1000$ and $d = 10$ the critical value curve $\rho_n(x, d)$ and the corresponding rejection curve $r_n(x, d)$ are given by the lowest and highest curves in $x = 0.8$, respectively; $\tilde{\rho}_n(x, d)$ and $\tilde{r}_n(x, d)$ are given by the highest and lowest curves in $x = 0.2$, respectively. The straight line is the diagonal.

(ii) *if* $c_n \equiv u_n/i_n \to c$ *for some* $c > 0$ *we obtain*

$$\frac{nh_{i_n,n}}{i_n} = \delta(c_n)\left(1 - \frac{\log(u_n) + 2t - \log(\pi)}{2i_n\sqrt{c_n^2 + 2c_n}}\right) + \mathrm{O}\left(\frac{\log(u_n)^2}{u_n^2}\right), \tag{2.12}$$

where $\delta(c_n) = 1 + c_n - \sqrt{c_n^2 + 2c_n} \in (0, 1)$;

(iii) *if* $i_n(1 - i_n/n)/u_n \to \infty$, *that is*, $i_n/u_n \to \infty$ *and* $(n - i_n)/u_n \to \infty$, *then*

$$\begin{aligned}\frac{nh_{i_n,n}}{i_n} &= 1 - \sqrt{\frac{2u_n}{i_n}\left(1 - \frac{i_n}{n}\right)} - \frac{\log(u_n) + 2t - \log(\pi)}{2\sqrt{2i_n u_n}}\sqrt{1 - \frac{i_n}{n}} \\ &\quad + \left(1 - \frac{2i_n}{n}\right)\frac{u_n}{i_n} + \mathrm{o}\left(\frac{u_n}{i_n} + \frac{1}{\sqrt{i_n u_n}}\sqrt{1 - \frac{i_n}{n}}\right);\end{aligned} \tag{2.13}$$

(iv) *if* $c_n \equiv (n - i_n)/u_n \to c$ *for some* $c \geq 0$ *we obtain*

$$\begin{aligned}\frac{nh_{i_n,n}}{i_n} &= 1 - \frac{u_n}{i_n}(1 + \sqrt{1 + 2c_n}) - \frac{\log(u_n) + 2t - \log(\pi)}{2i_n} \\ &\quad \times \left(1 + \frac{1 + c_n}{\sqrt{1 + 2c_n}}\right) + \mathrm{O}\left(\frac{\log(u_n)^2}{i_n u_n}\right).\end{aligned} \tag{2.14}$$

In the next sections, we provide asymptotics of local levels $\alpha_{i,n}$ of HC tests for all growth rates of i. To be precise, we are considering so-called extreme ranks i, where i or $n-i$ are fixed, and increasing ranks $i = i_n \to \infty$ as $n \to \infty$. We split the latter into central ranks, which are such that $i_n/n \to \zeta \in (0,1)$ as $n \to \infty$, and intermediate ranks, which are such that $i_n/n \to \zeta \in \{0,1\}$. For these concepts see, for example, [23].

3. Normal and Poisson approximations for local levels

Due to representation (2.8), we can approximate local levels of a HC test by applying Poisson and/or normal approximations for the binomial distribution. Below, let $Y_n \sim \mathcal{P}(nh_{i_n,n})$ and $\tilde{Y}_n \sim \mathcal{P}(n(1-h_{i_n,n}))$, where $\mathcal{P}(\lambda)$ denotes the Poisson distribution with parameter $\lambda > 0$. Thereby, $h_{i_n,n} = h_{i_n,n}(d_n(t))$ is given in Lemma 2.1.

The following theorem shows that for large values of n local levels $\alpha_{i_n,n}$ of HC tests based on critical values $h_{i_n,n}$, can be calculated by means of Poisson approximations for a wide range of ranks i_n.

Theorem 3.1 (Poisson approximation of local levels). *Let $i_n \in \mathbb{N}$, $i_n \le n$, be a sequence of non-decreasing numbers. For i_n such that $i_n = \mathrm{o}(u_n)$ we obtain*

$$\alpha_{i_n,n} = \mathbb{P}(Y_n = i_n)\big[1 + \mathrm{o}(1)\big], \tag{3.1}$$

for i_n such that $u_n/i_n \to c$ for some $c > 0$

$$\alpha_{i_n,n} = 1/\big(\sqrt{c^2 + 2c} - c\big)\mathbb{P}(Y_n = i_n)\big[1 + \mathrm{o}(1)\big] \tag{3.2}$$

and for i_n such that $i_n/u_n \to \infty$ and $i_n = \mathrm{o}(\sqrt{n/u_n})$

$$\alpha_{i_n,n} = \sqrt{i_n/(2u_n)}\mathbb{P}(Y_n = i_n)\big[1 + \mathrm{o}(1)\big]. \tag{3.3}$$

Analogously, for i_n with $n - i_n = \mathrm{o}(u_n)$, we get

$$\alpha_{i_n,n} = \mathbb{P}(\tilde{Y}_n = n - i_n)\big[1 + \mathrm{o}(1)\big], \tag{3.4}$$

for i_n fulfilling $(n - i_n)/u_n \to c$ for some $c > 0$ we obtain

$$\alpha_{i_n,n} = \big(1 + c/(1 + \sqrt{1 + 2c})\big)\mathbb{P}(\tilde{Y}_n = n - i_n)\big[1 + \mathrm{o}(1)\big] \tag{3.5}$$

and if $(n - i_n)/u_n \to \infty$ and $n - i_n = \mathrm{o}(\sqrt{n/u_n})$, then

$$\alpha_{i_n,n} = \sqrt{(n - i_n)/(2u_n)}\mathbb{P}(\tilde{Y}_n = n - i_n)\big[1 + \mathrm{o}(1)\big]. \tag{3.6}$$

The following theorem shows that local levels of HC tests corresponding to central ranks and to intermediate ranks close to central ones can be calculated in terms of the density of the standard normal distribution ϕ.

Theorem 3.2 (Normal approximation of local levels). *Let $i_n \in \mathbb{N}$ be such that $i_n(1 - i_n/n)/u_n^3 \to \infty$, $\sigma_n = \sqrt{nh_{i_n,n}(1-h_{i_n,n})}$ with $h_{i_n,n}$ given in (2.13) and $x_n = (i_n - nh_{i_n,n})/\sigma_n$. Then $x_n \to \infty$, $x_n^3/\sigma_n \to 0$ as $n \to \infty$ and*

$$\alpha_{i_n,n} = \phi(x_n)/x_n\bigl[1 + \mathrm{O}\bigl(x_n^3/\sigma_n + 1/x_n^2\bigr)\bigr]. \tag{3.7}$$

Proof. We can derive (3.7) by following the proof in [32], where he considered the case $p_n \equiv p$. Since $\alpha_{i_n,n} = \mathbb{P}(Z_n \geq i_n)$, where $Z_n \sim \mathcal{B}(n, h_{i_n,n})$, it suffices to show $x_n^3/\sigma_n \to 0$ if $i_n(1 - i_n/n)/u_n^3 \to \infty$. This can easily be proved by applying (2.13), which implies $\sigma_n = \sqrt{i_n(1-i_n/n)}[1 + \mathrm{o}(1)]$ and $x_n = \sqrt{2u_n}[1 + \mathrm{o}(1)]$. □

Remark 3.1. Note that for i_n satisfying $i_n(1 - i_n/n)/u_n^3 \to \infty$ as $n \to \infty$ and $i_n(1 - i_n/n) = \mathrm{o}(\sqrt{n/u_n})$, Theorems 3.1 and 3.2 provide two alternative approximations for local levels of HC tests.

4. Asymptotic expressions of local levels of HC tests

By means of Theorem 3.1 and the Stirling formula

$$i! = \sqrt{2\pi} i^{i+1/2} \exp(-i)\bigl[1 + \mathrm{O}(1/i)\bigr] \tag{4.1}$$

as well as Theorem 3.2, we are now able to calculate local levels $\alpha_{i,n}$ for various ranks i. Local levels $\alpha_{i,n}$ of HC tests with critical values $h_{i,n}(d_n(t))$ are given in Lemmas 4.1–4.5. For the sake of simplicity, we introduce the following notation for the different growth rates of i_n. We define the following sets of ranks $i_n \leq n$, $n \in \mathbb{N}$,

$$\begin{aligned}
A_c &\mathrel{\widehat{=}} i_n/u_n \to c \qquad \text{as } n \to \infty,\\
B_0 &\mathrel{\widehat{=}} i_n/u_n \to \infty \quad \text{and} \quad i_n/u_n^3 \to 0 \qquad \text{as } n \to \infty,\\
B_c &\mathrel{\widehat{=}} i_n/u_n^3 \to c > 0 \qquad \text{as } n \to \infty,\\
C &\mathrel{\widehat{=}} i_n(1 - i_n/n)/u_n^3 \to \infty \qquad \text{as } n \to \infty,\\
\bar{B}_c &\mathrel{\widehat{=}} (n - i_n)/u_n^3 \to c > 0 \qquad \text{as } n \to \infty,\\
\bar{B}_0 &\mathrel{\widehat{=}} (n - i_n)/u_n \to \infty \quad \text{and} \quad (n - i_n)/u_n^3 \to 0 \qquad \text{as } n \to \infty,\\
\bar{A}_c &\mathrel{\widehat{=}} (n - i_n)/u_n \to c \qquad \text{as } n \to \infty.
\end{aligned} \tag{4.2}$$

For example, for a sequence of ranks i_n, $n \in \mathbb{N}$, corresponding to A_c with $c = 0$ we write $i_n \in A_0$. Figure 4 summarizes which ranks i_n correspond to each lemma. In the next two lemmas, we state local levels of HC tests for extreme ranks and intermediate ranks close to extreme ones, that is, $i_n \in A_0 \cup \bar{A}_0$.

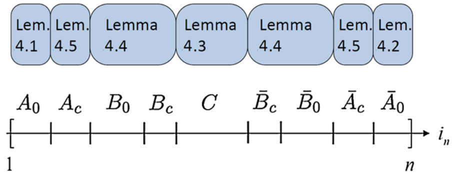

Figure 4. Diagram of the sets of ranks as defined in (4.2) and the corresponding lemmas in Section 4 which provide expressions for the local levels $\alpha_{i_n,n}$ for the different growth rates of i_n.

Lemma 4.1. *For $i_n \in A_0$, we obtain*

$$\alpha_{i_n,n} = \frac{1}{\sqrt{2\pi i_n}} \left(\gamma \frac{i_n}{u_n}\right)^{i_n} \exp(-i_n v_n)\left[1 + \mathrm{O}\left(\frac{1}{i_n}\right) + \mathrm{o}(1)\right], \tag{4.3}$$

where $\gamma = \exp(1)/2$ and

$$v_n = \big(\log(u_n) + 2t - \log(\pi) + 3i_n\big)/(2u_n)\big[1 + \mathrm{o}(1)\big]. \tag{4.4}$$

Alternatively, for $i_n \in A_0$ such that $i_n = \mathrm{o}(\sqrt{u_n})$ we get

$$\alpha_{i_n,n} = \left(\frac{i_n^2}{2u_n}\right)^{i_n} \frac{1}{i_n!}\big[1 + \mathrm{o}(1)\big]. \tag{4.5}$$

Lemma 4.2. *For $i_n \in \bar{A}_0$ we obtain*

$$\begin{aligned} \alpha_{i_n,n} = {} & \frac{\sqrt{\pi}\exp(-2t)}{\sqrt{2(n-i_n)}} \frac{1}{u_n \log(n)^2} \left(\frac{u_n}{\gamma(n-i_n)}\right)^{n-i_n} \exp\big((n-i_n)w_n\big) \\ & \times \big[1 + \mathrm{O}\big(1/(n-i_n)\big) + \mathrm{o}(1)\big], \end{aligned} \tag{4.6}$$

where $\gamma = \exp(1)/2$ and

$$w_n = \big(\log(u_n) + 2t - \log(\pi) + 3(n-i_n)\big)/(2u_n)\big[1 + \mathrm{o}(1)\big]. \tag{4.7}$$

Moreover, for $i_n \in \bar{A}_0$ such that $n - i_n = \mathrm{o}(\sqrt{u_n})$ we get

$$\alpha_{i_n,n} = \frac{\pi \exp(-2t)}{u_n \log(n)^2} \left(\frac{2u_n}{\exp(2)}\right)^{n-i_n} \frac{1}{(n-i_n)!}\big[1 + \mathrm{o}(1)\big]. \tag{4.8}$$

The following lemma contains an expression for local levels of HC tests for central ranks and intermediates close to central ranks, that is, $i_n \in C$.

Lemma 4.3. *Let $i_n \in C$. Then*

$$\alpha_{i_n,n} = \frac{\exp(-t)}{2u_n \log(n)} \left[1 + \mathrm{O}\left(\frac{\log(u_n)}{u_n} + \frac{u_n^{3/2}}{\sqrt{i_n(1 - i_n/n)}}\right)\right], \tag{4.9}$$

that is, local levels $\alpha_{i_n,n}$ with aforementioned i_n-values are asymptotically equal. Moreover, for a sequence $k_n \in \{1,\dots,n\}$ such that $k_n(1-k_n/n)/u_n^3 \to \infty$ as $n \to \infty$ and all $i_n = k_n,\dots,n-k_n$, local levels $\alpha_{i_n,n}$ converge uniformly.

The next lemma provides local levels of HC tests corresponding to intermediate ranks $i_n \in B_0 \cup B_c, \bar{B}_0 \cup \bar{B}_c$.

Lemma 4.4. *Let $i_n \in B_0 \cup B_c$ or $i_n \in \bar{B}_0 \cup \bar{B}_c$. Then*

$$\alpha_{i_n,n} = \frac{\exp(-t)}{2u_n \log(n)} \exp\left(\frac{\sqrt{2}\zeta_n}{3}\big(u_n + \mathrm{o}(u_n)\big)\right), \tag{4.10}$$

where $\zeta_n = \sqrt{u_n/i_n}$ if $i_n \in B_0 \cup B_c$ and $\zeta_n = -\sqrt{u_n/(n-i_n)}$ if $i_n \in \bar{B}_0 \cup \bar{B}_c$.

Finally, we give representations for local levels of HC tests for the remaining intermediate ranks $i_n \in A_c$ and $i_n \in \bar{A}_c$.

Lemma 4.5. *Let $i_n \in A_c$ for a $c > 0$ and set $c_n \equiv u_n/i_n$. Then*

$$\begin{aligned}\alpha_{i_n,n} = {} & \frac{\sqrt{c_n}}{(1-\delta(c_n))\sqrt{2\pi u_n}} \left[\frac{\delta(c_n)\exp(1)}{\exp(\delta(c_n))}\right]^{u_n/c_n} \\ & \times \big[\sqrt{\pi}\exp(-t)/\sqrt{u_n}\big]^{(1-\delta(c_n))/\sqrt{c_n^2+2c_n}}\big[1+\mathrm{o}(1)\big]\end{aligned} \tag{4.11}$$

with $\delta(c) = 1 + c - \sqrt{c^2+2c}$. If $i_n \in \bar{A}_c$, $c > 0$, and $c_n \equiv (n-i_n)/u_n$, then

$$\begin{aligned}\alpha_{i_n,n} = {} & \left(1 + \frac{c_n}{1+\sqrt{1+2c_n}}\right)\frac{1}{\sqrt{2\pi c_n u_n}}\left(\frac{1+c_n+\sqrt{1+2c_n}}{c_n}\right)^{c_n u_n} \\ & \times \big(\sqrt{\pi}\exp(-t)/\sqrt{u_n}\big)^{1+1/\sqrt{1+2c_n}} \log(n)^{-(1+\sqrt{1+2c_n})} \\ & \times \big[1+\mathrm{o}(1)\big].\end{aligned} \tag{4.12}$$

5. Monotonicity of HC local levels and related results

First, we briefly illustrate the behavior of one-sided local levels of HC tests for finite n-values. Figure 5 provides exactly calculated local levels $\alpha_{i,n} = \mathbb{P}(U_{i:n} < h_{i,n}(d))$ of HC tests φ^{HC} (say) with critical values $h_{i,n}(d)$, $i = 1,\dots,n$, for $n = 1000$ and $d = 1.5, 2.5, 3.5, 4.736$. For $d = 1.5, 2.5, 3.5, 4.736$ we get $\mathbb{E}_0(\varphi^{\mathrm{HC}}) = 0.803, 0.322, 0.111, 0.05$, respectively. That is, the HC test based on $d = 4.736$ is a level α GOF test for $\alpha = 0.05$. Figure 5 illustrates that local levels are decreasing for larger d-values. Noting that our asymptotic investigations are given for $d \equiv d_n$ tending to infinity, it seems that asymptotic results related to HC tests should be in accordance with the corresponding finite results for larger values of d.

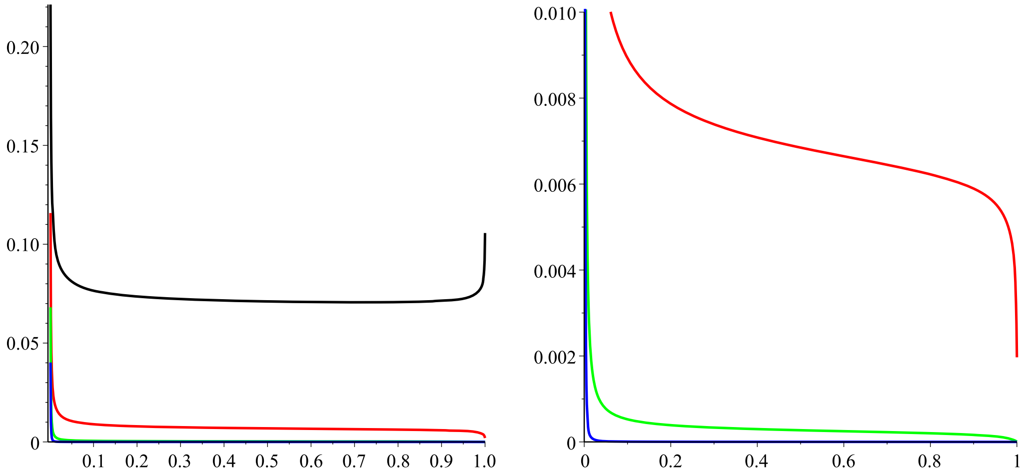

Figure 5. The left graph: exact local levels curves calculated for one-sided HC-tests with $n = 1000$ and $d = 1.5, 2.5, 3.5, 4.736$ (from top to bottom). The right graph is zoomed.

Indeed, the following theorem shows that local levels $\alpha_{i,n}$ of a HC test with critical values $h_{i,n}(d_n(t))$, $i = 1, \dots, n$, are asymptotically ($n \to \infty$) non-increasing in i in the following sense. For non-decreasing sequences $i_n^{(1)}$ and $i_n^{(2)}$ fulfilling $i_n^{(1)} < i_n^{(2)}$ for all $n \in \mathbb{N}$, we get $\lim_{n\to\infty} \alpha_{i_n^{(2)},n}/\alpha_{i_n^{(1)},n} \le 1$. More precisely, $\alpha_{i_n^{(2)},n}/\alpha_{i_n^{(1)},n}$ depends on the difference $i_n^{(2)} - i_n^{(1)}$ and/or the ratio $i_n^{(1)}/i_n^{(2)}$. Typically, the larger the difference $i_n^{(2)} - i_n^{(1)}$, the smaller the ratio $\alpha_{i_n^{(2)},n}/\alpha_{i_n^{(1)},n}$.

Theorem 5.1 (Asymptotic monotonicity of HC local levels). *Let $i_n^{(1)}$ and $i_n^{(2)}$ be non-decreasing sequences that satisfy $i_n^{(1)} < i_n^{(2)}$ for all $n \in \mathbb{N}$. Let $\alpha_{i,n}$ denote the ith local level corresponding to a HC test with critical values $h_{i,n}(d_n(t))$, $i = 1, \dots, n$. Then*

$$\lim_{n\to\infty} \alpha_{i_n^{(2)},n}/\alpha_{i_n^{(1)},n} = 1 \tag{5.1}$$

if the tuple $(i_n^{(1)}, i_n^{(2)})$ satisfies:

(i) $(i_n^{(1)}, i_n^{(2)}) \in C \times C$,
(ii) $(i_n^{(1)}, i_n^{(2)}) \in B_c \times B_c$,
(iii) $(i_n^{(1)}, i_n^{(2)}) \in \bar{B}_c \times \bar{B}_c$,
(iv) $(i_n^{(1)}, i_n^{(2)}) \in B_0 \times B_0$ *and* $i_n^{(1)}/i_n^{(2)} = 1 + \mathrm{o}(\sqrt{i_n^{(1)}/u_n^3})$,
(v) $(i_n^{(1)}, i_n^{(2)}) \in \bar{B}_0 \times \bar{B}_0$ *and* $(n - i_n^{(2)})/(n - i_n^{(1)}) = 1 + \mathrm{o}(\sqrt{(n - i_n^{(2)})/u_n^3})$.

Moreover, we have

$$0 < \lim_{n\to\infty} \alpha_{i_n^{(2)},n}/\alpha_{i_n^{(1)},n} < 1 \tag{5.2}$$

if one of the following conditions is satisfied:

(vi) $(i_n^{(1)}, i_n^{(2)}) \in B_{c_1} \times (B_{c_2} \cup C \cup \bar{B}_c)$ *with* $c_1 < c_2$,

(vii) $(i_n^{(1)}, i_n^{(2)}) \in (C \cup \bar{B}_{c_1}) \times \bar{B}_{c_2}$ *with* $c_2 < c_1$,

(vi) $(i_n^{(1)}, i_n^{(2)}) \in B_{c_1} \times (B_{c_2} \cup C \cup \bar{B}_c)$ *with* $c_1 < c_2$,

(vii) $(i_n^{(1)}, i_n^{(2)}) \in (C \cup \bar{B}_{c_1}) \times \bar{B}_{c_2}$ *with* $c_2 < c_1$,

(viii) $(i_n^{(1)}, i_n^{(2)}) \in B_0 \times B_0$ *and* $i_n^{(1)}/i_n^{(2)} = 1 - c_n\sqrt{i_n^{(1)}/u_n^3}$ *with* $c_n \to c > 0$,

(ix) $(i_n^{(1)}, i_n^{(2)}) \in \bar{B}_0 \times \bar{B}_0$ *and* $(n - i_n^{(2)})/(n - i_n^{(1)}) = 1 - c_n\sqrt{n - i_n^{(2)}/u_n^3}$, *with* $c_n \to c > 0$,

(x) $(i_n^{(1)}, i_n^{(2)}) \in A_c \times A_c$ *with* $c > 0$ *and* $i_n^{(2)} - i_n^{(1)} \equiv m$ *for an* $m \in \mathbb{N}$,

(xi) $(i_n^{(1)}, i_n^{(2)}) \in \bar{A}_c \times \bar{A}_c$ *with* $c > 0$ *and* $i_n^{(2)} - i_n^{(1)} \equiv m$ *for an* $m \in \mathbb{N}$.

Finally,

$$\lim_{n\to\infty} \alpha_{i_n^{(2)},n}/\alpha_{i_n^{(1)},n} = 0 \tag{5.3}$$

for all other tuples with $i_n^{(1)} < i_n^{(2)}$ *when this limit exists*.

Figure 6 illustrates the regions of validity of (i)−(xi) in Theorem 5.1. Since $\alpha_{1,n} \to 0$ as $n \to \infty$, cf. (4.5), and local levels $\alpha_{i,n}$, $i = 2, \dots, n$ are smaller than $\alpha_{1,n}$ for n large enough, cf. Theorem 5.1, the following result is obvious.

Theorem 5.2. *For the local levels of the HC test it holds that*

$$\lim_{n\to\infty} \max_{1\le i\le n} \alpha_{i,n} = 0.$$

Theorem 5.2 implies that local levels corresponding to a HC test show a completely different limiting behavior than the local levels corresponding to KS tests, cf. Figure 1. Moreover, the statement of Theorem 5.2 on the local levels of HC tests vanishing asymptotically allows us to

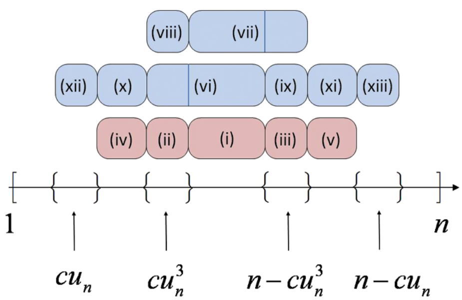

Figure 6. Diagram of the sets of ranks as defined in (4.2) and the corresponding regions covered by Theorem 5.1.

deduce a result on the more general case of asymptotic level α GOF tests with prespecified local levels $\alpha_{i,n}$, $i = 1, \dots, n$.

Remark 5.1. For a level α GOF test with $\alpha_{i,n}$ satisfying (1.4) or (1.5), we get

$$\lim_{n\to\infty} \min_{1\le i\le n} \alpha_{i,n} = 0 \quad \text{or} \quad \lim_{n\to\infty} \min_{1\le i\le n} \alpha_{i,n}^{=} = 0,$$

respectively. Thus, it is impossible to construct an asymptotic level α GOF test with local levels which are all asymptotically bounded away from zero.

Lemmas 4.1–4.5 and Theorem 5.1 lead to the next lemma that provides the asymptotics of level α GOF tests with equal local levels.

Lemma 5.1. *For one- or two-sided GOF tests with local levels equal to α_n^{loc}, $n \in \mathbb{N}$, we obtain an asymptotic level α test iff*

$$\lim_{n\to\infty} \alpha_n^{\text{loc}} \cdot \frac{2\log(\log(n))\log(n)}{-\log(1-\alpha)} = 1.$$

The rather technical and straightforward proof will be presented in a forthcoming paper.

Remark 5.2. Lemma 5.1 is up to now the most precise result concerning the asymptotics of the test with equal local levels. For example, adapting Theorem 4.1 in the third version of [27] leads to an asymptotic interval for α_n^{loc}. Moreover, results in [20] and [21] can be seen as a very rough approximation for the rate given in Lemma 5.1.

6. Comparison of GOF tests in the finite case

In this section, we compare one-sided versions of KS tests φ^{KS}, HC tests φ^{HC} and GOF tests $\varphi(\alpha_n^{\text{loc}})$ with equal local levels for a finite sample size n. In order to compare these tests in a fair way, all considered tests will be of exact level α. That is, for fixed $n \in \mathbb{N}$ and $\alpha \in (0, 1)$ we determine parameters of the considered tests, that is, find c for the KS test with critical values $i/n - c/\sqrt{n}$, $i = 1, \dots, n$, a parameter d for the HC test based on $h_{i,n}(d)$, $i = 1, \dots, n$, given in (2.6) and α_n^{loc} for the GOF test with equal local levels, so that

$$\mathbb{E}_0(\varphi^{\text{KS}}) = \mathbb{E}_0(\varphi^{\text{HC}}) = \mathbb{E}_0(\varphi(\alpha_n^{\text{loc}})) = \alpha.$$

Clearly, such parameters can be found numerically, for example, via some search algorithm, whenever the probability to reject the true null hypothesis can be numerically calculated. Thereby, the computation of the joint c.d.f. of the order statistics $U_{1:n}, \dots, U_{n:n}$, that is, $\mathbb{P}(U_{i:n} \le c_i, i = 1, \dots, n)$, plays the key role in the one-sided case, while the computation of $\mathbb{P}(c_i < U_{i:n} < \tilde{c}_i, i = 1, \dots, n)$ is crucial in the two-sided case. Probabilities of the first type can be calculated by Noe's, Bolshev's, Steck's or Khmaladze's recursions, $\mathbb{P}(c_i < U_{i:n} < \tilde{c}_i, i = 1, \dots, n)$ can be

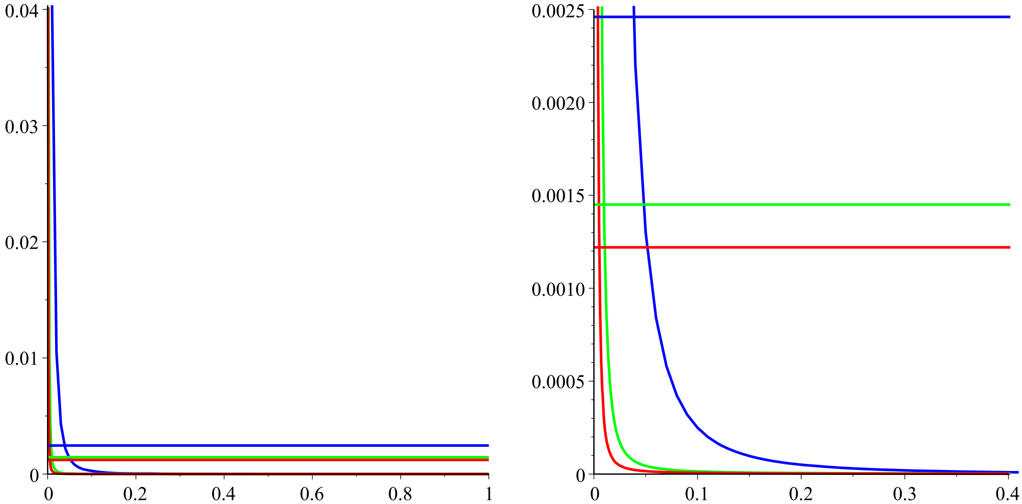

Figure 7. Local levels curves corresponding to HC tests φ^{HC} based on $h_{i,n}(d)$ with $d = 4.725$, 4.734, 4.736 (curves from top to bottom) for $n = 100, 500, 1000$, respectively, leading to $\mathbb{E}_0(\varphi^{\mathrm{HC}}) = 0.05$, and the corresponding local levels $\alpha_n^{\mathrm{loc}} = 0.00246, 0.00145, 0.00122$ (straight lines from top to bottom) that imply $\mathbb{E}_0(\varphi(\alpha_n^{\mathrm{loc}})) = 0.05$ for GOF tests with local levels equal α_n^{loc}. The right graph is zoomed.

calculated by Noe's, Ruben's or Khmaladze's recursions, for example, cf. [22] and pages 357–370 in [31]. If the sample size n is so large that exact computations are no longer possible, that is, $n \gg 10^4$, the parameters d and α_n^{loc} can approximately be calculated via numerical simulations.

For example, for $\alpha = 0.05$ and $n = 100$, 500, 1000 we get by numerical calculations $\mathbb{E}_0(\varphi^{\mathrm{HC}}) = \alpha$ for $d = 4.725$, 4.734, 4.736, respectively, $\mathbb{E}_0(\varphi^{\mathrm{KS}}) = \alpha$ for $c = 1.22387$ and $\mathbb{E}_0(\varphi(\alpha_n^{\mathrm{loc}})) = \alpha$ for $\alpha_n^{\mathrm{loc}} = 0.00246, 0.00145, 0.00122$, respectively. The asymptotic local level in Lemma 5.1 is equal to 0.00365, 0.00226, 0.00192 for $\alpha = 0.05$ and $n = 100, 500, 1000$, respectively, so that the asymptotic local level seems to be larger than the finite counterpart α_n^{loc}.

Figure 7 shows local levels curves of the level α HC tests together with equal local levels α_n^{loc} (straight lines) for $n = 100, 500, 1000$. Local levels of the corresponding KS tests are given in Figure 1. Note that almost all local levels of the HC tests are smaller than the corresponding α_n^{loc} and only the first ones are larger, for example, for $n = 100, 500, 1000$ we get $\alpha_{i,n} \geq \alpha_n^{\mathrm{loc}}$ if $i \leq 3, 4, 5$, respectively, and $\alpha_{i,n} < \alpha_n^{\mathrm{loc}}$ else. This indicates higher sensitivity of the GOF test with equal local levels in a specific intermediate range than by the HC tests.

Now we consider the aforementioned level α GOF tests in terms of their rejection curves. Figure 8 shows rejection curves for $n = 100$. Here, critical values induced by $\varphi(\alpha_n^{\mathrm{loc}})$ are larger than the corresponding HC critical values for $i \geq 4$ and only slightly smaller than the KS critical values in a specific central range, while the latter are considerably smaller in tails. Moreover, although all considered tests are level α tests, almost all of the HC critical values are considerably smaller than the corresponding critical values of the GOF test $\varphi(\alpha_n^{\mathrm{loc}})$ with equal local levels. It indicates that the smallest critical values have the biggest impact on $\mathbb{E}_0(\varphi)$ for any GOF test φ

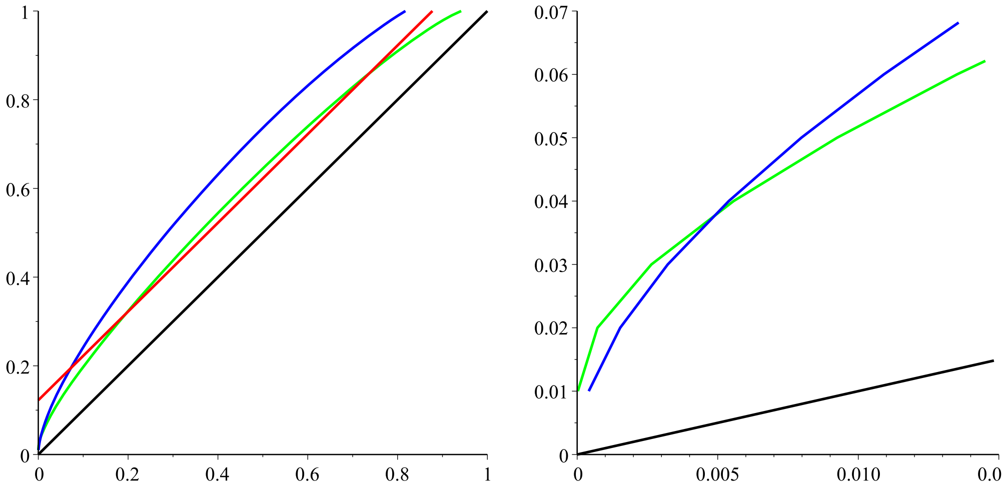

Figure 8. Rejection curves of the level α GOF tests φ^{HC}, $\varphi(\alpha_n^{\mathrm{loc}})$ and φ^{KS} together with the diagonal (from top to bottom in 0.5, respectively) for $n = 100$, $\alpha = 0.05$, φ^{HC} based on $h_{i,n}(d)$, $i = 1, \ldots, n$, with $d = 4.725$, $\varphi(\alpha_n^{\mathrm{loc}})$ based on $\alpha_n^{\mathrm{loc}} = 0.00246$ and φ^{KS} based on $i/n - c/\sqrt{n}$, $i = 1, \ldots, n$ with $c = 1.22387$. The right graph is zoomed.

while other critical values influence $\mathbb{E}_0(\varphi)$ only slightly. Further exact calculation showed that a similar picture is observed for various n-values.

Altogether, it seems that the level α GOF tests with equal local levels offer a good alternative to the classical GOF tests especially if it is not clear what kind of deviation from the null hypothesis may occur. For power comparisons between GOF tests with equal local levels and other GOF tests see [1,14] and [21].

7. Concluding remarks

In this paper, we introduced the concept of local levels $\alpha_{i,n}$ for a certain class of GOF tests. These quantities serve as an indicator of regions of high/low local sensitivity of a test and thus provide a method to compare tests with respect to areas of sensitivity. For example, the classical KS test has higher power for alternatives that differ from the null distribution in the central range. This coincides with the fact that local levels of the KS tests are considerably larger in the central range and are even equal to zero for extremes and smaller intermediates. In high-dimensional data with only sparse signals that are to be detected, it would be advisable to perform a GOF test (or related multiple tests) which is sensitive in the tails. In such situations performing HC tests, which are asymptotically sensitive only in the moderate tails, would be an advantage. Due to the fact that the number of local levels corresponding to central ranks is considerably higher than the number of local levels corresponding to intermediate ranks, one may guess that the HC local levels $\alpha_{i_n,n}$ for central ranks are much smaller than their counterparts in the moderate tails. Therefore, it is a rather striking result that central local levels are indeed asymptotically as large as the ones in the moderate tails. The reason for this may be hidden in the complex dependence structure of order statistics, so that a further investigation in this direction is needed. In general, it seems to be an interesting issue to analyze local levels of other multiple testing related GOF tests, thus gaining deeper insight into their nature. Figure 2 suggests that the Berk–Jones test comes close to the equal local levels test. It might be of interest to compare the asymptotic local levels of these tests as outlined in this paper for the HC test. An additional difficulty is that explicit critical values needed in (1.2) and (1.3) are hard to obtain for the Berk–Jones test as well as for most of the other phi-divergence tests.

Furthermore, the concept of local levels may be used to construct new tailored GOF tests if one has an idea in which region, that is, for which kinds of alternatives, a test needs to be sensitive. Given a set of suitable local levels we illustrated a way how to construct the corresponding GOF test. Moreover, by means of results related to the HC tests we showed that there is no level α GOF tests with local levels asymptotically uniformly bounded away from zero. In view of the fact that most of the HC local levels are asymptotically equal and that the first HC local levels are much too large so that the remaining ones are too small in the finite case, the GOF test with equal local levels $\alpha_{i,n} \equiv \alpha_n^{\mathrm{loc}}$ seems to be a good alternative for the classical HC test, which is known for its extremely slow asymptotics. Although we do not have any explicit formula for the local level α_n^{loc} as a function of the sample size n and predefined level α, we provide an asymptotic rate for α_n^{loc} leading to the asymptotic level α test.

Appendix A: Proofs of Sections 2 and 3

Proof of Lemma 2.1. Setting $A_n = 4(d_n^2 + n)i_n^2/(n(d_n^2 + 2i_n)^2)$, a critical value $h_{i_n,n} \equiv h_{i_n,n}(d_n(t))$ can be represented as

$$h_{i_n,n} = \bigl(d_n^2 + 2i_n\bigr)(1 - \sqrt{1 - A_n})/\bigl(2\bigl(d_n^2 + n\bigr)\bigr). \tag{A.1}$$

(i) Let i_n be such that $i_n = \mathrm{o}(u_n)$. Since $A_n = \mathrm{O}(i_n^2/u_n^2)$, $A_n \to 0$ as $n \to \infty$. Applying the Taylor series $1 - \sqrt{1-x} = x/2 + \mathrm{O}(x^2)$ for $x \in (0, 1)$, we get

$$\begin{aligned} h_{i_n,n} &= \bigl(d_n^2 + 2i_n\bigr)/\bigl(2\bigl(d_n^2 + n\bigr)\bigr)\bigl[2\bigl(d_n^2 + n\bigr)i_n^2/n/\bigl(d_n^2 + 2i_n\bigr)^2 + \mathrm{O}\bigl(i_n^4/u_n^4\bigr)\bigr] \\ &= i_n^2/\bigl(n\bigl(d_n^2 + 2i_n\bigr)\bigr) + \mathrm{O}\bigl(i_n^4/\bigl(nu_n^3\bigr)\bigr) \end{aligned}$$

and hence $nu_n h_{i_n,n}/i_n^2 = u_n/(d_n^2 + 2i_n) + \mathrm{O}(i_n^2/u_n^2)$. Noting that

$$d_n^2 = 2u_n + \log(u_n) + 2t - \log(\pi) + \mathrm{O}\bigl(\log(u_n)^2/u_n\bigr) \tag{A.2}$$

and $1/(2 + x) = 1/2 - x/4 + \mathrm{O}(x^2)$ for $x \in (0, 1)$, we get

$$\frac{u_n}{d_n^2 + 2i_n} = \frac{1}{2} - \frac{\log(u_n) + 2t - \log(\pi) + 2i_n}{4u_n} + \mathrm{O}\biggl(\frac{\log(u_n)^2 + i_n^2}{u_n^2}\biggr)$$

and consequently (2.11) follows.

(ii) Let $c_n \equiv u_n/i_n \to c$ as $n \to \infty$ for some fixed $c > 0$. Obviously, $A_n = [(d_n^2 + 2i_n)/(2i_n)]^{-2}(1 + \mathrm{O}(u_n/n))$. Since $(1+x)^{-2} = 1 - 2x + \mathrm{O}(x^2)$ for $x \in (0,1)$, (A.2) leads to

$$\left[\frac{d_n^2 + 2i_n}{2i_n}\right]^{-2} = \left[(1+c_n) + \frac{\log(u_n) + 2t - \log(\pi)}{2i_n} + \mathrm{O}\left(\frac{\log(u_n)^2}{u_n^2}\right)\right]^{-2}$$
$$= \frac{1}{(1+c_n)^2}\left[1 - \frac{\log(u_n) + 2t - \log(\pi)}{i_n(1+c_n)} + \mathrm{O}\left(\frac{\log(u_n)^2}{u_n^2}\right)\right].$$

Since $1 - \sqrt{1 - a(1-x)} = 1 - \sqrt{1-a} - ax/(2\sqrt{1-a}) + \mathrm{O}(x^2)$ for $a > 0$ and $x \in (0,1)$, we get

$$1 - \sqrt{1 - A_n} = 1 - \frac{\sqrt{c_n^2 + 2c_n}}{1 + c_n} - \frac{\log(u_n) + 2t - \log(\pi)}{2i_n(1+c_n)^2\sqrt{c_n^2 + 2c_n}} + \mathrm{O}\left(\frac{\log(u_n)^2}{u_n^2}\right).$$

Furthermore,

$$\frac{d_n^2 + 2i_n}{2(d_n^2 + n)} = \left(d_n^2 + 2i_n\right)/(2n)\left[1 + \mathrm{O}\left(d_n^2/n\right)\right] = (i_n/n)\left[1 + d_n^2/(2i_n) + \mathrm{O}(u_n/n)\right]$$
$$= (i_n/n)\left[1 + c_n + \frac{\log(u_n) + 2t - \log(\pi)}{2i_n} + \mathrm{O}\left(\log(u_n)^2/u_n^2\right)\right].$$

Formula (A.1) immediately leads to

$$h_{i_n,n} = i_n\left(1 + c_n - \sqrt{c_n^2 + 2c_n}\right)/n$$
$$\times \left[1 + \mathrm{O}\left(\log(u_n)^2/u_n^2\right) - \left(\log(u_n) + 2t - \log(\pi)\right)/\left(2i_n\sqrt{c_n^2 + 2c_n}\right)\right]$$

and hence, we get (2.12).

(iii), (iv) Now, let $u_n = \mathrm{o}(i_n)$. Due to $1/(1+x)^2 = 1 - 2x + 3x^2 + \mathrm{O}(x^3)$ for $x \in (0,1)$, we get

$$A_n = \left(d_n^2 + n\right)/n\left[\left(d_n^2 + 2i_n\right)/(2i_n)\right]^{-2} = \left(1 + \frac{d_n^2}{n}\right)\left[1 - \frac{d_n^2}{i_n} + \frac{3d_n^4}{4i_n^2} + \mathrm{O}\left(\frac{d_n^6}{i_n^3}\right)\right]$$
$$= 1 - \frac{d_n^2}{i_n}\left(1 - \frac{i_n}{n}\right) + \frac{d_n^4}{i_n^2}\left(\frac{3}{4} - \frac{i_n}{n}\right) + \mathrm{O}\left(\frac{d_n^6}{i_n^3}\right).$$

Hence, for i_n such that $u_n = \mathrm{o}(i_n(1 - i_n/n))$ we arrive at

$$A_n = 1 - 2u_n(1 - i_n/n)/i_n - \left(\log(u_n) + 2t - \log(\pi)\right)(1 - i_n/n)/i_n$$
$$+ \mathrm{O}\left(u_n^2/i_n^2 + \log(u_n)^2(1 - i_n/n)/(i_n u_n)\right)$$

and for i_n such that $n - i_n = \mathrm{O}(u_n)$ we obtain

$$A_n = 1 - u_n^2\left(1 + 2(n - i_n)/u_n\right)/i_n^2 - u_n\left(\log(u_n) + 2t - \log(\pi)\right)/i_n^2$$
$$\times \left(1 + (n - i_n)/u_n\right) + \mathrm{O}\left(\log(u_n)^2/i_n^2\right).$$

Then

$$1-\sqrt{1-A_n}=1-\sqrt{\frac{2u_n}{i_n}\left(1-\frac{i_n}{n}\right)}-\frac{\log(u_n)+2t-\log(\pi)}{2\sqrt{2i_nu_n}}\sqrt{1-i_n/n}$$
$$+\mathrm{O}\left(\frac{u_n^{3/2}}{i_n^{3/2}\sqrt{1-i_n/n}}+\frac{\log(u_n)^2}{\sqrt{i_n}u_n^{3/2}}\sqrt{1-i_n/n}\right)$$

for i_n such that $u_n=\mathrm{o}(i_n(1-i_n/n))$ and

$$1-\sqrt{1-A_n}=1-\frac{u_n}{i_n}\sqrt{1+2\frac{n-i_n}{u_n}}-\frac{\log(u_n)+2t-\log(\pi)}{2i_n}$$
$$\times\frac{1+(n-i_n)/u_n}{\sqrt{1+2(n-i_n)/u_n}}+\mathrm{O}\left(\frac{\log(u_n)^2}{i_nu_n}\right)$$

if $n-i_n=\mathrm{O}(u_n)$. Since

$$\frac{d_n^2+2i_n}{2(d_n^2+n)}=\left(\frac{i_n}{n}+\frac{d_n^2}{2n}\right)\left[1-\frac{d_n^2}{n}+\mathrm{O}\left(\frac{u_n^2}{n^2}\right)\right]$$
$$=\frac{i_n}{n}\left[1+\left(1-\frac{2i_n}{n}\right)\frac{2u_n+\log(u_n)+2t-\log(\pi)}{2i_n}\right.$$
$$\left.+\mathrm{O}\big(u_n^2/(ni_n)+(1-2i_n/n)\log(u_n)^2/(u_ni_n)\big)\right]$$

for all $i_n\le n$, we get for i_n such that $u_n=\mathrm{o}(i_n(1-i_n/n))$

$$h_{i_n,n}=\frac{i_n}{n}\left[1-\sqrt{\frac{2u_n}{i_n}\left(1-\frac{i_n}{n}\right)}-\frac{\log(u_n)+2t-\log(\pi)}{2\sqrt{2i_nu_n}}\sqrt{1-\frac{i_n}{n}}\right.$$
$$+(1-2i_n/n)\big(2u_n+\log(u_n)+2t-\log(\pi)\big)/(2i_n) \tag{A.3}$$
$$\left.+\mathrm{O}\left(\frac{u_n^{3/2}}{i_n^{3/2}\sqrt{1-i_n/n}}+\frac{\log(u_n)^2}{\sqrt{i_n}u_n^{3/2}}\sqrt{1-\frac{i_n}{n}}\right)\right]$$

and

$$h_{i_n,n}=\frac{i_n}{n}\left[1-\frac{u_n}{i_n}\big(1+\sqrt{1+2(n-i_n)/u_n}\big)-\frac{\log(u_n)+2t-\log(\pi)}{2i_n}\right.$$
$$\left.\times\left(1+\frac{1+(n-i_n)/u_n}{\sqrt{1+2(n-i_n)/u_n}}\right)+\mathrm{O}\left(\frac{\log(u_n)^2}{i_nu_n}\right)\right]$$

for i_n with $n-i_n=\mathrm{O}(u_n)$. □

Proof of Theorem 3.1. We only prove (3.1)–(3.3). The cases (3.4)–(3.6) can be handled analogously. For $n \geq 4$, $h_{i_n,n} \leq 1/4$ and $|i_n - nh_{i_n,n}| \leq \sqrt{n}/2$ we obtain

$$\frac{\mathbb{P}(X_n = i_n)}{\mathbb{P}(Y_n = i_n)} = 1 - \frac{(i_n - nh_{i_n,n})^2}{2n(1-h_{i_n,n})} + \frac{i_n}{2n} + \mathrm{O}\left(\frac{(i_n - nh_{i_n,n})^3}{n^2} + \frac{i_n^2}{n^2}\right),$$

cf. formula (17) in [29]. Therefore, (2.11)–(2.13) in Lemma 2.1 lead to

$$\alpha_{i_n,n} = \mathbb{P}(Y_n \geq i_n)\big[1 + \mathrm{o}(1)\big] \qquad \text{at least for } i_n = \mathrm{o}(\sqrt{n/u_n}). \tag{A.4}$$

Moreover, for $i_n = \mathrm{o}(\sqrt{n/u_n})$ we get

$$\alpha_{i_n,n} = \mathbb{P}(Y_n = i_n)\left(1 + \sum_{k=1}^{\infty}\prod_{j=1}^{k}\frac{nh_{i_n,n}}{i_n + j}\right)\big[1 + \mathrm{o}(1)\big],$$

cf. [31], page 485. Since $nh_{i_n,n}/i_n < 1$ for larger n-values, we get

$$1 < 1 + \sum_{k=1}^{\infty}\prod_{j=1}^{k}\frac{nh_{i_n,n}}{i_n + j} \leq \sum_{k=0}^{\infty}\left(\frac{nh_{i_n,n}}{i_n + 1}\right)^k = \frac{i_n + 1}{i_n - nh_{i_n,n} + 1}$$

for $i_n = \mathrm{o}(\sqrt{n/u_n})$. Hence, $\lim_{n\to\infty}(i_n + 1)/(i_n - nh_{i_n,n} + 1) = 1$ in case $i_n = \mathrm{o}(u_n)$, that is, (3.1) follows. For i_n such that $u_n/i_n \to c > 0$, $n \to \infty$,

$$\lim_{n\to\infty}\frac{i_n + 1}{i_n - nh_{i_n,n} + 1} = \frac{1}{\sqrt{c^2 + 2c} - c}$$

and for a fixed $k \in \mathbb{N}$

$$\lim_{n\to\infty}\left(\prod_{j=1}^{k}\frac{nh_{i_n,n}}{i_n + j}\right)\Big/\left(\frac{nh_{i_n,n}}{i_n + 1}\right)^k = 1,$$

which implies (3.2). Furthermore, from known asymptotic decompositions for the incomplete gamma function (e.g., cf. [12,17] and [2], page 140) and from the fact that for $i_n = \mathrm{o}(\sqrt{n/u_n})$ such that $u_n = \mathrm{o}(i_n)$ it holds $i_n \to \infty$, $nh_{i_n,n} \to \infty$ and $(i_n - nh_{i_n,n})/\sqrt{nh_{i_n,n}} \to \infty$, $n \to \infty$, we obtain

$$\mathbb{P}(Y_n \geq i_n) = i_n/(i_n - nh_{i_n,n} - 1)\mathbb{P}(Y_n = i_n)\big[1 + \mathrm{o}(1)\big].$$

This together with (2.13) and (A.4) imply (3.3). □

Appendix B: Proofs of Section 4

Proof of Lemma 4.1. With respect to Theorem 3.1, it suffices to calculate $\mathbb{P}(Y_n = i_n)$, where $Y_n \sim \mathcal{P}(nh_{i_n,n})$. Obviously,

$$\mathbb{P}(Y_n = i_n) = \big(i_n^{i_n}/i_n!\big)\big[(nh_{i_n,n}/i_n)\exp(-nh_{i_n,n}/i_n)\big]^{i_n}.$$

Since (2.11) implies $nh_{i_n,n}/i_n = \mathrm{o}(1)$ and $x\exp(-x) = x - x^2 + \mathrm{O}(x^3)$ for $x \in (0, 1)$, we obtain

$$\mathbb{P}(Y_n = i_n) = \big(i_n^{i_n}/i_n!\big)\big[(nh_{i_n,n}/i_n) - (nh_{i_n,n}/i_n)^2 + \mathrm{O}(nh_{i_n,n}/i_n)^3\big]^{i_n}.$$

Setting representation (2.11) for a critical value $h_{i_n,n}$ in the equation above, we get

$$\mathbb{P}(Y_n = i_n) = (B_n/i_n!)\big(i_n^2/2u_n\big)^{i_n}, \tag{B.1}$$

where

$$\begin{aligned}
B_n &= \left(1 - \frac{\log(u_n) + 2t - \log(\pi) + 3i_n}{2u_n} + \mathrm{O}\left(\frac{\log(u_n)^2 + i_n^2}{u_n^2}\right)\right)^{i_n} \\
&= \exp\Big(i_n \log\Big(1 - \big(\log(u_n) + 2t - \log(\pi) + 3i_n\big)/(2u_n) + \mathrm{O}\big(\big(\log(u_n)^2 + i_n^2\big)/u_n^2\big)\Big)\Big).
\end{aligned}$$

Since $i_n/u_n \to 0$ as $n \to \infty$ and $\log(1 - x) = -x + \mathrm{O}(x^2)$ as $x \to 0$, it follows

$$B_n = \exp\left(-i_n\left[\frac{\log(u_n) + 2t - \log(\pi) + 3i_n}{2u_n} + \mathrm{O}\left(\frac{\log(u_n)^2 + i_n^2}{u_n^2}\right)\right]\right). \tag{B.2}$$

Particularly, for $i_n = \mathrm{o}(\sqrt{u_n})$ we get $B_n = 1 + \mathrm{o}(1)$, so that (B.1) immediately leads to (4.5). Finally, the Stirling formula (4.1), (B.1) and (B.2) imply (4.3). □

Proof of Lemma 4.2. Due to Theorem 3.1, we have to calculate

$$\mathbb{P}(\tilde{Y}_n = n - i_n) = (n - nh_{i_n,n})^{n-i_n}\exp(-n + nh_{i_n,n})/(n - i_n)!.$$

Setting $c_n \equiv (n - i_n)/u_n$, we get $c_n \to 0$ as $n \to \infty$. In order to simplify (2.14), we obtain $1 + \sqrt{1 + 2c_n} = 2 + c_n - c_n^2/2 + \mathrm{O}(c_n^3)$ and $1 + (1 + c_n)/\sqrt{1 + 2c_n} = 2 + \mathrm{O}(c_n^2)$. Then

$$\begin{aligned}
nh_{i_n,n} &= i_n - u_n\big(2 + c_n - c_n^2/2\big) - \log(u_n) - 2t + \log(\pi) \\
&\quad + \mathrm{O}\big(u_n c_n^3 + \log(u_n)c_n^2 + \log(u_n)^2/u_n\big), \\
\exp(-n + nh_{i_n,n}) &= \frac{\pi\exp(-2t)}{u_n\log(n)^2\exp(2(n - i_n))}\big[1 + \mathrm{O}\big(\log(u_n)^2/u_n\big)\big] \\
&\quad \times \exp\big((n - i_n)\big[c_n/2 + \mathrm{O}\big(c_n^2 + c_n\log(u_n)/u_n\big)\big]\big)
\end{aligned} \tag{B.3}$$

and

$$\begin{aligned}
&(n - nh_{i_n,n})^{n-i_n} \\
&\quad = (2u_n)^{n-i_n}\exp\big((n - i_n)\log\big((n - nh_{i_n,n})/(2u_n)\big)\big) \\
&\quad = (2u_n)^{n-i_n}\exp\left((n - i_n)\log\left(1 + c_n + \frac{\log(u_n) + 2t - \log(\pi)}{2u_n} + \mathrm{O}\big(c_n^2 + \log(u_n)^2/u_n^2\big)\right)\right).
\end{aligned}$$

Taylor's series $\log(1+x) = x + \mathrm{O}(x^2)$ for $x \in (0, 1)$ leads to

$$\begin{aligned}&(n - nh_{i_n,n})^{n-i_n}\\ &\quad = (2u_n)^{n-i_n} \exp\biggl((n - i_n)\biggl(c_n + \frac{\log(u_n) + 2t - \log(\pi)}{2u_n} + \mathrm{O}\bigl(c_n^2 + \log(u_n)^2/u_n^2\bigr)\biggr)\biggr).\end{aligned}$$

Combining (3.4), (B.3) and the last expression we get (4.8) in case $n - i_n = \mathrm{o}(\sqrt{u_n})$ and applying Stirling's formula (4.1) to $(n - i_n)!$ we get (4.6). □

Proof of Lemma 4.3. Formula (3.7) in Theorem 3.2 implies

$$\alpha_{i_n,n} = \exp\bigl(-x_n^2/2\bigr)/(\sqrt{2\pi}x_n)\bigl[1 + \mathrm{O}\bigl(1/x_n^2 + x_n^3/\sqrt{nh_{i_n,n}(1 - h_{i_n,n})}\bigr)\bigr]. \tag{B.4}$$

First, we have to calculate x_n. From (2.13), we get

$$nh_{i_n,n}(1 - h_{i_n,n}) = i_n(1 - i_n/n)\bigl[1 + \mathrm{O}\bigl(\sqrt{u_n/\bigl(i_n(1 - i_n/n)\bigr)}\bigr)\bigr]$$

and hence

$$\sqrt{nh_{i_n,n}(1 - h_{i_n,n})} = \sqrt{i_n(1 - i_n/n)}\bigl[1 + \mathrm{O}\bigl(\sqrt{u_n/\bigl(i_n(1 - i_n/n)\bigr)}\bigr)\bigr].$$

Regarding to (A.3), we arrive at

$$i_n - nh_{i_n,n} = \sqrt{2i_n(1 - i_n/n)u_n}\bigl[1 + \bigl(\log(u_n) + 2t - \log(\pi)\bigr)/(4u_n) + \mathrm{O}\bigl(\varepsilon_n(i_n)\bigr)\bigr],$$

where $\varepsilon_n(i_n) = \sqrt{u_n/(i_n(1 - i_n))} + \log(u_n)^2/u_n^2$. Hence,

$$\begin{aligned}x_n &= \frac{\sqrt{2i_n(1 - i_n/n)u_n}[1 + (\log(u_n) + 2t - \log(\pi))/(4u_n) + \mathrm{O}(\varepsilon_n(i_n))]}{\sqrt{i_n(1 - i_n/n)}[1 + \mathrm{O}(\sqrt{u_n/(i_n(1 - i_n/n))})]}\\ &= \sqrt{2u_n}\bigl[1 + \bigl(\log(u_n) + 2t - \log(\pi)\bigr)/(4u_n) + \mathrm{O}\bigl(\varepsilon_n(i_n)\bigr)\bigr]\end{aligned}$$

and

$$x_n^2 = 2u_n + \log(u_n) + 2t - \log(\pi) + \mathrm{O}\bigl(u_n\varepsilon_n(i_n)\bigr).$$

This, the fact that $1/x_n = 1/\sqrt{2u_n}[1 + \mathrm{O}(\log(u_n)/u_n)]$ and (B.4) lead to

$$\begin{aligned}\alpha_{i_n,n} &= 1/(2\sqrt{\pi u_n}) \exp\bigl(-u_n - \log(u_n)/2 - t + \log(\pi)/2\bigr)\\ &\quad \times \bigl[1 + \mathrm{O}\bigl(\log(u_n)/u_n + u_n^{3/2}/\sqrt{i_n(1 - i_n/n)}\bigr)\bigr]\end{aligned}$$

and hence (4.9) follows. □

Proof of Lemma 4.4. We restrict our attention to $i_n \in B_0 \cup B_c$. The other case can be proved similarly. Combining (3.3) and (4.1), we get

$$\alpha_{i_n,n} = C_n/(2\sqrt{\pi u_n})\bigl[1 + \mathrm{o}(1)\bigr] \tag{B.5}$$

with $C_n \equiv [(nh_{i_n,n}/i_n)\exp(1 - nh_{i_n,n}/i_n)]^{i_n}$. It holds

$$C_n = \exp\bigl(i_n \log\bigl((nh_{i_n,n}/i_n)\exp(1 - nh_{i_n,n}/i_n)\bigr)\bigr).$$

From (2.13), we get $1 - nh_{i_n,n}/i_n = \mathrm{O}(\sqrt{u_n/i_n})$. Applying $\log((1-x)\exp(x)) = -x^2/2 - x^3/3 + \mathrm{O}(x^4)$ for $x \in (0,1)$, we arrive at

$$C_n = \exp\bigl(i_n\bigl\{-(1 - nh_{i_n,n}/i_n)^2/2 - (1 - nh_{i_n,n}/i_n)^3/3 + \mathrm{O}\bigl((u_n/i_n)^2\bigr)\bigr\}\bigr).$$

Lemma 2.1 leads to

$$1 - nh_{i_n,n}/i_n = \sqrt{2u_n/i_n} + \frac{\log(u_n) + 2t - \log(\pi)}{2\sqrt{2i_n u_n}} - \frac{u_n}{i_n} + \mathrm{o}(u_n/i_n),$$

$$(1 - nh_{i_n,n}/i_n)^2 = \frac{2u_n}{i_n} + \frac{\log(u_n) + 2t - \log(\pi)}{i_n} - \frac{2\sqrt{2}u_n^{3/2}}{i_n^{3/2}} + \mathrm{o}\bigl(u_n^{3/2}/i_n^{3/2}\bigr)$$

and

$$(1 - nh_{i_n,n}/i_n)^3 = 2\sqrt{2}u_n^{3/2}/i_n^{3/2} + \mathrm{o}\bigl(u_n^{3/2}/i_n^{3/2}\bigr).$$

Then

$$\begin{aligned} C_n &= \exp\biggl(-u_n - \frac{\log(u_n) + 2t - \log(\pi)}{2} + \frac{\sqrt{2}u_n^{3/2}}{3\sqrt{i_n}} + \mathrm{o}\biggl(\frac{u_n^{3/2}}{\sqrt{i_n}}\biggr)\biggr) \\ &= \exp(-t)\sqrt{\pi}/\bigl(\log(n)\sqrt{u_n}\bigr)\exp\bigl(\sqrt{2}u_n^{3/2}/(3\sqrt{i_n})\bigl(1 + \mathrm{o}(1)\bigr)\bigr) \end{aligned}$$

and hence (B.5) yields (4.10) for i_n fulfilling $u_n = \mathrm{o}(i_n)$ and $i_n = \mathrm{O}(u_n^3)$. □

Proof of Lemma 4.5. Formulas (3.2) and (3.5) provide that in order to find $\alpha_{i_n,n}$ we have to calculate $\mathbb{P}(Y_n = i_n)$ and $\mathbb{P}(\tilde{Y}_n = n - i_n)$.

We start with the case $c_n \equiv u_n/i_n \to c > 0$. Noting that $(1-x)^k = \exp(-kx + \mathrm{O}(kx^2))$ for $x \in (0,1)$ and $k \in \mathbb{N}$, formula (2.12) implies

$$\biggl(\frac{nh_{i_n,n}}{i_n}\biggr)^{i_n} = \bigl(\delta(c_n)\bigr)^{i_n}\exp\biggl(-\frac{\log(u_n) + 2t - \log(\pi)}{2\sqrt{c_n^2 + 2c_n}} + \mathrm{O}\biggl(\frac{\log(u_n)^2}{u_n}\biggr)\biggr).$$

Applying the Stirling formula (4.1) and (2.12), we arrive at

$$\begin{aligned} \mathbb{P}(Y_n = i_n) &= (nh_{i_n,n}/i_n)^{i_n}/\sqrt{2\pi i_n}\exp(i_n - nh_{i_n,n}) \\ &= \bigl[\delta(c_n)\exp\bigl(1 - \delta(c_n)\bigr)\bigr]^{i_n}/\sqrt{2\pi i_n}\bigl(\sqrt{\pi}\exp(-t)/\sqrt{u_n}\bigr)^{(1-\delta(c_n))/\sqrt{c_n^2+2c_n}} \\ &\quad \times \bigl(1 + \mathrm{O}\bigl(\log(u_n)^2/u_n\bigr)\bigr). \end{aligned}$$

Therefore, (3.2) implies (4.11).

Now let i_n be such that $c_n \equiv (n-i_n)/u_n \to c > 0$. Similarly as above, (2.14) implies

$$\left(\frac{n-nh_{i_n,n}}{n-i_n}\right)^{n-i_n} = \left[(1+c_n+\sqrt{1+2c_n})/c_n\right]^{n-i_n} \times \exp\left(c_n/\sqrt{1+2c_n}\left(\log(u_n)+2t-\log(\pi)\right)/2+\mathrm{O}\left(\log(u_n)^2/u_n\right)\right).$$

Then

$$\begin{aligned}\mathbb{P}(\tilde{Y}_n = n-i_n) &= \left[\frac{n-nh_{i_n,n}}{n-i_n}\right]^{n-i_n} \Big/ \sqrt{2\pi(n-i_n)}\exp(nh_{i_n,n}-i_n) \\ &= \left[(1+c_n+\sqrt{1+2c_n})/c_n\right]^{n-i_n}/\sqrt{2\pi(n-i_n)} \\ &\quad\times \left(\sqrt{\pi}\exp(-t)/\sqrt{u_n}\right)^{1+1/\sqrt{1+2c_n}}\log(n)^{-(1+\sqrt{1+2c_n})} \\ &\quad\times \left(1+\mathrm{O}\left(\log(u_n)^2/u_n\right)\right)\end{aligned}$$

and (3.5) lead to (4.12). □

Appendix C: Proofs of Section 5

Proof of Theorems 5.1. Formula (5.1) for the case (i) immediately follows from Lemma 4.3. Here we prove (5.1) for (ii)–(v), (5.2) for (vi)–(ix), (5.3) for $(i_n^{(1)}, i_n^{(2)}) \in B_0 \times B_0$ such that (iv), (viii) are not fulfilled and (5.3) for $(i_n^{(1)}, i_n^{(2)}) \in \bar{B}_0 \times \bar{B}_0$ such that (v), (ix) are not fulfilled. Lemma C.2 shows (5.2) for (x), (xi), (5.3) for $(i_n^{(1)}, i_n^{(2)}) \in A_c \times A_c$, $c > 0$, such that (x) is not fulfilled and (5.3) for $(i_n^{(1)}, i_n^{(2)}) \in \bar{A}_c \times \bar{A}_c$, $c > 0$ such that (xi) is not fulfilled. The remaining cases for (5.3) are proved in Lemmas C.1, C.3, C.4, C.5 and C.6.

For $(i_n^{(1)}, i_n^{(2)}) \in (B_0 \cup B_c) \times (B_0 \cup B_c)$ Lemma 4.4 yields

$$\alpha_{i_n^{(2)},n}/\alpha_{i_n^{(1)},n} = \exp\left(-\sqrt{2}u_n^{3/2}/\left(3\sqrt{i_n^{(1)}}\right)\left(1-\sqrt{i_n^{(1)}/i_n^{(2)}}\right)\left[1+\mathrm{o}(1)\right]\right).$$

This implies (5.1) for (ii) and (5.2) for $(i_n^{(1)}, i_n^{(2)}) \in B_{c_1} \times B_{c_2}$ with $c_1 < c_2$, that is, (5.2) for a partial case of (vi). Moreover, Lemmas 4.3 and 4.4 immediately yield the remaining cases of (vi).

For $i_n^{(1)} \in B_0$ define $b_n \equiv u_n^{3/2}/\sqrt{i_n^{(1)}}(1-\sqrt{i_n^{(1)}/i_n^{(2)}})$. Clearly, we get (5.1) if $b_n \to 0$, (5.2) if $b_n \to c > 0$ and (5.3) in case $b_n \to \infty$. Note that

$$i_n^{(1)}/i_n^{(2)} = 1 - b_n\sqrt{i_n^{(1)}}/u_n^{3/2}\left(2+b_n\sqrt{i_n^{(1)}}/u_n^{3/2}\right).$$

If $b_n \to 0$ as $n \to \infty$, that is, $i_n^{(1)}/i_n^{(2)} = 1+\mathrm{o}(\sqrt{i_n^{(1)}}/u_n^{3/2})$, then we get (5.1) for (iv). We get (5.2) for (viii), when $b_n \to b$ for some $b > 0$ and (5.3) in case $b_n \to \infty$.

Finally, (iii), (v), (vii) and (ix) can be proved in a similar way. □

Lemma C.1. *Let* $\{i_n^{(1)}\}_{n\in\mathbb{N}}$ *and* $\{i_n^{(2)}\}_{n\in\mathbb{N}}$ *be such that* $i_n^{(1)} < i_n^{(2)}$, $n \in \mathbb{N}$, *and either* $(i_n^{(1)}, i_n^{(2)}) \in A_0 \times A_0$ *or* $(i_n^{(1)}, i_n^{(2)}) \in \bar{A}_0 \times \bar{A}_0$. *Then* (5.3) *is fulfilled.*

Proof. First, let $i_n^{(j)} \in A_0$, $j = 1, 2$. If $i_n^{(2)} \equiv i_2 \in \mathbb{N}$ for all $n \in \mathbb{N}$ and n is large enough, representation (4.5) immediately yields $\alpha_{i_2,n}/\alpha_{i_1,n} = \mathrm{O}((2u_n)^{i_1-i_2})$, and hence (5.3) follows. Furthermore, let $i_n^{(2)} \to \infty$ as $n \to \infty$. Since $1/\sqrt{2\pi i}$ and $\exp(-i v_n)$ in representation (4.3) decrease as i increases for a fixed larger n, in order to prove (5.3) it suffices to show that

$$B_n \equiv \log\bigl(\bigl(\gamma i_n^{(2)}/u_n\bigr)^{i_n^{(2)}} / \bigl(\gamma i_n^{(1)}/u_n\bigr)^{i_n^{(1)}}\bigr)$$

converges to $-\infty$ as $n \to \infty$. Setting $x_n \equiv i_n^{(2)} - i_n^{(1)}$, we obtain

$$B_n = i_n^{(1)}\bigl[\bigl(-x_n/i_n^{(1)}\bigr)\log(u_n/\gamma) + \bigl(1 + x_n/i_n^{(1)}\bigr)\log\bigl(i_n^{(2)}\bigr) - \log\bigl(i_n^{(1)}\bigr)\bigr].$$

If $d_n \equiv x_n/i_n^{(1)} \to d$ for a $d > 0$ or $d = \infty$, we get $B_n = -i_n^{(1)} d_n \log(u_n/\gamma)(1 + \mathrm{o}(1))$, that is, $B_n \to -\infty$ as $n \to \infty$. Hence, (5.3) is fulfilled.

For $i_n^{(j)}$, $j = 1, 2$, such that $x_n/i_n^{(1)} = \mathrm{o}(1)$ we get

$$B_n = -x_n \log(u_n/\gamma) + x_n \log\bigl(i_n^{(2)}\bigr) + i_n^{(1)} \log\bigl(1 + x_n/i_n^{(1)}\bigr).$$

Applying $\log(1 + x) = x + \mathrm{O}(x^2)$ for $x \in (0, 1)$ and the fact that $i_n^{(2)} = \mathrm{o}(u_n)$, we obtain $B_n = -x_n \log(u_n/\gamma)(1 + \mathrm{o}(1))$, and hence (5.3) follows.

Now, let $i_n^{(j)} \in \bar{A}_0$, $j = 1, 2$. For $i_n^{(1)} < i_n^{(2)}$ such that $n - i_n^{(1)}$ is fixed, formula (4.8) in Lemma 4.2 immediately leads to the assertion. For the case $n - i_n^{(1)} \to \infty$ as $n \to \infty$, due to (4.6) it suffices to consider

$$D_n \equiv \bigl(u_n/\bigl(\gamma(n - i_n)\bigr)\bigr)^{n-i_n} / \sqrt{n - i_n} \exp\bigl((n - i_n)w_n\bigr).$$

Since

$$D_n = \exp\bigl((n - i_n)\bigl[-\log(n - i_n)/\bigl(2(n - i_n)\bigr) + \log\bigl(u_n/\bigl(\gamma(n - i_n)\bigr)\bigr) + w_n\bigr]\bigr),$$

$\log(u_n/(\gamma(n - i_n))) \to \infty$ as $n \to \infty$, $\log(x)/x < 1$ for $x \geq 1$ and $w_n = \mathrm{o}(1)$, we arrive at $D_n = \exp((n - i_n)\log(u_n/(\gamma(n - i_n)))[1 + \mathrm{o}(1)])$. Thus, it suffices to show that

$$\log\bigl(\bigl(u_n/\bigl(\gamma\bigl(n - i_n^{(2)}\bigr)\bigr)\bigr)^{n-i_n^{(2)}} / \bigl(u_n/\bigl(\gamma\bigl(n - i_n^{(1)}\bigr)\bigr)\bigr)^{n-i_n^{(1)}}\bigr)$$

converges to $-\infty$ for $n \to \infty$. This can be proved similarly as before. □

Lemma C.2. *Let* $\{i_n^{(1)}\}_{n\in\mathbb{N}}$ *and* $\{i_n^{(2)}\}_{n\in\mathbb{N}}$ *be such that* $i_n^{(1)} < i_n^{(2)}$ *for* $n \in \mathbb{N}$. *We suppose that either* $\lim_{n\to\infty} u_n/i_n^{(j)} = c_j$, $j = 1, 2$, *or* $\lim_{n\to\infty}(n - i_n^{(j)}) = c_j$, $j = 1, 2$, *for arbitrary but fixed* $c_j > 0$. *Moreover, let* $m_n \equiv i_n^{(2)} - i_n^{(1)}$, $n \in \mathbb{N}$. *If* $m_n = m$ *for some fixed* $m \in \mathbb{N}$ *and all* $n \in \mathbb{N}$, (5.2) *is fulfilled and if* $m_n \to \infty$ *as* $n \to \infty$, (5.3) *is fulfilled.*

Proof. First, let $c_{j,n} \equiv u_n/i_n^{(j)} \to c_j > 0$, $j = 1, 2$. For i_n such that $c_n \equiv u_n/i_n \to c > 0$ formula (4.11) in Lemma 4.5 implies

$$\alpha_{i_n,n} = \exp\bigl(f_1(c_n)u_n + f_2(c_n)\log(u_n) + f_3(c_n) + \mathrm{o}(1)\bigr),$$

where

$$f_1(c) = (1/c)\log\bigl(\delta(c)/\exp\bigl(\delta(c)-1\bigr)\bigr), \qquad f_2(c) = -1 + c/\bigl(2\sqrt{c^2+2c}\bigr),$$
$$f_3(c) = \log\bigl(\sqrt{c}/\bigl(\bigl(1-\delta(c)\bigr)\sqrt{2\pi}\bigr)\bigr) + \bigl(1-\delta(c)\bigr)/\sqrt{c^2+2c}\log\bigl(\sqrt{\pi}\exp(-t)\bigr).$$

It follows

$$\alpha_{i_n^{(2)},n}/\alpha_{i_n^{(1)},n} = \exp\bigl(\bigl(f_1(c_{2,n}) - f_1(c_{1,n})\bigr)u_n + \bigl(f_2(c_{2,n}) - f_2(c_{1,n})\bigr)\log(u_n)$$
$$+ f_3(c_{2,n}) - f_3(c_{1,n}) + \mathrm{o}(1)\bigr).$$

Since $f_1(\cdot)$ is strictly increasing, (5.3) is fulfilled in case $c_1 > c_2$. Now, let $c_1 = c_2$. Hence,

$$\frac{\alpha_{i_n^{(2)},n}}{\alpha_{i_n^{(1)},n}} = \exp\bigl(\bigl(f_1(c_{2,n}) - f_1(c_{1,n})\bigr)u_n + \bigl(f_2(c_{2,n}) - f_2(c_{1,n})\bigr)\log(u_n) + \mathrm{o}(1)\bigr).$$

Setting $x_n \equiv i_n^{(2)} - i_n^{(1)}$ and noting that $x_n = \mathrm{o}(u_n)$, we get

$$c_{1,n} = c_{2,n}/\bigl(1 - x_n/i_n^{(2)}\bigr) = c_{2,n}\bigl[1 + c_{2,n}x_n/u_n + \mathrm{O}\bigl(x_n^2/u_n^2\bigr)\bigr],$$
$$f_1(c_{1,n}) = f_1(c_{2,n}) + c_{2,n}\Bigl[c_{2,n}/\sqrt{c_{2,n}^2 + 2c_{2,n}} - 1 - f_1(c_{2,n})\Bigr]x_n/u_n + \mathrm{O}\bigl(x_n^2/u_n^2\bigr)$$

and $f_2(c_{1,n}) = f_2(c_{2,n}) + \mathrm{O}(x_n/u_n)$. Therefore,

$$\alpha_{i_n^{(2)},n}/\alpha_{i_n^{(1)},n} = \exp\Bigl(x_n c_{2,n}\Bigl[1 + f_1(c_{2,n}) - c_{2,n}\Big/\sqrt{c_{2,n}^2 + 2c_{2,n}} + \mathrm{o}(1)\Bigr]\Bigr).$$

Since $1 + f_1(c) - c/\sqrt{c^2+2c} < 0$ for $c \in (0,\infty)$, we get (5.2) if $x_n = x$ for some $x \in \mathbb{N}$ and (5.3) if $x_n \to \infty$ as $n \to \infty$.

The case $c_{j,n} \equiv (n - i_n^{(j)})/u_n \to c_j$, $j = 1, 2$, can be proved similarly. □

Lemma C.3. *Let $\{i_n^{(1)}\}_{n\in\mathbb{N}}$ and $\{i_n^{(2)}\}_{n\in\mathbb{N}}$ be such that $i_n^{(1)} = \mathrm{o}(u_n)$ and $c_n \equiv u_n/i_n^{(2)} \to c$, $n \to \infty$, for some $c > 0$. Then (5.3) is fulfilled.*

Proof. Formulas (4.3) and (4.11) yield

$$\log\biggl(\frac{\alpha_{i_n^{(2)},n}}{\alpha_{i_n^{(1)},n}}\biggr) = u_n\biggl[-\biggl(\frac{1}{2} + \frac{1-\delta(c_n)}{2\sqrt{c_n^2+2c_n}}\biggr)\frac{\log(u_n)}{u_n} + \frac{\log(i_n^{(1)})}{2u_n} + \frac{i_n^{(1)}}{u_n}v_n$$
$$+ \frac{1}{c_n}\log\biggl(\frac{\delta(c_n)}{\exp(\delta(c_n)-1)}\biggr) - \frac{i_n^{(1)}}{u_n}\log\biggl(\gamma\frac{i_n^{(1)}}{u_n}\biggr) + \mathrm{o}(1)\biggr].$$

Since $i_n^{(1)}/u_n = \mathrm{o}(1)$ and $\lim_{x\to 0} x\log(x) = 0$ and $\log(\delta(c)/\exp(\delta(c)-1)) < 0$ for all $c \in (0,\infty)$, we obtain $\log(\alpha_{i_n^{(2)},n}/\alpha_{i_n^{(1)},n}) \to -\infty$ as $n \to \infty$ and hence (5.3) follows. □

Lemma C.4. *Let* $\{i_n^{(1)}\}_{n\in\mathbb{N}}$ *and* $\{i_n^{(2)}\}_{n\in\mathbb{N}}$ *be such that* $c_n \equiv u_n/i_n^{(1)} \to c$, $n \to \infty$, *for some* $c > 0$, $i_n^{(2)} = \mathrm{O}(u_n^3)$ *and* $u_n = \mathrm{o}(i_n^{(2)})$. *Then* (5.3) *is fulfilled.*

Proof. Formulas (4.10) and (4.11) imply

$$\log(\alpha_{i_n^{(2)},n}/\alpha_{i_n^{(1)},n}) = -\bigl[1 + (1/c_n)\log\bigl(\delta(c_n)/\exp\bigl(\delta(c_n)-1\bigr)\bigr)\bigr]u_n + \mathrm{o}(u_n).$$

Since $c_n \to c > 0$ as $n \to \infty$ and $\log(\delta(c)/\exp(\delta(c)-1))/c > -1$ for all $c \in (0,\infty)$, we immediately obtain (5.3). □

Lemma C.5. *Let* $\{i_n^{(1)}\}_{n\in\mathbb{N}}$ *and* $\{i_n^{(2)}\}_{n\in\mathbb{N}}$ *be such that* $n - i_n^{(1)} = \mathrm{O}(u_n^3)$, $u_n = \mathrm{o}(n - i_n^{(1)})$ *and* $c_n \equiv (n - i_n^{(2)})/u_n \to c$, $n \to \infty$, *for some* $c > 0$. *Then* (5.3) *is fulfilled.*

Proof. Formulas (4.10) and (4.12) lead to

$$\log\left(\frac{\alpha_{i_n^{(2)},n}}{\alpha_{i_n^{(1)},n}}\right) = -\left[\sqrt{1+2c_n} - c_n \log\left(\frac{1 + c_n + \sqrt{1+2c_n}}{c_n}\right)\right]u_n + \mathrm{o}(u_n).$$

Noting that $\sqrt{1+2c} - c\log((1+c+\sqrt{1+2c})/c) > 0$ for all $c \in (0,\infty)$, we get (5.3). □

Lemma C.6. *Let* $\{i_n^{(1)}\}_{n\in\mathbb{N}}$ *and* $\{i_n^{(2)}\}_{n\in\mathbb{N}}$ *satisfy* $c_n \equiv (n - i_n^{(1)})/u_n \to c$, $n \to \infty$, *for some* $c > 0$ *and* $n - i_n^{(2)} = \mathrm{o}(u_n)$. *Then* (5.3) *is fulfilled.*

Proof. Formulas (4.6) and (4.12) lead to

$$\log\left(\frac{\alpha_{i_n^{(2)},n}}{\alpha_{i_n^{(1)},n}}\right) = \left[-1 + \sqrt{1+2c_n} - c_n \log\left(\frac{1 + c_n + \sqrt{1+2c_n}}{c_n}\right)\right]u_n + \mathrm{o}(u_n).$$

Since $-1 + \sqrt{1+2c} - c\log((1+c+\sqrt{1+2c})/c) < 0$ for all $c \in (0,\infty)$, (5.3) is fulfilled. □

Acknowledgements

The authors sincerely thank the referees for numerous helpful and constructive comments and suggestions and several hints for additional references. Special thanks are due to the Editor E. Moulines for his patience in handling the manuscript. This work was supported by the Ministry of Science and Research of the State of North Rhine-Westphalia (MIWF NRW) and the German Federal Ministry of Health (BMG).

References

[1] Aldor-Noiman, S., Brown, L., Buja, A., Rolke, W. and Stine, R. (2013). The power to see: A new graphical test of normality. *Am. Stat.* **67** 249–260. MR3280623

[2] Bateman, H. and Erdelyi, A. (1953). *Higher Transcendental Functions*, *Vol*. 2. New York: McGraw-Hill.

[3] Berk, R.H. and Jones, D.H. (1978). Relatively optimal combinations of test statistics. *Scand. J. Stat.* **5** 158–162. MR0509452

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

[5] Buja, A. and Rolke, W. (2006). Calibration for simultaneity: (Re)Sampling methods for simultaneous inference with applications to function estimation and functional data. Unpublished.

[6] Cai, T.T., Jin, J. and Low, M.G. (2007). Estimation and confidence sets for sparse normal mixtures. *Ann. Statist.* **35** 2421–2449. MR2382653

[7] David, H.A. and Nagaraja, H.N. (2003). *Order Statistics*, 3rd ed. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley. MR1994955

[8] Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195

[9] Donoho, D. and Jin, J. (2008). Higher criticism thresholding: Optimal feature selection when useful features are rare and weak. *Proc. Natl. Acad. Sci. USA* **105** 14790–14795.

[10] 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

[11] Eicker, F. (1979). The asymptotic distribution of the suprema of the standardized empirical processes. *Ann. Statist.* **7** 116–138. MR0515688

[12] Ferreira, C., López, J.L. and Pérez Sinusía, E. (2005). Incomplete gamma functions for large values of their variables. *Adv. in Appl. Math.* **34** 467–485. MR2123546

[13] Finner, H., Dickhaus, T. and Roters, M. (2009). On the false discovery rate and an asymptotically optimal rejection curve. *Ann. Statist.* **37** 596–618. MR2502644

[14] Gontscharuk, V., Landwehr, S. and Finner, H. (2014). The intermediates take it all: Asymptotics of higher criticism statistics and a powerful alternative based on equal local levels. *Biom. J.* **57** 159–180.

[15] Hall, P. and Jin, J. (2008). Properties of higher criticism under strong dependence. *Ann. Statist.* **36** 381–402. MR2387976

[16] Hall, P. and Jin, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732. MR2662357

[17] Ivchenko, G.I. (1974). On comparison of binomial and Poisson distributions. *Theory Probab. Appl.* **19** 584–587.

[18] Jaeschke, D. (1979). The asymptotic distribution of the supremum of the standardized empirical distribution function on subintervals. *Ann. Statist.* **7** 108–115. MR0515687

[19] Jager, L. and Wellner, J.A. (2007). Goodness-of-fit tests via phi-divergences. *Ann. Statist.* **35** 2018–2053. MR2363962

[20] Kaplan, D.M. and Goldman, M. (2014). Evenly sensitive KS-type inference on distributions. Preprint. Available at http://economics.missouri.edu/working-papers/2008.shtml.

[21] Kaplan, D.M. and Goldman, M. (2014). True equality (of pointwise sensitivity) at last: A Dirichlet alternative to Kolmogorov–Smirnov inference on distributions. Preprint. Avaliable at http://scholar.google.com/citations?user=c4yK72IAAAAJ&hl=en.

[22] Khmaladze, E. and Shinjikashvili, E. (2001). Calculation of noncrossing probabilities for Poisson processes and its corollaries. *Adv. in Appl. Probab.* **33** 702–716. MR1860097

[23] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. *Springer Series in Statistics*. New York: Springer. MR0691492

[24] Lockhart, R.A. (1991). Overweight tails are inefficient. *Ann*. *Statist*. **19** 2254–2258. MR1135175

[25] Mary, D. and Ferrari, A. (2014). A non-asymptotic standardization of binomial counts in higher criticism. In 2014 *IEEE International Symposium on Information Theory* (*ISIT*) 561–565. Piscataway, NJ: IEEE.

[26] Mason, D.M. and Schuenemeyer, J.H. (1983). A modified Kolmogorov–Smirnov test sensitive to tail alternatives. *Ann*. *Statist*. **11** 933–946. MR0707943

[27] Moskovich Eiger, A., Nadler, B. and Spiegelman, C. (2014). The calibrated Kolmogorov–Smirnov test. Preprint. Available at arXiv:1311.3190.

[28] Nikitin, Y. (1995). *Asymptotic Efficiency of Nonparametric Tests*. Cambridge: Cambridge Univ. Press. MR1335235

[29] Prohorov, Yu.V. (1953). Asymptotic behavior of the binomial distribution [Russian]. *Uspekhi Mat*. *Nauk* **8** 135–142. MR0056861

[30] Révész, P. (1982/83). A joint study of the Kolmogorov–Smirnov and the Eicker–Jaeschke statistics. *Statist*. *Decisions* **1** 57–65. MR0685588

[31] Shorack, G. and Wellner, J. (2009). *Empirical Processes with Applications to Statistics*. Philadelphia: Society for Industrial and Applied Mathematics.

[32] Smirnov, N.V. (1933). On probabilities of greater deviations [Russian]. *Mathem*. *Sb*. **40** 443–454.

Received February 2014 and revised October 2014