The intermediates take it all: Asymptotics of higher criticism statistics and a powerful alternative based on equal local levels

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, Universitätsstr. 1, 40225 Düsseldorf, Germany

[2] German Diabetes Center, Institute for Biometrics and Epidemiology, Auf'm Hennekamp 65, 40225 Düsseldorf, Germany

Received 7 November 2013; revised 11 April 2014; accepted 14 April 2014

The higher criticism (HC) statistic, which can be seen as a normalized version of the famous Kolmogorov–Smirnov statistic, has a long history, dating back to the mid seventies. Originally, HC statistics were used in connection with goodness of fit (GOF) tests but they recently gained some attention in the context of testing the global null hypothesis in high dimensional data. The continuing interest for HC seems to be inspired by a series of nice asymptotic properties related to this statistic. For example, unlike Kolmogorov–Smirnov tests, GOF tests based on the HC statistic are known to be asymptotically sensitive in the moderate tails, hence it is favorably applied for detecting the presence of signals in sparse mixture models. However, some questions around the asymptotic behavior of the HC statistic are still open. We focus on two of them, namely, why a specific intermediate range is crucial for GOF tests based on the HC statistic and why the convergence of the HC distribution to the limiting one is extremely slow. Moreover, the inconsistency in the asymptotic and finite behavior of the HC statistic prompts us to provide a new HC test that has better finite properties than the original HC test while showing the same asymptotics. This test is motivated by the asymptotic behavior of the so-called local levels related to the original HC test. By means of numerical calculations and simulations we show that the new HC test is typically more powerful than the original HC test in normal mixture models.

Keywords: Goodness of fit test; Higher criticism; Local levels; Normal and Poisson approximation; Order statistics.

1 Introduction and summary

Many important theoretical results related to the so-called *higher criticism* (HC) test statistic have been obtained during the past three decades, where it has typically been applied in the context of goodness of fit (GOF) and detection problems. The HC statistic can be seen as a normalized or standardized version of the well-known Kolmogorov–Smirnov test statistic. The asymptotics of the HC statistic was extensively investigated by Jaeschke (1979) and Eicker (1979) in the late 1970s. For earlier references see Anderson and Darling (1952). However, neither Jaeschke nor Eicker made use of the term *higher criticism* statistic. The notion of the *higher criticism* was first introduced by J. W. Tukey in the mid 1970s. Later, Tukey (1989) wrote:

> *"If we look at many comparisons, say n, and assess the significance of each at 5% individually [...]. We know that, even if the underlying value of each comparison is blah [...] we will get an average of $n/20$ (i.e. 5% of n) apparent significance.*

*Corresponding author: e-mail: veronika@gontscharuk.de, Phone: +49-211-3382-231, Fax: +49-211-3382-677

Various ways of relating the observed number, k, of individual-5% significances to $n/20$ are mnemonically referred to as "the higher criticism" [...].

It was not until a decade ago that Donoho and Jin (2004) termed the normalized Kolmogorov–Smirnov statistic as *Tukey's higher criticism* and rediscovered the HC in the context of detecting signals that are both sparse and weak.

Among others, Donoho and Jin (2004) showed the optimality of the HC statistic in the sense that a test based on the HC statistic asymptotically mimics the performance of an oracle likelihood ratio test under several conditions. A substantial contribution to results in Donoho and Jin (2004) was made by Ingster (1997, 1999). Jager and Wellner (2007) provided a family of GOF test statistics based on ϕ-divergences that have the same optimal detection boundary in sparse normal mixtures as the HC statistic. Hall and Jin (2008) focused on HC in the case of dependent data. Later, Hall and Jin (2010) modified the standard HC statistic used by Donoho and Jin (2004) to account for correlated noise.

Instead of approaching HC tests in terms of their test statistics a viewpoint from so-called local levels was introduced in Gontscharuk et al. (2013). There has been a lot of further interest and developments in connection with this statistic, e.g. cf. Cai et al. (2007), Donoho and Jin (2008), Hall et al. (2008), Donoho and Jin (2009), and Cai et al. (2011).

Practical applications of the HC statistic can be found in several areas. Often, these applications result from questions arising in the context of large-scale multiple testing. In particular, scientific areas such as genomics, astronomy, or image processing, have seen a growing need for statistical tools to analyze high-dimensional data. In these areas the aim is often to identify whether there are signals present in the data. For example, Parkhomenko et al. (2009) employed the HC test to detect the presence of small effects in a genome-wide association study (GWAS) on rheumatoid arthritis. In Sabatti et al. (2009), the authors analyzed GWAS data to determine genetic influences on certain metabolic traits and tested the global null hypothesis of no genetic effect using the HC statistic. Besides these applications the HC is generally applicable in any areas in which there emerges an interest in GOF testing. Here, a number of questions and issues in the context of HC goodness of fit tests remain, three of which will be addressed in this paper.

1. It is well known that a GOF test based on the HC statistic is asymptotically sensitive for some special kinds of alternatives that differ from the null distribution in the moderate tails. However, it is not clear how to explain this behavior, since the proofs of related results are typically of pure technical nature. Based on the theory of stochastic processes we will show why a specific intermediate range plays a crucial role.
2. It is known that the convergence of the distribution of the HC statistic to the limiting distribution is extremely slow so that the application of asymptotic results may lead to doubtful outcomes for a finite sample size. Results based on simulations, cf. Donoho and Jin (2004) or Hall and Jin (2010), show that this irregular behavior is frequently caused by the smallest order statistics of the underlying sample. However, there seems no theoretical result justifying this observation. In this paper, we will provide a simple condition how to check whether the asymptotic HC distribution approximates the finite one quite well for a given sample size.
3. Due to a huge discrepancy between the HC's finite and asymptotic behavior, it is desirable to construct a new level α test that has the same asymptotic properties as the original HC test but shows an improved finite sample behavior. It was at the MCP2011 Conference where we introduced such a test for the first time. Eventually, at the MCP2013, we presented asymptotic as well as finite properties and power considerations of this new HC test, which will be addressed as a final topic of this paper.

This paper is organized as follows. In Section 2, we provide the basic notation and necessary background and illustrate the key issues in detail. Section 3 is devoted to the first problem. We introduce continuous stochastic processes that are related to the HC statistic. More precisely, we consider the normalized Brownian bridge and its approximation property in the region where the HC statistic is particularly sensitive. Further, we introduce the Ornstein–Uhlenbeck process that results from a suitable time transformation of the normalized Brownian bridge and helps us in the investigation of the phenomena that appear in the asymptotics of HC statistics. Section 4 addresses the second key issue of this paper, that is a discussion about the quality of the HC asymptotics applied in the finite sample size case. Thereby, we study the performance of various truncated versions of the HC statistic. An essential observation is that the left tail in form of the smallest order statistics involved is key in contributing to the distribution of the HC statistic. In Section 5, we refer to the third key aspect. We construct a new HC test by considering so-called *local levels* and by setting these quantities to be all equal. We compare the new and original HC tests under the null hypothesis as well as under alternatives. Concluding remarks are given in Section 6.

2 Background and notation

Let $U_{1:n}, \ldots, U_{n:n}$ be the order statistics of n independent and identically distributed (iid) random variables. Typically, HC statistics can be expressed in terms of the order statistics $U_{i:n}$, $i = 1, \ldots, n$, from the viewpoint of the abscissa or in terms of the corresponding empirical cumulative distribution function (ecdf) $\hat{F}_n$ from the ordinate viewpoint. Below we restrict our attention to the standardized versions of the HC test statistic, that is,

$$\mathrm{HC}_{0,1} = \sup_{0<t<1} \sqrt{n}\frac{\hat{F}_n(t) - t}{\sqrt{t(1-t)}} \tag{1}$$

$$\overset{a.s.}{=} \max_{1\le i\le n} \sqrt{n}\frac{i/n - U_{i:n}}{\sqrt{U_{i:n}(1-U_{i:n})}}. \tag{2}$$

Depending on the research question posed, either definition will be considered. Under the assumption that the given order statistics $U_{i:n}$, $i = 1, \ldots, n$, stem from the uniform distribution on the interval [0, 1], the limit distribution of $\mathrm{HC}_{0,1}$ is given by the Gumbel distribution in the sense that

$$\lim_{n\to\infty} \mathrm{P}(\mathrm{HC}_{0,1} \le b_n(x)) = \exp(-\exp(-x))\,, \tag{3}$$

where $x \in \mathbb{R}$,

$$b_n(x) = \sqrt{2\log_2(n)} + (\log_3(n) - \log(\pi) + 2x)/(2\sqrt{2\log_2(n)}), \tag{4}$$

$\log_2(n) = \log(\log(n))$ and $\log_3(n) = \log(\log(\log(n)))$, cf. Jaeschke (1979) and Eicker (1979). Then we can define a GOF test based on the HC test statistic, which we will call the *HC test*, for testing the null hypothesis H_0 that the underlying sample is, in fact, a realization of iid standard uniform distributed random variables. We say that the HC test rejects H_0 if the test statistic $\mathrm{HC}_{0,1}$ is larger than the (asymptotic) critical value $b_n(x)$. Setting $x \equiv x_\alpha = -\log_2(1/(1-\alpha))$, the corresponding HC test is an asymptotic level α test, that is

$$\lim_{n\to\infty} \mathrm{P}(\mathrm{HC}_{0,1} \le b_n(x_\alpha)) = 1 - \alpha.$$

It is known that such HC tests are typically more powerful than the classical (asymptotic level α) Kolmogorov–Smirnov tests if an alternative distribution deviates from the null distribution in moderate tails. This is due to the fact that under H_0 the supremum and/or maximum in (1) and/or (2), respectively, is asymptotically taken only over a specific intermediate range. More precisely, let us consider a truncated version of the HC statistic defined by

$$\mathrm{HC}_{d_n,e_n} = \sup_{d_n<t<e_n} \sqrt{n}\,\frac{\hat{F}_n(t) - t}{\sqrt{t(1-t)}}, \quad 0 \le d_n < e_n \le 1. \tag{5}$$

Thereby, we say (d_n, e_n) lies in the central range of $(0, 1)$ if there exist $d > 0$ and $e < 1$ such that $d < d_n < e_n < e$, $n \in \mathbb{N}$. If $e_n \to 0$ or $d_n \to 1$ as $n \to \infty$, we say (d_n, e_n) lies in the left or right tail, respectively. Moreover, (d_n, e_n) belongs to the intermediate range if either $e_n \to 0$ and $nd_n \to \infty$ or $d_n \to 1$ and $n(1 - e_n) \to \infty$ as $n \to \infty$. Note that the intermediate range is a part of the respective tail.

For what follows, we assume that H_0 is true, that is, $U_1, \dots, U_n$ are iid uniformly distributed on [0, 1], if nothing else is mentioned. Jaeschke (1979) and Eicker (1979) showed that the distribution of HC_{d_n,e_n} is asymptotically degenerated in the sense that

$$\lim_{n\to\infty} \mathrm{P}\big(\mathrm{HC}_{d_n,e_n} \le b_n(x)\big) = 1,$$

whenever d_n and e_n meet one of the following conditions:

(i) $d_n \to d \in (0, 1)$ and $e_n \to e \in (0, 1)$ as $n \to \infty$,
(ii) $d_n \to 0$ and $e_n \to 1$ too slowly as $n \to \infty$,
(iii) $d_n \equiv 0$ and $e_n = n^{s_n - 1}$ for a $0 < s_n = o(1)$,
(iv) $d_n = 1 - n^{s_n - 1}$ for a $0 < s_n = o(1)$ and $e_n \equiv 1$.

For example, it follows that the maximum taken over the central range or the maximum over extreme tails is asymptotically stochastically smaller than the asymptotic critical value $b_n(x)$ for any $x \in \mathbb{R}$. Furthermore, applying results in Jaeschke (1979), we even get that the maximum taken over a specific intermediate range is asymptotically (stochastically) equal to the maximum taken over the entire interval $(0, 1)$, that is

$$\lim_{n\to\infty} \mathrm{P}\big(\max\big\{\mathrm{HC}_{d_n,e_n}, \mathrm{HC}_{1-e_n,1-d_n}\big\} \le b_n(x)\big) = \exp(-\exp(-x)) \tag{6}$$

for $d_n = n^{s_n - 1}$ and $e_n = n^{-r_n}$ with $0 < s_n = o(1)$ and $0 < r_n = o(1)$. This is why we say that intermediates take it all. In view of (6) we denote

$$R_n(s_n, r_n) = (n^{s_n - 1}, n^{-r_n}) \cup (1 - n^{-r_n}, 1 - n^{s_n - 1}) \tag{7}$$

with $0 < s_n = o(1)$ and $0 < r_n = o(1)$ as a sensitivity range related to the HC statistic. Roughly speaking, the sensitivity range is given by

$$R_n \equiv (n^{o(1)-1}, n^{-o(1)}) \cup (1 - n^{-o(1)}, 1 - n^{o(1)-1}).$$

Surprisingly, the sensitivity range R_n is very small compared to the entire interval $(0, 1)$ but crucial for the HC asymptotics. Unfortunately, proofs related to HC results are typically of technical nature so that it is not clear why the supremum taken over R_n is asymptotically stochastically larger than the supremum taken over the remaining area. In Section 3, we will give an explanation for this phenomenon by considering the HC statistic as the maximum of a stochastic process.

Another interesting issue that is addressed in this paper is the behavior of the HC statistic in the finite case. First, we note that available formulas for the cumulative distribution function (cdf) of the HC statistic $F_n(y)$ (say) for a given $n \in \mathbb{N}$ are not easy to handle analytically. One way out would be the numeric calculation or simulation of $F_n(y)$ for a given n. However, such computations take much time even for $n \approx 10^5$. On the other hand, due to (3) we can approximate $F_n(y)$ by the (asymptotic) cdf

$$F_n^{\infty}(y) = \exp\big(-\exp\big(-b_n^{-1}(y)\big)\big),$$

where b_n^{-1} is the inverse function of b_n, that is

$$b_n^{-1}(y) = y\sqrt{2\log_2(n)} - 2\log_2(n) - (\log_3(n) - \log(\pi))/2.$$

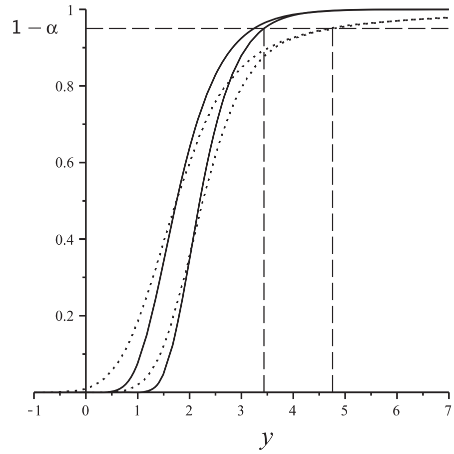

Figure 1 The cdf of the HC statistic $\mathrm{HC}_{0,1}$ (dotted curves) simulated by 10^5 repetitions, that is $F_n(y)$, and its respective asymptotic version $F_n^{\infty}(y)$ (solid curves) for $n = 10^2$ (left curves) and $n = 10^4$ (right curves). The vertical dashed lines locate the $(1-\alpha)$-quantiles of F_n^{∞} and F_n for $n = 10^4$ and $\alpha = 0.05$.

Unfortunately, it seems that the asymptotic distribution does not yield a good approximation even for larger n-values. For example, Jaeschke (1979) already noted that he would not recommend to use the confidence intervals that are computed on the basis of the asymptotic theory. Also Khmaladze and Shinjikashvili (2001) showed that the exact (finite) distributions of the HC statistic are quite far from the limiting distribution. Even worse, the authors note that for sample sizes up to $n = 10^4$ the exact distributions even diverge from the limiting one. The discrepancy between the finite and asymptotic behavior becomes clear in Fig. 1, where the (simulated) cdf of the HC statistic F_n and its asymptotic law F_n^{∞} are shown for various n-values.

As compared to the critical values $b_n(x)$ of the limiting distribution, it is apparent that larger quantiles of the finite $\mathrm{HC}_{0,1}$ statistic, which are typically relevant for testing purposes, are too large even for large sample sizes. As an example, we consider the 0.95-quantile of the limiting distribution, that is $b_n(x_\alpha)$ with $x_\alpha = -\log_2(1/(1-\alpha))$ and $\alpha = 0.05$. For $n = 10^4$ the asymptotic critical value $b_n(x_\alpha) = 3.434$ should approximate the 0.95-quantile of the exact $\mathrm{HC}_{0,1}$-distribution; however, this quantity is equal to the 0.876-quantile. The 0.95-quantile of the exact distribution turns out to be approximately 4.74, showing the discrepancy to the asymptotic value, which is also visible in Fig. 1. Therefore, using the critical values of the asymptotic distribution in the context of testing may lead to a considerable exceedance of the prechosen level α.

Moreover, it is known that for a finite sample size unusual large values of $\mathrm{HC}_{0,1}$ are most frequently caused by the smallest order statistic $U_{1:n}$. This is why truncated HC versions were considered in the literature. For example, Donoho and Jin (2004) proposed to restrict the range of the maximum by applying

$$\mathrm{HC}_n^{+} = \max_{1<i\leq n/2,\, U_{i:n}\geq 1/n} \sqrt{n}\frac{i/n - U_{i:n}}{\sqrt{U_{i:n}(1-U_{i:n})}}$$

and Hall and Jin (2010) considered

$$\mathrm{HC}_n^{*} = \max_{i:1/n<U_{i:n}<1/2} \sqrt{n}\frac{i/n - U_{i:n}}{\sqrt{U_{i:n}(1-U_{i:n})}}.$$

Of course, HC_n^+ and HC_n^* are not larger than the original HC statistic so that the distributions of these truncated versions are closer to the limiting distribution. Consequently, the asymptotic critical value $b_n(x)$ approximates the corresponding quantiles of HC_n^+ and HC_n^* better than the same quantile related to the $\mathrm{HC}_{0,1}$. On the other hand, to apply a truncated HC statistic in the context of GOF tests, that is, to exclude the smallest values of the underlying sample, which typically can be seen as an indicator that the null hypothesis is false, seems to be too wasteful from a statistical point of view. This is why, in the finite case, we restrict our attention to the original HC statistic $\mathrm{HC}_{0,1}$. Thereby, the second focus of this paper lies on addressing the question why we observe a rather poor agreement of the asymptotic and finite distributions of the $\mathrm{HC}_{0,1}$ statistic even in a large sample size case.

The final aspect we focus on is a modification of the HC test that seems to be essential in view of the previous topic of this paper. In Section 5, we derive the new (better) HC test motivated by results in Gontscharuk et al. (2013) and show that this test is typically more powerful than the original HC test in a normal mixture model with rather sparse signals as studied by, for example, Donoho and Jin (2004).

3 Why do intermediates take it all?

This section is intended to provide an answer to the first issue raised in Section 1 by studying the behavior of the supremum of truncated versions of the HC statistic and related approximations in various ranges.

3.1 Empirical HC process and its approximations

HC statistics defined in (5) can be seen as a supremum of the normalized (uniform) empirical process $\mathbb{Z}_n = \mathbb{U}_n/\sqrt{I(1-I)}$ over the interval $(d_n, e_n) \subseteq (0, 1)$, where $\mathbb{U}_n = \sqrt{n}(\hat{F}_n - I)$ is a natural normalization of the empirical cdf $\hat{F}_n$, and I is the identity function on [0, 1]. Below, we denote $\mathbb{Z}_n(t)$ as the *empirical HC process*. By definition,

$$\mathrm{HC}_{d_n,e_n} = \sup_{d_n<t<e_n} \mathbb{Z}_n(t).$$

It is well known that the empirical process $\mathbb{U}_n(t)$ can be approximated by a Brownian bridge as well as by a Poisson process. The Brownian bridge is a suitable approximation in a central range, while the Poisson approximation can be more appropriate in the tails. Both approximations lead to interesting results concerning the asymptotics of the HC statistic.

In order to separate regions with different approximations, for a suitable $d_n \in (0, 0.5)$ we chose the area $(0, d_n) \cup (1 - d_n, 1)$ for the Poisson approximation and the interval $(d_n, 1 - d_n)$ for the normal approximation. Note that there are infinitely many sequences d_n, $n \in \mathbb{N}$, that lead to the same results concerning the asymptotics of the HC statistic. For technical reasons, we choose

$$d_n = \frac{\log^5(n)}{n}, \quad n \in \mathbb{N}.$$

Thereby, $d_n < 0.5$ is fulfilled for $n \geq 1010389$.

First, we consider the HC process evaluated in the range $(0, d_n) \cup (1 - d_n, 1)$. Let $\nu(t)$, $t > 0$, denote a Poisson process with parameter $\lambda = 1$ and paths constant except for upward jumps of height one.

Lemma 3 in Jaeschke (1979) provides that for any $k > 0$ and $x \in \mathbb{R}$ the corresponding normalized Poisson process $(\nu(t) - t)/\sqrt{t}$ fulfills

$$\lim_{n\to\infty} \mathrm{P}\left(\sup_{0<t<\log^k(n)} \frac{\nu(t)-t}{\sqrt{t}} \leq b_n(x)\right) = 1.$$

Moreover, following the proof of Lemma 4 in Jaeschke (1979), we get

$$\lim_{n\to\infty} \mathrm{P}\left(\sup_{t\in(0,d_n)\cup(1-d_n,1)} \mathbb{Z}_n(t) \leq b_n(x)\right) = 1 \tag{8}$$

so that values of the HC process taken in $(0, d_n) \cup (1 - d_n, 1)$ are asymptotically negligible with respect to their contribution to the distribution of the original HC statistic. Hence, it suffices to study the HC process and its suprema taken over subsets of the interval $(d_n, 1 - d_n)$.

Denoting a *Brownian bridge* on [0, 1] by $\mathbb{U}$, a continuous process related to the HC statistic is given by

$$\mathbb{Z}(t) = \frac{\mathbb{U}(t)}{\sqrt{t(1-t)}}, \quad 0 < t < 1.$$

Clearly, $\mathbb{Z}_n(t)$ converges in distribution to $\mathbb{Z}(t)$ for any $t \in (0, 1)$. Moreover, it holds

$$\sup_{d_n<t<1-d_n} |\mathbb{Z}_n(t) - \mathbb{Z}(t)| \overset{a.s.}{=} o(1/\sqrt{2\log_2(n)}), \tag{9}$$

cf. Shorack and Wellner (2009), p. 601. This allows us to work with the continuous process $\mathbb{Z}(t)$ instead of the discrete analog $\mathbb{Z}_n(t)$ on any subinterval of $(d_n, 1 - d_n)$. Therefore, we denote the normalized Brownian bridge $\mathbb{Z}(t)$ as the *continuous HC process*. In conclusion, we summarize the observations to get the following result.

Result 1. *The cdf of the original HC statistic $HC_{0,1}$ can be approximated with any accuracy (for n large enough) by the cdf of the supremum of the continuous HC process $\mathbb{Z}(t)$, that is by the cdf of the supremum of the normalized Brownian bridge, taken over $(d_n, 1 - d_n)$ with $d_n = \log^5(n)/n$.*

3.2 Relation to Gaussian stationary processes

First, we summarize results from Shorack and Wellner (2009) related to the continuous HC process $\mathbb{Z}(t)$ that show that $\mathbb{Z}(t)$ can be seen as a transformation of a Gaussian stationary process. Let $\mathbb{S}$ denote a *Brownian motion* on $[0, \infty)$ and

$$\mathbb{X}(r) = e^{-r}\,\mathbb{S}(e^{2r}) \quad \text{for all} \quad r \in \mathbb{R}.$$

The process $\mathbb{X}$ is known as the *Ornstein–Uhlenbeck* (OU) process. The OU process $\mathbb{X}$ is a stationary zero-mean Gaussian process with unit variance and

$$\mathrm{Cov}[\mathbb{X}(s), \mathbb{X}(r)] = \exp(-|s-r|) \quad \text{for} \quad s, r \in \mathbb{R}. \tag{10}$$

Hence, the dependence between $\mathbb{X}(s)$ and $\mathbb{X}(r)$ for $s, r \in \mathbb{R}$ is determined by their distance $|s - r|$ only. Then Doob's transformation leads to

$$\mathbb{Z}(t) \overset{\mathcal{D}}{=} \sqrt{\frac{1-t}{t}}\,\mathbb{S}\left(\frac{t}{1-t}\right) = \mathbb{X}\left(\frac{1}{2}\log\left(\frac{t}{1-t}\right)\right), \quad 0 < t < 1, \tag{11}$$

where $\overset{\mathcal{D}}{=}$ denotes equality in distribution, cf. Shorack and Wellner (2009, p. 20 and p. 598). Moreover, in view of the stationarity of $\mathbb{X}$, we obtain

$$\sup_{d<t<e} \mathbb{Z}(t) \overset{\mathcal{D}}{=} \sup_{1<t<a} \frac{\mathbb{S}(t)}{\sqrt{t}} = \sup_{0<t<\gamma} \mathbb{X}(t) \tag{12}$$

with

$$a \equiv a(d,e) = \frac{e(1-d)}{d(1-e)} \quad \text{and} \quad \gamma \equiv \gamma(a) = \frac{1}{2}\log(a). \tag{13}$$

Thereby, according to (11) an interval (d, e) related to the continuous HC process $\mathbb{Z}$ leads to the interval $(\log(d/(d-1))/2, \log(e/(e-1))/2)$ related to the OU process $\mathbb{X}$. Noting that $\mathbb{X}$ is stationary, the supremum of $\mathbb{X}$ over the latter interval is equal in distribution to the supremum over any interval of the same length, which is $\gamma = \log(e(1-d)/(d(1-e)))/2$. Hence, we get $\sup_{d<t<e} \mathbb{Z}(t) \overset{\mathcal{D}}{=} \sup_{0<t<\gamma} \mathbb{X}(t)$. Moreover, due to (11) an interval $(0, \gamma)$ related to $\mathbb{X}$ immediately leads to the interval $(1, \exp(2\gamma))$ related to the Brownian motion $\mathbb{S}$. Finally, we set $a \equiv \exp(2\gamma)$.

In view of (12) and (13) we obtain the following result.

Result 2. *The supremum of the continuous HC process* $\mathbb{Z}(t)$ *over an interval* $(d, e) \subseteq (0, 1)$ *is equal in distribution to the supremum of the OU process* $\mathbb{X}(t)$ *over* $(0, \gamma)$ *with* $\gamma \equiv \gamma(a(d, e))$ *defined in* (13).

This observation implies, for example, that suprema of $\mathbb{Z}(t)$ over intervals, which are symmetric about $t = 0.5$, are equal in distribution, that is, for $0 < d < e < 1$ we get

$$\sup_{d<t<e} \mathbb{Z}(t) \overset{\mathcal{D}}{=} \sup_{1-e<t<1-d} \mathbb{Z}(t).$$

Furthermore, for any intervals (d, e) and (d', e') satisfying $d, d' > 0$ and $e, e' < 1$, we obtain

$$\sup_{d<t<e} \mathbb{Z}(t) \overset{\mathcal{D}}{=} \sup_{d'<t<e'} \mathbb{Z}(t) \quad \text{if and only if} \quad \gamma(a(d,e)) = \gamma(a(d',e')),$$

with $\gamma(a(d, e))$ defined in (13). For example, Fig. 2 shows values of $e \equiv e(d)$ (left picture) and the length of intervals (d, e), that is $e - d$, (right picture) fulfilling $\gamma(a(d, e)) \equiv \gamma$ for all $d \in (0, 1)$ and $\gamma = 0.5, 1, 2$. Note that for a fixed γ the distribution of the supremum of $\mathbb{Z}(t)$ over intervals $(d, e(d))$ is the same for all values of d.

Surprisingly, Fig. 2 illustrates that the supremum of $\mathbb{Z}(t)$ over small intervals that are close to zero, can be stochastically equal to and even larger than the supremum over much larger intervals, which are further away from zero. In order to reach a deeper understanding of the nature of this phenomenon we will study the OU process in more detail.

3.3 Distribution of the supremum of the OU process

First, we turn to the OU process on a fixed interval. An exact formula for the cdf of the supremum of the OU process $\mathbb{X}(t)$ on $(0, \gamma)$ with $\gamma > 0$ is given by

$$F_\gamma^{OU}(y) = \mathrm{P}\left(\sup_{0<t<\gamma} \mathbb{X}(t) \le y\right) = -\varphi(y) \sum_{i=1}^{\infty} \frac{D_{\nu_i - 1}(-y)}{\nu_i D'_{\nu_i}(-y)} \exp(-\nu_i \gamma), \tag{14}$$

where $D_\nu(\cdot)$ is a parabolic cylinder function, ν_i is the ith root of the equation $D_\nu(-y) = 0$ as a function of ν, $D'_{\nu_i}(-y)$ is the derivative with respect to ν evaluated at ν_i and $\varphi(y)$ is the standard normal density function, cf. Estrella and Rodrigues (2005) and De Long (1981). To the best of our knowledge the cdf

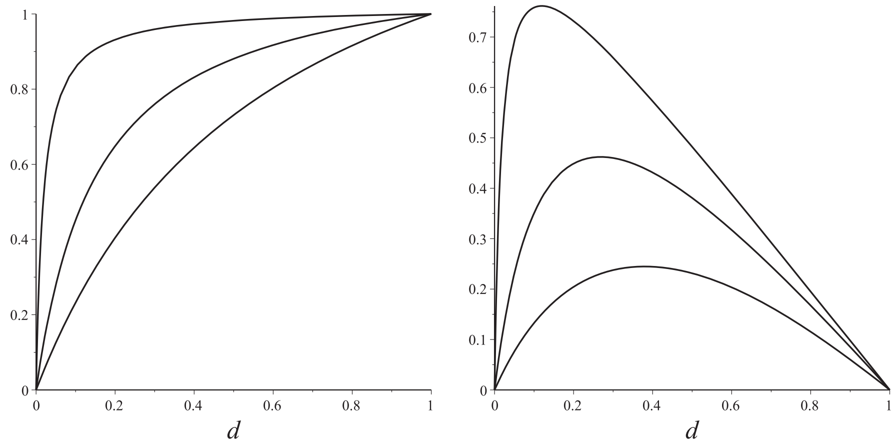

Figure 2 Values of the upper boundary e (left graph) and values of the interval's length $e - d$ as a function of d such that the distribution of $\sup_{d<t<e} \mathbb{Z}(t)$ is the same for all intervals (d, e) fulfilling $\gamma(a(d, e)) = \gamma$ (right graph). Here, $\gamma = 0.5, 1, 2$ (from the bottom to the top).

of the supremum of the OU process over a fixed interval and, particularly, the formula (14) have not been considered in connection with the HC statistic until now. At first glance, Results 1 and 2 offer a ray of hope that for a suitable $\gamma > 0$ the expression in (14) leads to a better approximation of the cdf of $\mathrm{HC}_{0,1}$ in the finite case than the well known Gumbel-related cdf F_n^∞. Although the expression given in (14) seems to be hard to handle analytically, the cdf $F_\gamma^{OU}(y)$ can be calculated numerically as precisely as one may wish. All in all, F_γ^{OU} offers an alternative approximation for the finite (exact) distribution of the original HC statistic.

The limiting behavior of the supremum of $\mathbb{X}(t)$ over an increasing interval is given by the Gumbel distribution in the sense that

$$\lim_{\gamma\to\infty} \mathrm{P}\left(\sup_{0<t<\gamma} \mathbb{X}(t) \le \tilde{b}(x, \gamma)\right) = \exp(-\exp(-x)), \tag{15}$$

where

$$\tilde{b}(x, \gamma) = \sqrt{2\log(\gamma)} + (\log_2(\gamma) - \log(\pi) + 2x)/(2\sqrt{2\log(\gamma)}), \tag{16}$$

cf. Theorem 12.3.5 in Leadbetter et al. (1989, p. 237). Setting $\gamma = \log(n)$, $n \in \mathbb{N}$, we obtain

$$\tilde{b}(x, \log(n)) = b_n(x), \quad x \in \mathbb{R},$$

where $b_n(x)$ is an asymptotic critical value of the HC test, cf. (4). Then the asymptotic distribution of the supremum of the OU process on the interval $(0, \log(n))$ is the same as the asymptotic distribution of $\mathrm{HC}_{0,1}$, hence we have

$$F_{\log(n)}^{OU}(y) \approx F_n^\infty(y) \ \text{ for any } \ y \in \mathbb{R} \ \text{ and large } \ n \in \mathbb{N}. \tag{17}$$

As a consequence, $F_{\log(n)}^{OU}(y)$ is a candidate for a better approximation of the cdf $F_n(y)$ in a finite case. For example, Fig. 3 shows the asymptotic and finite cdf's of the supremum of $\mathbb{X}(t)$ over $(0, \log(n))$ for $n = 10^2, 10^8, 10^{32}$. Unfortunately, the cdf $F_{\log(n)}^{OU}$ is an even worse approximation of the cdf of $\mathrm{HC}_{0,1}$ than the approximation by the limit cdf F_n^∞ in the sense that critical values, which can be calculated via $F_{\log(n)}^{OU}(y)$, are typically smaller than the corresponding critical values related to F_n^∞, that is $b_n(x)$, and

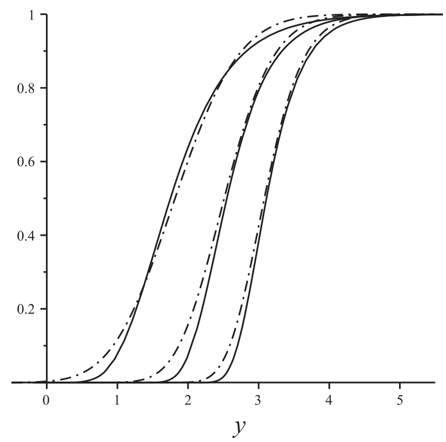

Figure 3 Asymptotic cdf $F_n^{\infty}(y)$ (solid curves) and finite cdf $F_{\log(n)}^{OU}(y)$ (dash-dotted curves) of the supremum of the OU process on the interval $(0, \log(n))$ for $n = 10^2, 10^8, 10^{32}$ (from left to right).

hence much smaller than the test statistic $\mathrm{HC}_{0,1}$. Consequently, the search for a suitable approximation of the finite cdf of $\mathrm{HC}_{0,1}$ has to be continued.

Note that due to Results 1 and 2 the original HC statistic is also asymptotically equal in distribution to the supremum of the OU process over the interval $(0, \gamma_n)$ with $\gamma_n = \gamma(a(d_n, 1 - d_n))$ and $d_n = \log(n)^5/n$. That is, for such γ_n, which particularly fulfills $\gamma_n = \log(n) + o(\log(n))$, we get

$$F_{\gamma_n}^{OU}(y) \approx F_n^{\infty}(y) \quad \text{for any} \quad y \in \mathbb{R} \quad \text{and large} \quad n \in \mathbb{N}. \tag{18}$$

Expressions (17) and (18) indicate that values of the OU process $\mathbb{X}(t)$ over a smaller interval may not contribute much to the distribution of the supremum of $\mathbb{X}(t)$ over distinctly larger intervals. Indeed, the next lemma provides the corresponding result.

Lemma 3.1. *Let $0 < \gamma_n' < \gamma_n$ be such that $\gamma_n' \to \infty$ as $n \to \infty$ and $\epsilon_n \equiv \gamma_n - \gamma_n'$ fulfill $\epsilon_n = o(\gamma_n)$. Then for any $x \in \mathbb{R}$ we get*

$$\lim_{n\to\infty} F_{\gamma_n}^{OU}(\tilde{b}(x, \gamma_n)) = \lim_{n\to\infty} F_{\gamma_n'}^{OU}(\tilde{b}(x, \gamma_n)). \tag{19}$$

Proof. Note that (19) can be written as

$$\lim_{n\to\infty} \mathrm{P}\left(\sup_{0<t<\gamma_n} \mathbb{X}(t) \le \tilde{b}(x, \gamma_n)\right) = \lim_{n\to\infty} \mathrm{P}\left(\sup_{0<t<\gamma_n'} \mathbb{X}(t) \le \tilde{b}(x, \gamma_n)\right).$$

In view of (15)–(16), it suffices to show that $\tilde{b}(x, \gamma_n) = \tilde{b}(x, \gamma_n') + o(1/\sqrt{\log(\gamma_n)})$, which implies $\tilde{b}(x, \gamma_n) = \tilde{b}(x + o(1), \gamma_n')$. Applying Taylor series expansions, we get

$$\begin{aligned}\sqrt{\log(\gamma_n)} &= \sqrt{\log(\gamma_n'(1 + \epsilon_n/\gamma_n'))}\\ &= \sqrt{\log \gamma_n'} \cdot \sqrt{1 + o(1/\log(\gamma_n'))}\\ &= \sqrt{\log \gamma_n'} + o(1/\sqrt{\log(\gamma_n')}),\end{aligned}$$

which completes the proof. □

Hence, we get the following result.

Result 3. *Distributions of the supremum of the OU process over increasing intervals with lengths* γ_n *and* γ_n'*, respectively, are asymptotically equal if* $\gamma_n/\gamma_n' \to 1$ *as* $n \to \infty$.

3.4 The supremum of the HC process over the intermediate range

Thanks to Results 1–3 we are now able to give an explanation why the intermediate range R_n plays a key role for the original HC statistic $\mathrm{HC}_{0,1}$. Results 1–3 show that $\mathrm{HC}_{0,1}$ can be approximated as well as one may wish (for n large enough) by the supremum of the maximum of the OU process over any interval of the length $\log(n) + o(\log(n))$. That is, any interval $(d_n, e_n) \subset (0, 1)$ would be crucial for the statistic $\mathrm{HC}_{0,1}$ in the sense that Eq. (6) is fulfilled for this interval (d_n, e_n), if the length of the corresponding OU-interval $\gamma_n \equiv \gamma(a(d_n, e_n))$ is approximately $\log(n)$, that is $\gamma_n = \log(n) + o(\log(n))$. Since the sensitivity range R_n defined in (7) corresponds to the OU-interval $(0, \gamma(R_n))$ with the length

$$\gamma(R_n) = \log(n)[1 + o(1)],$$

it becomes clear why the intermediates asymptotically take it all. It remains to take a look at the region between the two parts of $R_n(s_n, r_n)$, that is, $(n^{-r_n}, 1 - n^{-r_n})$ with a suitable $r_n = o(1)$. Although the length of this interval tends to one, the length of the corresponding OU-interval is

$$\gamma_n(a(n^{-r_n}, 1 - n^{-r_n})) = r_n \log(n) + O(n^{-r_n}) = o(\log(n)).$$

For this, Lemma 3.1 yields that values of the continuous (and hence empirical) HC process on $(n^{-r_n}, 1 - n^{-r_n})$ asymptotically have no influence on the supremum on the entire interval $(0, 1)$.

3.5 Asymptotic dependence structure of the HC process

In order to round out our study of the HC sensitivity range, we investigate the asymptotic dependence structure of the HC process. We show that the HC process in the central region of the interval (0, 1) is more dependent than in the sensitivity range R_n.

Let $f : [0, 1] \to [-\infty, \infty]$ be a function related to Doob's transformation, that is,

$$f(t) = \log(t/(1-t))/2.$$

Applying (11), we arrive at

$$\mathbb{Z}(t) \overset{\mathcal{D}}{=} \mathbb{X}(f(t)), \quad 0 < t < 1.$$

Then the dependence structure of the continuous HC process $\mathbb{Z}(t)$ can be described by the covariance between two points of this process as follows:

$$\mathrm{Cov}(\mathbb{Z}(t), \mathbb{Z}(t+\delta)) = \exp(f(t) - f(t+\delta)), \quad 0 < t < t + \delta < 1. \tag{20}$$

Remember that the empirical and continuous HC processes asymptotically coincide on the interval $(d_n, 1 - d_n)$ for at least $d_n = \log(n)^5/n$ so that in this interval also the dependence structure of the continuous process asymptotically coincides with the dependence structure of the corresponding empirical process.

To illustrate the dependence visually, we consider the covariance structure of $\mathbb{Z}(t)$ evaluated in equidistant points $t_i \equiv i/(n+1)$, $i = 1, \ldots, n$. This choice can be motivated by noting that under the null hypothesis H_0 the expectation of the ith smallest order statistic is $i/(n+1)$.

Figure 4 shows the transformation function $f(t)$ together with points $(t_i, f(t_i))$, $i = 1, \ldots, n$, $n = 100$, on this curve, which are equidistant on the horizontal axis. Note that $f(t)$ maps intervals lying

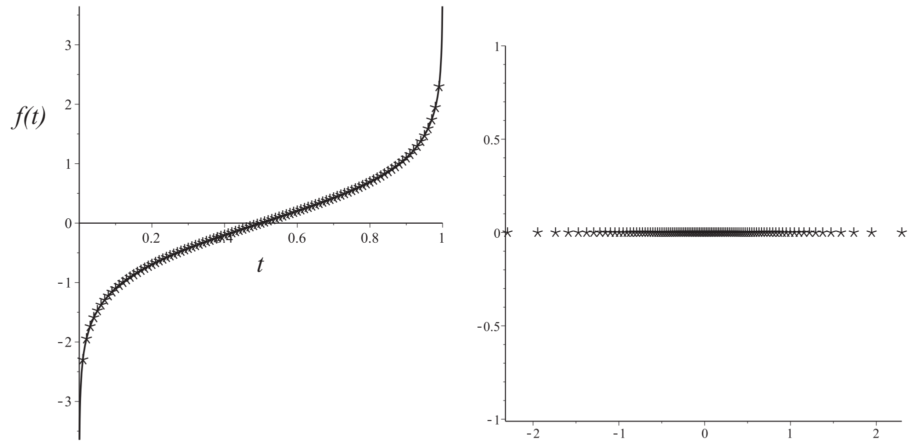

Figure 4 Left graph: The function $f(t)$ (solid curve), for which the processes $\mathbb{Z}(t)$ and $\mathbb{X}(f(t))$ are equal in distribution for all $t \in (0, 1)$, together with points $(t_i, f(t_i))$, $t_i = i/(n+1)$, $i = 1, \ldots, n$ for $n = 100$ (asterisks). Right graph: Points $(t_i, f(t_i))$, $t_i = i/(n+1)$, $i = 1, \ldots, n$ for $n = 100$.

in any $[d, e]$ with $0 < d < e < 1$ (i.e. intervals lying in the central range) to intervals with bounded lengths, while intervals lying in the tails are mapped to increasing ones. For example, $(0.1, 0.9)$ is mapped to $(-1.0986, 1.0986)$, that is, any subinterval in $(0.1, 0.9)$ is mapped to an interval with length not larger than 2.197. On the other hand, $(1/n, 0.1)$ with the length smaller than 0.1 is mapped to $(-\log(n-1)/2, 1.0986)$, which has a length larger than $\log(n-1)/2$.

What is more, Fig. 4 shows that the distance on the vertical axis (i.e. the distance on the OU-scale) between two adjacent points in the central range is considerably smaller than the corresponding distance in the tails. Therefore, (20) implies that for two points t and $t+\delta$, which lie in the central range of the interval $(0, 1)$, the corresponding random variables $\mathbb{Z}(t)$ and $\mathbb{Z}(t+\delta)$ are more dependent than for t and $t+\delta$ lying in the tails.

More precisely,

$$\mathrm{Cov}(\mathbb{Z}(t), \mathbb{Z}(t+\delta)) = \sqrt{\frac{1/(t+\delta)-1}{1/t-1}} \quad \text{for} \quad 0 < t < t+\delta < 1.$$

For example, it follows for $0 < t_n < t_n + \delta_n < 1/2$ that

$$\lim_{n\to\infty} \mathrm{Cov}(\mathbb{Z}(t_n), \mathbb{Z}(t_n+\delta_n)) = 0 \;\text{ if }\; \lim_{n\to\infty} \delta_n/t_n = \infty,$$

and hence for such a choice of t_n and δ_n the random quantities $\mathbb{Z}(t_n)$ and $\mathbb{Z}(t_n+\delta_n)$ are asymptotically independent. Note that asymptotic uncorrelation and hence asymptotic independence of $\mathbb{Z}(t_n)$ and $\mathbb{Z}(t_n+\delta_n)$ is possible only if either $\lim_{n\to\infty} t_n = 0$ or $\lim_{n\to\infty}(t_n+\delta_n) = 1$, that is, either t_n, $n \in \mathbb{N}$, lie in the left tail or $t_n+\delta_n$, $n \in \mathbb{N}$, lie in the right tail. Therefore, there exists a considerable amount of points t in the sensitivity range R_n such that $\mathbb{Z}(t)$ evaluated in these points are asymptotically independent. On the other hand, we get for any central interval (d, e), that is for $0 < d < e < 1$, that

$$\mathrm{Cov}(\mathbb{Z}(t_n), \mathbb{Z}(t_n+\delta_n)) > \sqrt{\frac{d(1-e)}{e(1-d)}} \quad \text{if} \quad d < t_n < t_n+\delta_n < e \quad \text{for all} \quad n \in \mathbb{N}$$

and consequently there are no asymptotically independent points related to $\mathbb{Z}(t)$ in the central range.

Altogether, we obtain that the continuous as well as empirical HC process taken in the central range is asymptotically more (positively) dependent than in the sensitivity range R_n. In general, the supremum of any positively correlated normal variables is stochastically smaller than the supremum of the corresponding uncorrelated variables. This supports the fact that intermediates play the key role for the asymptotics of the HC statistic.

4 Why the asymptotics of the HC statistic is so poor

In this section, we approach the question why the asymptotics of the HC statistic is extremely slow, or, in other words, why $\mathrm{HC}_{0,1}$ is too large compared to the asymptotic critical value $b_n(x)$ given in (4). Below, we restrict attention to the HC statistic represented in the form (2) and study the quality of various asymptotic results related to truncated versions of the HC statistic when applied in the finite sample size case.

4.1 Left-truncated HC statistics and the role of the smallest order statistics

Now we show that the finite distribution of a left-truncated HC statistic is typically dominated by the asymptotic HC distribution F_n^∞ so that, in contrast to the original HC statistic, the asymptotic critical value $b_n(x)$ is not too small for such a left-truncated statistic.

First, we provide a simple numerical example. For $\alpha = 0.05$ and $n = 10^5$ we get that the asymptotic critical value $b_n(x_\alpha) = 3.497$ is too small for the HC statistic, that is $\mathrm{P}\big(\mathrm{HC}_{0,1} > b_n(x_\alpha)\big) \approx 0.123$. By means of numerical simulations we obtain that the critical value $b^* = 4.76$ leads to an (approximately) level α HC test for $\alpha = 0.05$ and $n = 10^5$, that is $\mathrm{P}\big(\mathrm{HC}_{0,1} > b^*\big) \approx \alpha$. Thereby, an interesting observation is that the first (three) smallest order statistics take most of the level α in the sense that $\mathrm{P}(\mathbb{Z}_n(U_{1:n}) > b^*) \approx 0.042$, $\mathrm{P}(\max_{1\le i\le 2} \mathbb{Z}_n(U_{i:n}) > b^*) \approx 0.047$ and $\mathrm{P}(\max_{1\le i\le 3} \mathbb{Z}_n(U_{i:n}) > b^*) \approx 0.049$. Note that a similar phenomenon can be observed for various other α- and n-values. All this indicates that the first (e.g. three) order statistics are crucial for the HC statistic in the finite case that leads to the question whether left-truncated versions of the HC statistics are (stochastically) considerably smaller than $\mathrm{HC}_{0,1}$ so that the limiting cdf F_n^∞ approximates (or even dominates) the distribution of these left-truncated statistics much better than the $\mathrm{HC}_{0,1}$-distribution.

A left-truncated HC statistic we are dealing with is defined as the maximum of the empirical HC process $\mathbb{Z}_n(t)$ evaluated at the order statistics $U_{i:n}$ except the first k ones, that is

$$\mathrm{HC}_{k+1,n} = \max_{k+1\le i\le n} \sqrt{n}\frac{i/n - U_{i:n}}{\sqrt{U_{i:n}(1-U_{i:n})}}.$$

Since $\mathrm{HC}_{k+1,n}$ can be seen as the maximum of just $n-k$ points of the empirical process $\mathbb{Z}_n(t)$, one can expect that for k large enough such a maximum is not larger than the supremum of the continuous process $\mathbb{Z}(t)$ evaluated in a suitable interval (d, e) with $0 < d < e < 1$. Clearly, the larger k, the smaller should be the corresponding interval (d, e). Remembering that the cdf of the supremum of $\mathbb{Z}(t)$ over (d, e) is given by F_γ^{OU} for $\gamma = \gamma(d, e)$ defined in (13) and that for γ small enough F_γ^{OU} leads to smaller critical values than the limiting cdf F_n^∞, cf. Section 3, one may expect that the asymptotic critical value $b_n(x_\alpha)$ is not too small for $\mathrm{HC}_{k+1,n}$ if k is large enough. Moreover, it seems that even for smaller k-values we are "on the safe side" due to the discreteness of the HC statistic.

For illustration we choose $k = 3$. Figure 5 shows the cdf of $\mathrm{HC}_{4,n}$ ($k = 3$) for $n = 10^4, 10^6$ together with the corresponding asymptotic cdf's F_n^∞. Apparently, almost all quantiles related to the truncated HC versions are smaller than the corresponding quantiles of the corresponding cdf $F_n^\infty(y)$ and hence the exceedance probability $\mathrm{P}(\mathrm{HC}_{k+1,n} > b_n(x_\alpha))$ is smaller than α for a lot of α-values. For example, for $\alpha = 0.05$ and $n = 10^4, 10^6$ we get $b_n(x_\alpha) = 3.434, 3.549$ and $\mathrm{P}(\mathrm{HC}_{k+1,n} > b_n(x_\alpha)) = 0.047, 0.042,$

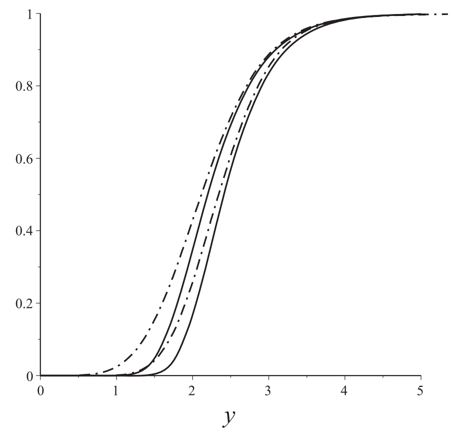

Figure 5 The cdf of $\mathrm{HC}_{k+1,n}$ for $k = 3$ (dash-dotted curves) simulated by 10^5 repetitions together with the asymptotic cdf $F_n^{\infty}(y)$ (solid curves) for $n = 10^4$ (left curves) and $n = 10^6$ (right curves).

respectively. For other choices of n the distribution of left-truncated HC statistics seems to behave similarly. Our simulations showed that for $n = 10^3, 10^4, 10^5, 10^6$, and $\alpha = 0.05$ the exceedance probability $\mathrm{P}(\mathrm{HC}_{k+1,n} > b_n(x_\alpha))$ is larger than α for $k = 0, 1, 2$, whereas $\mathrm{P}(\mathrm{HC}_{k+1,n} > b_n(x_\alpha)) \leq \alpha$ for $k \geq 3$.

Altogether, we have seen that for a finite (appropriate) $n \in \mathbb{N}$, the fact that the $\mathrm{HC}_{0,1}$ statistic is too large compared to the asymptotic critical value $b_n(x)$ results from the contribution of the smallest order statistics. Thus, in order to get some insight into the reasons for the poor HC asymptotics, we now focus on the behavior of the empirical HC process $\mathbb{Z}_n(t)$ evaluated at the first few order statistics.

4.2 Condition for the quality of the asymptotic approximation of the HC statistic

In this section, we provide a simple condition in terms of the first order statistic how to check for given $n \in \mathbb{N}$ and $\alpha \in (0, 1)$ whether the asymptotic critical value $b_n(x_\alpha)$ is too small for the test statistic $\mathrm{HC}_{0,1}$. For example, it turns out that for $\alpha = 0.05$ and $n \leq 10^{68}$ the exceedance probability $\mathrm{P}(\mathrm{HC}_{0,1} > b_n(x_\alpha))$ is strictly larger than α and hence $b_n(x_\alpha)$ is too small compared to $\mathrm{HC}_{0,1}$ for such n-values.

Firstly, we note that the finite cdf $F_n(y)$ of the statistic $\mathrm{HC}_{0,1}$ is not larger than the cdf of the HC process evaluated in any order statistic $U_{i:n}$, that is,

$$F_n(y) = \mathrm{P}\Big(\max_{1\leq i\leq n} \mathbb{Z}_n(U_{i:n}) \leq y\Big) \leq \mathrm{P}(\mathbb{Z}_n(U_{i:n}) \leq y).$$

An easy calculation yields

$$\mathrm{P}(\mathbb{Z}_n(U_{i:n}) \leq y) = \mathrm{P}(U_{i:n} \geq h_{i,n}(y)),$$

where

$$h_{i,n}(y) = (y^2 + 2i - y\sqrt{y^2 + 4i - 4i^2/n})/(2(y^2 + n)). \tag{21}$$

Setting $\bar{F}_{i,n}(y) \equiv \mathrm{P}(U_{i:n} \geq h_{i,n}(y))$, we get

$$F_n(y) \leq \min_{1\leq j\leq n} \bar{F}_{j,n}(y) \leq \bar{F}_{i,n}(y) \quad \text{for any} \quad i = 1, \ldots, n.$$

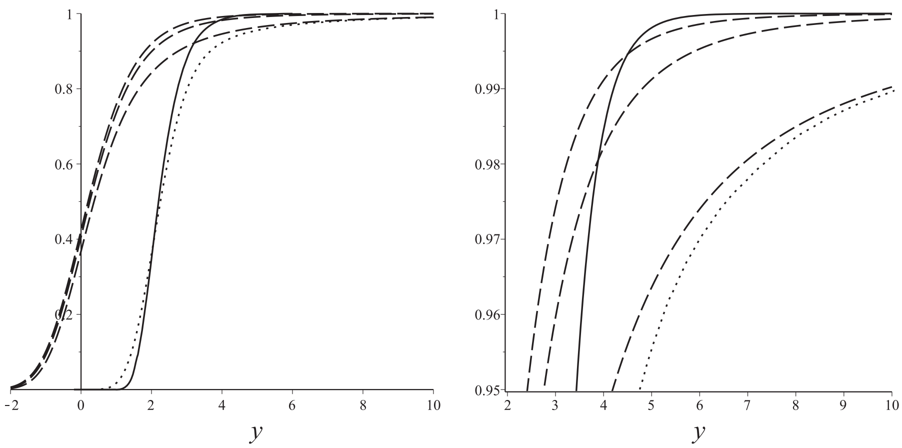

Figure 6 The cdf of $\mathrm{HC}_{0,1}$, that is $F_n(y)$, simulated by 10^5 repetitions (dotted curve), the cdf's of the HC process in the point $U_{i:n}$, that is $\bar{F}_{i,n}(y)$, for $i = 1, 2, 3$ (dashed curves from the bottom to the top in $y \geq 2$) and the corresponding asymptotic cdf $F_n^{\infty}(y)$ (solid curve) for $n = 10^4$. The right graph is zoomed.

An upper bound $\bar{F}_{i,n}(y)$ can be represented as the cdf of a binomially distributed random variable with parameters n and $h_{i,n}(y)$ so that $\bar{F}_{i,n}(y)$, $i = 1, \ldots, n$, can be calculated (at least numerically for a given $n \in \mathbb{N}$) by

$$\bar{F}_{i,n}(y) = \sum_{j=0}^{i-1} \binom{n}{j} h_{i,n}(y)^j (1 - h_{i,n}(y))^{n-j}, \quad i = 1, \ldots, n.$$

For example, Fig. 6 shows the first three upper bounds $\bar{F}_{i,n}(y)$, $i = 1, 2, 3$, together with the cdf $F_n(y)$ (simulated by 10^5 repetitions) and the asymptotic cdf $F_n^{\infty}(y)$ for $n = 10^4$. It looks like $\bar{F}_{1,n}(y)$ is the smallest of $\bar{F}_{i,n}(y)$, $i = 1, \ldots, n$, and hence the best (tighter) upper bound for $F_n(y)$. The picture on the right-hand side of Fig. 6 even reveals that $F_n(y)$ and $\bar{F}_{1,n}(y)$ nearly coincide in the right tail. What is more, based on results related to the theory of local levels (see Gontscharuk et al., 2013) we obtain that, in fact, the minimum of $\bar{F}_{i,n}(y)$, $i = 1, \ldots, n$, is asymptotically attained for $\bar{F}_{1,n}(y)$. All in all, it seems that $\bar{F}_{1,n}(y)$ is a good choice of an upper bound, which is pretty close to $F_n(y)$ in larger quantiles for appropriate values of n.

For a given $\alpha \in (0, 1)$ we show how to determine the sample size n_α (say) such that the asymptotic critical value $b_n(x_\alpha)$ is too small for the HC statistic $\mathrm{HC}_{0,1}$ for all $n \leq n_\alpha$. First, it is clear that $b_n(x_\alpha)$ is a good approximation for the exact critical value if

$$F_n(b_n(x_\alpha)) \approx 1 - \alpha.$$

On the other hand, the following inequality is always fulfilled:

$$\bar{F}_{1,n}(b_n(x_\alpha)) > F_n(b_n(x_\alpha)).$$

Therefore,

$$\bar{F}_{1,n}(b_n(x)) \geq 1 - \alpha \tag{22}$$

would be a necessary requirement that the asymptotic critical value $b_n(x_\alpha)$ is not too small compared to $\mathrm{HC}_{0,1}$, so that $b_n(x_\alpha)$ gives a good approximation to the $(1-\alpha)$-quantile of $\mathrm{HC}_{0,1}$. This condition can be verified for fixed values of n and α at least by means of numerical calculations. As expected, it turns out that a good approximation of the critical value by the asymptotic one can only be achieved by working with a very large sample size n. For example, for $\alpha = 0.05$ the inequality (22) is not fulfilled for $n \le 10^{68}$, while it is for $n = 10^{69}$. Hence, for practically relevant sample sizes, the asymptotic critical value $b_n(x_\alpha)$ with $\alpha = 0.05$ will always be too small compared to that of $\mathrm{HC}_{0,1}$.

In general, even for larger n the higher quantiles of the original HC statistic can be approximated much better by the corresponding quantiles of the HC process $\mathbb{Z}_n(t)$ evaluated only in the first order statistic $U_{1:n}$ than by those related to the asymptotic distribution. Hence, the $(1-\alpha)$-quantile related to $\mathbb{Z}_n(U_{1:n})$ is a better critical value for a test based on $\mathrm{HC}_{0,1}$. Clearly, it makes little sense to consider such a test, while instead of $\mathrm{HC}_{0,1}$ one may immediately choose $U_{1:n}$ as a test statistic.

5 New HC tests with improved finite properties

In view of the slow convergence of the HC statistic to the asymptotic distribution as discussed in the previous section, it is desirable to modify the HC test in order to improve its applicability in finite sample size settings. To this end, Gontscharuk et al. (2013) introduced the concept of so-called *local levels*.

For the HC test with critical value y a local level $\alpha_{i,n}(y)$ is defined as the probability that the ith order statistic $U_{i:n}$ falls below its respective critical value $h_{i,n}(y)$ defined in (21). Formally, local levels can be calculated via

$$\alpha_{i,n}(y) = \mathrm{P}(U_{i:n} \le h_{i,n}(y)) = 1 - \bar{F}_{i,n}(y), \quad i = 1, \ldots, n.$$

These quantities can be seen as an indicator as to where one would expect high/low local sensitivity of the test.

Theorem 5.1 in Gontscharuk et al. (2013) provides the result that the first local level $\alpha_{1,n}$ is asymptotically the largest. Even more, simulations in that paper show that $\alpha_{1,n}$ is much larger than the remaining local levels and takes up almost the entire α-level; for example, for the level α HC test in Subsection 4.1 we get $\alpha_{1,n}(y) \approx 0.042$ for $n = 10^5$ and $\alpha = 0.05$. From a practical point of view, this finite behavior is in contradiction with the asymptotic results related to the sensitivity range. In Theorem 5.1 in Gontscharuk et al. (2013) it is shown that HC local levels related to the sensitivity range are asymptotically equal in the sense that a fraction of two local levels tends to one as n increases. Motivated by this result, a new GOF test with equal local levels in the finite sample size case seems to be a good candidate for a better HC test. This test can be defined as follows. For a prechosen α_n^{loc} (say) we define a set of critical values $c_{i,n} \equiv c_{i,n}(\alpha_n^{loc})$ such that

$$\alpha_n^{loc} = \mathrm{P}(U_{i:n} \le c_{i,n}), \quad i = 1, \ldots, n. \tag{23}$$

Note that $c_{i,n}$, $i = 1, \ldots, n$, are determined in a unique way and can be calculated at least numerically. Then the new HC test rejects the null hypothesis H_0 that the underlying sample comes from the standard uniform distribution if $U_{i:n} \le c_{i,n}$ for at least one $i \in \{1, \ldots, n\}$. Thereby, in order to construct a level α test, $\alpha_n^{loc} \equiv \alpha_n^{loc}(\alpha)$ has to be such that

$$\mathrm{P}(U_{i:n} > c_{i,n},\ i = 1, \ldots, n) = 1 - \alpha. \tag{24}$$

Simple considerations imply $\alpha/n < \alpha_n^{loc} < \alpha$ for $n \in \mathbb{N}$. By means of Lemma 4.3 in Gontscharuk et al. (2013) it can be seen that GOF tests with all local levels equal to

$$\alpha_n^* \equiv \alpha_n^*(\alpha) = \frac{-\log(1-\alpha)}{2\log(\log(n))\log(n)} \tag{25}$$

yield an asymptotic level α test. Moreover, for level α GOF tests with local levels equal to α_n^{loc} we get

$$\lim_{n\to\infty} \alpha_n^*(\alpha)/\alpha_n^{loc} = 1.$$

The new HC level α test can alternatively be defined in terms of a test statistic. Let $F_{i,n}$ be the cdf of $U_{i:n}$, that is, $F_{i,n}$ is the cdf of the *Beta* distribution with parameters i and $n+1-i$. Noting that $F_{i,n}(c_{i,n}) = \alpha_n^{loc}$ via (23), condition (24) leads to

$$\begin{aligned} 1-\alpha &= \mathrm{P}(F_{i,n}(U_{i:n}) > \alpha_n^{loc},\ i=1,\dots,n) \\ &= \mathrm{P}(1-F_{i,n}(U_{i:n}) < 1-\alpha_n^{loc},\ i=1,\dots,n) \\ &= \mathrm{P}\left(\max_{1\le i\le n} \Phi^{-1}(1-F_{i,n}(U_{i:n})) < u_{1-\alpha_n^{loc}}\right), \end{aligned} \tag{26}$$

where Φ^{-1} is the inverse function of the standard normal cdf and $u_t = \Phi^{-1}(t)$. Once α_n^{loc} is determined, the new HC tests rejects if $F_{i,n}(U_{i:n}) \le \alpha_n^{loc}$ for some i. Note that $F_{i,n}(U_{i:n})$, $i = 1,\dots,n$, are uniformly distributed on [0, 1]. Alternatively, setting $W_i = \Phi^{-1}(1-F_{i,n}(U_{i:n}))$, $i = 1,\dots,n$, the new HC statistic can be represented as the maximum of standard normally distributed random variables, that is

$$\mathrm{HC}^{new} = \max_{1\le i\le n} W_i.$$

Consequently, the new HC test with equal local levels rejects H_0 if $\mathrm{HC}^{new} > u_{1-\alpha_n^{loc}}$. The p-value of this test can be calculated by

$$p(\mathrm{HC}^{new}) = 1 - \mathrm{P}\big(U_{i:n} > F_{i,n}^{-1}\big(\alpha_n^{obs}\big),\ i=1,\dots,n\big),$$

where α_n^{obs} is a realization of $\min_{1\le i\le n} F_{i,n}(U_{i:n})$. Thereby, the probability in the aforementioned expression can be calculated via one of the recursions provided in Shorack and Wellner (2009), p. 362–370.

Note that the transformation to the normal distribution in (26) may allow or facilitate the comparison of the new and original HC statistics due to the fact that $\mathrm{HC}_{0,1}$ can be approximated by the maximum of asymptotically standard normally distributed variables. Figure 7 shows the (simulated) cdfs of $\mathrm{HC}_{0,1}$ and HC^{new} together with the asymptotic cdf F_n^∞ and the cdf $F_{\log(n)}^{OU}$ of the maximum of the OU process. Here the cdf of the new HC statistic lies to the left of the asymptotic cdf F_n^∞, which leads to assume that the new HC test with the critical value $b_n(x_\alpha)$ instead of $u_{1-\alpha_n^{loc}}$ is a (conservative) level α test for a finite sample size. Moreover, it seems that the cdf of HC^{new} can be approximated by $F_{\log(n)}^{OU}$ much better than by F_n^∞. The behavior of the new and original HC tests under the null hypothesis as well as the concept of local levels are discussed in more detail in Gontscharuk et al (2013).

Finally, we briefly study and compare the power of the original and new HC test. We restrict attention to the following normal mixture model with sparse signals, a prominent example in the HC literature. Let $\epsilon_n = n^{-\beta}$, $\beta \in (1/2, 1)$, $\mu_n = \sqrt{2r\log(n)}$, $r \in (0, 1)$ and let $X_1,\dots,X_n$ be iid random variables with the cdf

$$F_n^X(x) = (1-\epsilon_n)\Phi(x) + \epsilon_n\Phi(x-\mu_n), \quad x \in \mathbb{R}.$$

We test whether any signals are present, that is, we test H_0 that $\epsilon_n = 0$ against the alternatives $H_{1,n}$ that $\epsilon_n > 0$ for $n \in \mathbb{N}$. This can be reworded as GOF testing for uniformity, where $U_i \equiv 1 - \Phi(X_i)$, $i = 1,\dots,n$, are iid uniformly distributed on [0, 1] under H_0. Donoho and Jin (2004) provided a detection boundary that separates the parameter plane (r, β) into two regions, where it is possible to

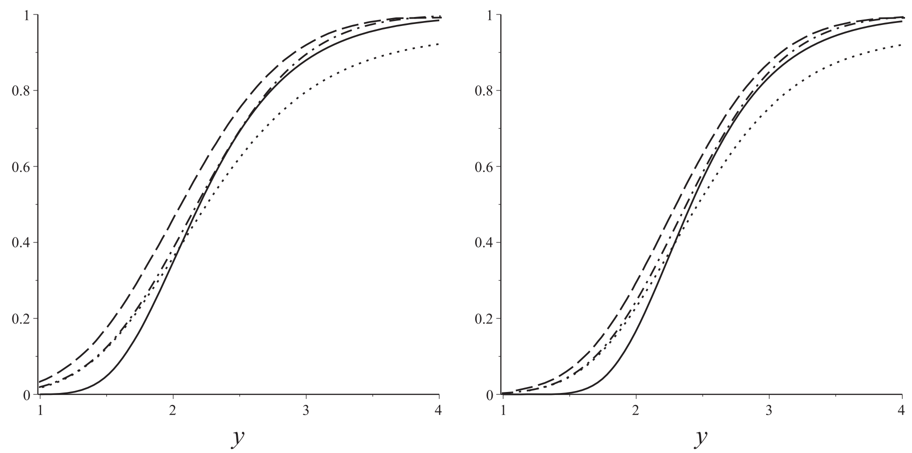

Figure 7 The cdf of HC^{new} (dashed curves) and the cdf of $\mathrm{HC}_{0,1}$ (dotted curves) simulated by 10^5 repetitions, asymptotic cdf $F_n^{\infty}(y)$ (solid curves) and finite OU cdf $F_{\log(n)}^{OU}(y)$ (dash-dotted curves) for $n = 10^4$ (left picture) and $n = 10^6$ (right picture).

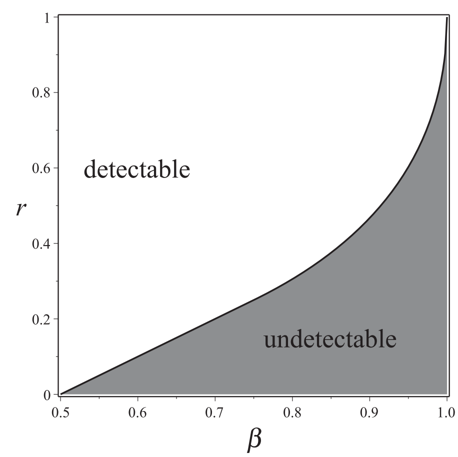

Figure 8 The detectable and undetectable regions of the (r, β) parameter plane of the HC test separated by the boundary function ρ^*.

reliably detect signals and where it is impossible to do so. They showed that depending on the considered parameters the power of the asymptotic level α HC test, which is defined as $\mathrm{P}(\mathrm{HC}_{0,1} > b_n(x_\alpha))$ under $H_{1,n}$, tends to one or α. More precisely, for the function

$$\rho^*(\beta) = \begin{cases} \beta - 1/2, & 1/2 < \beta \leq 3/4 \\ (1 - \sqrt{1-\beta})^2, & 3/4 < \beta < 1 \end{cases},$$

the power of the asymptotic level α HC test converges to one if $r > \rho^*(\beta)$ (detectable), and tends to α if $r < \rho^*(\beta)$ (undetectable), see also Ingster (1997, 1999). Figure 8 illustrates these two regions separated by the boundary function ρ^*.

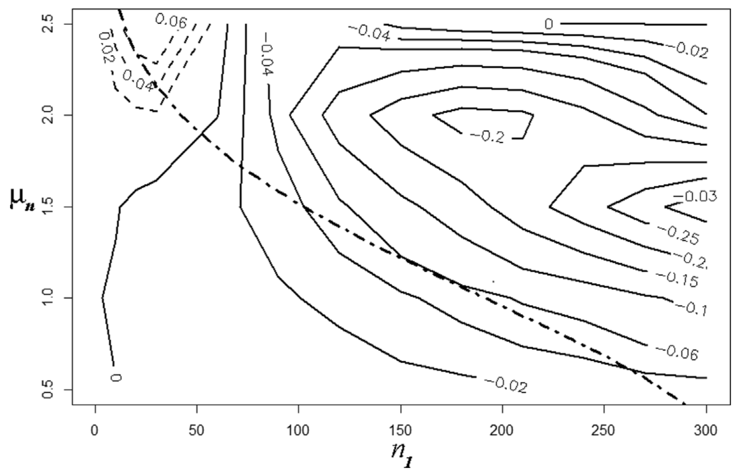

Figure 9 Difference in the power between the level α HC and new HC tests as a function of $n_1 = n\epsilon_n$ and μ_n for $n = 10^5$ and $\alpha = 0.05$ simulated by 4×10^4 repetitions. The area with a positive (negative) power difference, where the new HC test is more (less) powerful, is marked by solid (dashed) contours. The asymptotically detectable (undetectable) region of the HC test lies above (below) the dash-dotted curve.

Here, we take a look at the power of the HC and new HC tests in the finite case. Let $n = 10^5$ and $\alpha = 0.05$. Numerical calculations yield that the HC test with critical value $b^* \approx 4.76$ and the new HC test based on $\alpha_n^{loc} \approx 0.00056$ are level α tests and hence can be compared in a fair way. In contrast to the asymptotic case with parameter plane (β, r) we consider the power in the (n_1, μ_n)-plane, where $n_1 = n\epsilon_n$ is the number of signals. Figure 9 shows the power difference

$$\mathrm{P}(\mathrm{HC}_{0,1} > b^*) - \mathrm{P}\big(\mathrm{HC}^{new} > u_{1-\alpha_n^{loc}}\big) \quad \text{under} \quad H_{1,n}$$

for various values of n_1 and μ_n. It seems that the new HC test is more powerful in the largest part of the (n_1, μ_n)-plane with the power difference being considerable in a large part of the asymptotically detectable region, which lies above the dash-dotted line in Figure 9. On the contrary, the original HC test has larger power only if very few signals are present. For example, for $n_1 = 200$ and $\mu_n = 2$ the power of the HC test is ≈ 0.628 and the power of the new HC test is ≈ 0.855.

Finally, we consider “local powers” of the level α HC and new HC tests, which are defined as local levels under alternatives, that is

$$\mathrm{P}(U_{i:n} \le h_{i,n}(b^*)) \quad \text{and} \quad \mathrm{P}\big(F_{i:n}(U_{i:n}) \le \alpha_n^{loc}\big) \quad \text{under} \quad H_{1,n},$$

respectively. Hence, the ith “local power” is the probability that $U_{i:n}$ leads to the rejection of H_0 under $H_{1,n}$. Figure 10 shows these “local powers” of the level α HC tests for $n = 10^5$, $\alpha = 0.05$, $\varepsilon_n = 0.002$ (i.e. $n_1 = 200$) and $\mu_n = 2$. It is only the first 7 order statistics for which the “local powers” are larger for the original HC test than that for the new one while the “local power” for the new HC is much better than that for the original one for a large (especially the intermediate) range of i. It seems very likely that in the finite case the power of the original HC test is concentrated on the first few order statistics (extreme values), while the power of the new HC test is concentrated on the intermediates (moderate tail).

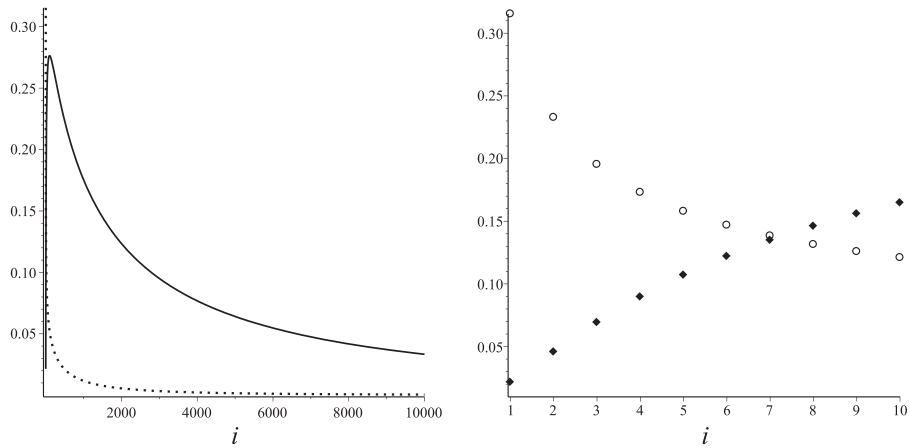

Figure 10 "Local powers" of the original level α HC test (dashed lines in the left graph, circles in the right graph) and the new level α HC test (solid lines in the left graph, solid diamonds in the right graph) for $n = 10^5$ and $\alpha = 0.05$ locally at each $i = 1, \ldots, 10^4$ (left graph) and $i = 1, \ldots, 10$ (right graph).

Altogether, it can be supposed that the new HC test finitely offers what the HC test asymptotically promises. Even more, beyond the context of testing the global null hypothesis in high dimensional data, the proposed new test with equal local levels is an attractive, easy to compute competitor to classical GOF tests and also allows for easy to compute simultaneous confidence bands for the related test statistic.

6 Concluding remarks

In this paper, we first focused on two topics that explain some important properties of the HC test statistic. The first one concerns the range of sensitivity of GOF tests based on the HC statistic, the second addresses the convergence of the finite HC distribution to its limiting counterpart. Eventually, we proposed a modification of the HC test showing better finite properties.

The fact that tests based on the HC statistic effectively detect signals which are very weak and very sparse, aroused a lot of interest among statisticians. Such tests are even asymptotically successful throughout the same region in the amplitude/sparsity plane where the oracle likelihood ratio test would succeed. The nature of this phenomenon becomes clearer by looking at the sensitivity region of the HC statistic. The explanation given in Section 3, why a certain intermediate range of order statistics plays a crucial role for the HC asymptotics, contributes to the knowledge about the behavior of the HC statistic. Unfortunately, due to the fact that the distribution of the HC statistic converges to the limiting one very slowly, the type I error by the corresponding HC tests is not controlled even for a very large sample size, cf. Section 4.

In view of this point, it would be favorable to have a "better" HC statistic at hand. It is worthy to note that the application of truncated HC versions leads to an exclusion of several order statistics of the underlying random sample and hence to a loss of information. This is why, in our opinion, truncated HC statistics are not suited to be "better" HC statistics. New approaches seem to be necessary in order to construct more favorable HC tests, that is a test with the same asymptotic properties as the original HC test, but with a more appropriate finite sample size behavior. The concept of the so-called local levels introduced in Section 5 seems to be such a promising approach. Local levels can be seen as an indicator as to where one would expect high/low local sensitivity, for more details see Gontscharuk et al. (2013). Motivated by a result in this work that almost all HC local levels are asymptotically equal, a new GOF test with equal local levels in the finite sample size case seems a good candidate that might show the properties mentioned. In Section 5, we provide new results concerning the asymptotics of the new HC test and show by means of simulations that the new test is typically more powerful than the original procedure in a normal mixture model. A more detailed study of the new HC test with equal local levels will be reported in our forthcoming work.

Acknowledgment 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).

Conflict of interest
The authors have declared no conflict of interest.

References

Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain "Goodness of fit" criteria based on stochastic processes. *Annals of Mathematical Statistics* **23**, 193–212.
Cai, T., Jin, J. and Low, M. (2007). Estimation and confidence sets for sparse normal mixtures. *Annals of Statistics* **35**, 2421–2449.
Cai, T., Jeng, J. and Jin, J. (2011). Optimal detection of heterogeneous and heteroscedastic mixtures. *Journal of the Royal Statistical Society B* **73**, 629–662.
De Long, D. M. (1981). Crossing probabilities for a square root boundary by a bessel process. *Communications in Statistics – Theory and Methods* **10**, 2197–2213.
Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Annals of Statistics* **32**, 962–994.
Donoho, D. and Jin, J. (2008). Higher criticism thresholding: optimal feature selection when useful features are rare and weak. *Proceedings of the National Academy of Sciences USA* **105**, 14790–14795.
Donoho, D. and Jin, J. (2009). Feature selection by higher criticism thresholding achieves the optimal phase diagram. *Philosophical Transactions of the Royal Society A* **367**, 4449–4470.
Eicker, F. (1979). The asymptotic distribution of the suprema of the standardized empirical processes. *Annals of Statistics* **7**, 116–138.
Estrella, A. and Rodrigues, A. P. (2005). One-sided test for an unknown breakpoint: theory, computation and application to monetary theory. Staff Report, no. 232.
Gontscharuk, V., Landwehr, S. and Finner, H. (2013). Goodness of fit tests in terms of local levels. Submitted for publication.
Hall, P. and Jin, J. (2008). Properties of higher criticism under strong dependence. *Annals of Statistics* **36**, 381–402.
Hall, P. and Jin, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Annals of Statistics* **38**, 1686–1732.
Hall, P., Pittelkow, Y. and Ghosh, M. (2008). Theoretical measures of relative performance of classifiers for high dimensional data with small sample sizes. *Journal of the Royal Statistical Society B* **70**, 159–173.
Ingster, Y. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distribution. *Mathematical Methods of Statistics* **6**, 47–69.
Ingster, Y. I. (1999). Minimax detection of a signal for l_n^p-balls. *Mathematical Methods of Statistics* **7**, 401–428.
Jaeschke, D. (1979). The asymptotic distribution of the supremum of the standardized empirical distribution function on subintervals. *Annals of Statistics* **7**, 108–115.
Jager, L. and Wellner, J. (2007). Goodness-of-fit tests via phi-divergences. *Annals of Statistics* **35**, 2018–2053.
Khmaladze, E. and Shinjikashvili, E. (2001). Calculation of noncrossing probabilities for Poisson processes and its corollaries. *Advances in Applied Probability* **33**, 702–716.
Leadbetter, M. R., Lindgren, G. and Rootzen, H. (1989). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York, NY.
Parkhomenko, E., Tritchler, D., Lemire, M., Hu, P. and Beyene, J. (2009). Using a higher criticism statistic to detect modest effects in a genome-wide study of rheumatoid arthritis. BMC Proceedings 3 Suppl 7, S40.

Sabatti, C., Service, S. K., Hartikainen, A.l., Pouta, A., Ripatti, S., Brodsky, J., Jones, C. G., Zaitlen, N. A., Varilo, T., Kaakinen, M., Sovio, U., Ruokonen, A., Laitinen, J., Jakkula, E., Coin, L., Hoggart, C., Collins, A., Turunen, H., Gabriel, S., Elliot, P., McCarthy, M. I., Daly, M. J., Järvelin, M. R., Freimer, N. B. and Peltonen, L. (2009). Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nature Genetics* **41**, 35–46.

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

Tukey, J. (1989). Higher criticism for individual significances in several tables or parts of tables. Internal Working Paper 89-9, Princeton University, Princeton, NJ.

