Distributions and Power of Optimal Signal-Detection Statistics in Finite Case

Hong Zhang, Jiashun Jin, and Zheyang Wu

***Abstract*—For detecting weak and sparse signals by a set of n input p-values, the Higher Criticism (HC) type statistics, the Berk-Jones (B-J) type statistics, and the phi-divergence statistics have the equivalent asymptotic optimality as n goes to infinity. However, they can have significantly different performance in practical data analysis, where n is always finite and even very small. To address this problem in a broader context, this paper introduces a general family of goodness-of-fit statistics, called the gGOF, which unifies a broad signal-detection statistics including these optimal ones. Efficient and accurate analytical calculations for the distributions of the gGOF statistics are provided under arbitrary *i.i.d.* continuous models of the null and the alternative hypotheses. Based on that, a systematic power study reveals that in finite case, the number of signals is often more relevant than the signal proportion. The HC and the reverse HC have advantages for relatively sparser and denser signals, respectively, while the B-J is more robust. A general framework is given to apply the gGOF into data analysis based on the generalized linear models. An application to the SNP-set based genome-wide association study (GWAS) for Crohn's disease shows that these optimal statistics have a good potential for detecting novel disease genes with weak SNP effects. The calculations have been implemented into an R package *SetTest* and published on the CRAN.**

***Index Terms*—Signal detection, hypothesis testing, statistical power, goodness-of-fit, genetic association.**

I. Introduction

Hypothesis-testing based statistical signal-detection method is an important approach for engineering and scientific researches. A theoretical study of this problem can be found in Arias-Castro, Donoho and Huo [1], and related applications in signal detection and processing are enormous [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12]. Here we take a simple example in genetic association studies. To determine whether a candidate gene is associated with a disease, we collect n p-values of single-nucleotide variants (SNVs) in this gene: $P_i, i = 1, ..., n$. These "input p-values" are then used to form a summary statistic. Testing this statistics will tell us whether there exists "signals", often represented by unexpectedly small p-values. Donoho and Jin's Higher Criticism (HC) statistic [13] is such a summary statistic. Denote $P_{(1)} \leq ... \leq P_{(n)}$ be the ordered p-values, the summary statistic is

$$HC = \sup_{\mathcal{R}=\{1\leq i\leq n/2\}} \sqrt{n}\frac{i/n - P_{(i)}}{\sqrt{P_{(i)}(1-P_{(i)})}}. \qquad (1)$$

Hong Zhang is with the Biostatistics and Research Decision Sciences, Merck Research Laboratories.

Jiashun Jin is with the Department of Statistics, Carnegie Mellon University.

Zheyang Wu is with the Department of Mathematical Sciences, Worcester Polytechnic Institute. E-mail: zheyangwu@wpi.edu. Zhang and Wu were partially supported by NSF grants DMS-1309960 and DMS-1812082.

Theoretical studies have revealed a collection of *asymptotic optimal tests* in the sense that they can asymptotically reach the boundary of the detectable region of signals. The HC type tests [13], [14], [15], [16], the Berk-Jones (B-J) type tests [17], and the ϕ-divergence type tests [18] are such optimal tests for example. These statistics share the same optimality property as $n \to \infty$ (see also Arias [19] and Cai [20]). However, when n is small these asymptotically equivalent statistics can have quite different power under various signal patterns. To study this problem in an even broader context, this paper unifies these optimal statistics within a general statistic family, referred as the gGOF. Novel methodology is developed to analytically calculate and compare power under general null and alternative hypotheses.

A. Our Contributions

This paper is largely motivated by practical problems and makes three main contributions.

- We propose a general family of goodness-of-fit test statistics, the gGOF, which covers any supremum-based one-side goodness-of-fit statistics with arbitrary truncation of the input p-values. The gGOF contains the asymptotic optimal statistics described above, and provides a general strategy for the signal detection problem.
- Novel analytical methods are developed to calculate the distributions and the statistical power of the gGOF. Both exact and approximate methods are studied. Comparing with relevant literature, this work improves generality and computational efficiency.
- Through careful power comparisons over various signal patterns, we reveal relative advantages among these optimal statistics under finite n. The results provide a useful guidance for practitioners to choose proper statistics based on their signal patterns and data properties.

B. Connection to Distribution Calculation Literature

Analytically calculating the distributions of relevant statistics has a critical advantage over empirical methods, such as Monte-Carlo simulation and permutation [2], [7]. Well-designed analytical calculation often provide a higher accuracy while requiring much less computation. Moreover, calculation for power can provide mathematical insights to elucidate the mechanism of statistical signal detection.

Our calculation methods are related to Denuit et al. [21] (which covers the result of a more recent work [22]). However, our work possesses significant advantages in both generality and efficiency. First, we address not only the size but also

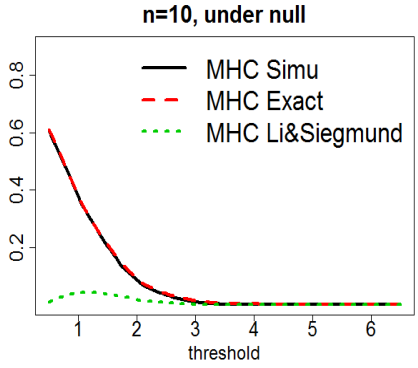


Fig. 1. Comparisons among different methods for calculating the MHC p-values over threshold b. Simu: curves obtained by simulation; Exact: by our Corollary 5.1; Li&Siegmund: by [23].

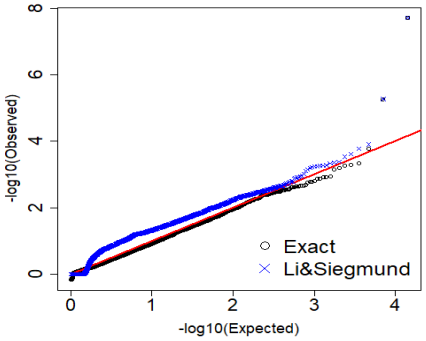


Fig. 2. The QQ plots of all genes' expected MHC p-values versus the observed ones by calculation. Exact: by Corollary 5.1; Li&Siegmund: [23].

the statistical power of relevant tests. Second, our calculations allow arbitrary truncation on input p-values. Such a truncation procedure is an important component of the test statistics. For example, the modified HC (MHC), where the truncation domain is modified to be $\mathcal{R} = \{1 \le i \le n/2, P_{(i)} \ge 1/n\}$ for improving performance [13], [23]. This truncation cannot be handled by Denuit's method. Third, our calculation has a lower computational complexity. These related new developments are technically non-trivial.

Another closely related work is by Li and Siegmund (LS) [23]. The LS method is an asymptotic approximation for HC and B-J type statistics, for which it requires a threshold $b = O(\sqrt{n})$ in p-value calculation. At $n = 10$, Fig. 1 shows the MHC p-values over b, for which the LS method is not accurate at small b. Meanwhile, LS method is easy to compute and is satisfiable at large n. Inspired by the LS, we also studied approximate calculations. There, the main difference is that we propose to use the gamma approximation, instead of the beta approximation by the LS. Our formula has the same asymptotic accuracy, while the performance could be improved sometimes. Also, the proof is simplified, which helps us to get a sufficient condition for addressing the whole gGOF family under general hypotheses.

Going back to the example of detecting genetic signals, Fig. 2 shows the QQ-plots in a GWAS of Crohn's disease. Each dot represents a gene for which its MHC p-value is calculated. By our calculation, most dots are aligned along the diagonal line, indicating that the genome-wide type I error is well controlled. The dots by the LS method are significantly off the diagonal line, indicating its limitation in real data analysis.

The remainder of the paper is organized as follows. In Section II the statistical signal detection problem is formulated. We review the literature of the asymptotic optimal tests in Section III, and define the gGOF family that covers these tests in Section IV. Both exact and approximate calculations for the null and the alternative distributions of the gGOF are presented in Section V. Section VI numerically evidences the calculation accuracy, and provides systematic power comparisons among the asymptotic optimal tests. We give a framework of applying the gGOF under the generalized linear models in Section VII, and illustrate the real GWAS of Crohn's disease in Section VIII. Section IX summarizes this work. Detailed proofs and supportive lemmas are given in the supplemental document.

II. Hypothesis-Testing Based Signal Detection

We consider statistical signal detection problem based on hypothesis testing. With a group of *input statistics*, each could be a data summary or a random observation, we aim to determine whether there is a significant statistical evidence for the existence of "signals". The null and the alternative hypotheses can be loosely stated as [2]

$$\begin{aligned} H_0 &: \text{Only background noise present.} \\ H_1 &: \text{Both noise and signals present.} \end{aligned} \tag{2}$$

In statistics, signals can be characterized by the contrast between the distributions under H_0 and H_1. Specifically, let $T_1, ..., T_n$ be the input statistics, the hypotheses are

$$H_0 : T_i \overset{i.i.d.}{\sim} F_0, \quad H_1 : T_i \overset{i.i.d.}{\sim} F_1, i = 1, ..., n, \tag{3}$$

where $F_j, j = 0, 1$, denote two continuous cumulative distribution functions (CDFs) under these hypotheses. The signals are modeled by the distinction between F_0 and F_1. As a special case, the classic Gaussian mixture model is defined as [13], [15], [24]:

$$H_0 : T_i \sim F_0 = \Phi, \quad H_1 : T_i \sim F_1 = \epsilon\Phi_\mu + (1-\epsilon)\Phi, \tag{4}$$

for $i = 1, ..., n$, where Φ and Φ_μ are the CDFs of $N(0, 1)$ and $N(\mu, 1)$, respectively. H_1 indicates that $\epsilon \in (0, 1)$ proportion of n input statistics are associated with the signals, with a signal strength characterized by μ.

Hypothesis testing based on the input statistics is equivalent to that based on the *input p-values* $P_1, ..., P_n$, where

$$P_i = 1 - F_0(T_i), \quad i = 1, ..., n. \tag{5}$$

Since F_0 and F_1 are allowed arbitrary, the input p-values could be two-sided. For example, if F_0 is symmetric around 0 (e.g., $N(0, 1)$), the input statistics can be replaced by $T_i' = T_i^2 \sim F_0'$. Therefore the framework allows detecting directional signals, e.g., both protective and deleterious effects of genetic mutations.

The advantage of using input p-values over T_i's for summary statistic is that they directly measure the significance of

T_i's, which could have different scales. In fact, the null hypothesis in (3) can be further generalized to allow heterogeneous distributions:

$$H_0 : T_i \overset{ind}{\sim} F_{0i}, \quad i = 1, ..., n. \tag{6}$$

For example, in meta-analysis or integrative analysis of heterogeneous data, the input test statistics could follow different distributions. As long as F_{0i} are continuous, we always have a homogeneous distribution of the input p-values under the null:

$$H_0 : P_i \overset{i.i.d.}{\sim} \text{Uniform}[0, 1], i = 1, ..., n. \tag{7}$$

The input p-values are used to form a summary *test statistic* for testing the hypotheses. The p-value of the test statistic, which we call the *test p-value* in order to distinguish from the term input p-values, is then used for statistical evidence against the null hypothesis.

III. Asymptotic optimal Tests for Weak-Sparse Signals

The *asymptotic optimal tests* are those that reach the *asymptotic detection boundary* indicating the minimal signal intensity required for reliable detection. Consider the asymptotic rare and weak (ARW) setting. The parameters in (4) are regulated as $\epsilon_n = n^{-\alpha}$, where $\alpha \in (1/2, 1)$ is the signal sparsity parameter, $\mu_n = \sqrt{2r\log(n)}$, where $r \in (0, 1)$ is the signal strength parameter. A few seminal studies [13], [25], [26], [20] discovered the asymptotic detection boundary in terms of a function curve of α:

$$\rho^\star(\alpha) = \begin{cases} \alpha - 1/2 & 1/2 < \alpha \le 3/4 \\ (1 - \sqrt{1-\alpha})^2 & 3/4 < \alpha < 1. \end{cases} \tag{8}$$

When the signals are too weak, i.e., $r < \rho^\star(\alpha)$, the power of any statistics will converge to its type I error rate as $n \to \infty$. That is, no reliable detection is available. Whenever the signal strength is above this lower bound, i.e., $r > \rho^\star(\alpha)$, the asymptotic optimal tests will be capable to assure the power converge to 1 at any fixed type I error rate.

Based on an initial idea of Tukey's [27], one type of optimal statistics are the HC type statistics:

$$\begin{aligned} HC^{2004}_{n,\mathcal{R}} &= \sup_{\mathcal{R}} \sqrt{n}\frac{i/n - P_{(i)}}{\sqrt{P_{(i)}(1-P_{(i)})}}, \\ HC^{2008}_{n,\mathcal{R}} &= \sup_{\mathcal{R}} \sqrt{n}\frac{i/n - P_{(i)}}{\sqrt{i/n(1-i/n)}}, \end{aligned} \tag{9}$$

which were defined in 2004 and 2008 [13], [14], respectively ($HC^{2008}_{n,\mathcal{R}}$ is also called the reverse HC). $\mathcal{R}$ denotes a truncation domain of the input p-values. For example, the original HC is defined by $HC^{2004}_{n,\mathcal{R}}$ with $\mathcal{R} = \{1 \le i \le n/2\}$ [13]. Note that HC^{2004} statistic is similar as the Anderson-Darling statistic [28], but is more general because it is based on input p-values rather than input statistics, and allows a truncation domain $\mathcal{R}$. Note also that in literature [13], [19], [29] the HC statistic was also written as

$$HC = \sup_{t \in \mathcal{R}^*} \frac{\sum_i \{T_i > t\} - n\bar{\Phi}(t)}{\sqrt{n\bar{\Phi}(t)\Phi(t)}}, \tag{10}$$

where $\bar{\Phi}(t) = 1 - \Phi(t)$. In this paper, however, we do not follow this formula because it is restricted to the hypothesis setting of $F_0 = \Phi$. Note also that the supremum domain $\mathcal{R}^*$ on t is equivalent to $\mathcal{R}$ on $P_{(i)}$, but not on the index i.

Besides the HC type statistics, the Berk-Jones (B-J) type statistics and a spectrum of ϕ-divergence statistics were also proven to be asymptotically optimal [13], [14], [30], [18]. All of them can be covered by the gGOF family as explained below.

IV. Generalization of Goodness-of-Fit Tests

The gGOF family follow the traditional idea of the goodness-of-fit test. We aim to determine whether the input p-values have a good "fit" with the uniform distribution under the null in (7). The fitness is measured by pair-wise comparisons between the ordered input p-values and their null expectations. If any pair is quite different, then the null is likely not true. Therefore, a gGOF statistic is defined based on the supremum of a generic contrast function f:

$$S_{n,\mathcal{R}} = \sup_{\mathcal{R}} f(\frac{i}{n}, P_{(i)}), \tag{11}$$

in which the truncation domain of the input p-values is

$$\mathcal{R} = \{i : k_0 \le i \le k_1\} \bigcap \{P_{(i)} : \alpha_0 < P_{(i)} < \alpha_1\}, \tag{12}$$

for given $k_0 \le k_1 \in \{1, ..., n\}$ and $\alpha_0 \le \alpha_1 \in [0, 1]$. If the null is untrue, $P_{(i)}$ will likely departure from their null expectations $E(P_{(i)}) = \frac{i}{n+1}$, which is to be captured by the contrast function f. The gGOF test should be one-sided for signal detection because the smaller, rather than the larger, input p-values are more likely associated with signals. Therefore $f(x, y)$ can be any decreasing function in y at fixed x so that the smaller the input p-values, the larger the statistic, and the stronger the evidence is. Following the tradition of GOF, $\frac{i}{n}$, instead of $\frac{i}{n+1}$, are used here to represent the null means, which imposes no practical difference in statistical power.

For both theoretical and practical reasons, it is important to allow a general truncation domain $\mathcal{R}$ to restrict both the index i and the magnitude of the input p-values $P_{(i)}$. Besides the benefit of computational efficiency (e.g., big input p-values can be truncated since they are likely not signal related), the performance could also be improved by excluding some input p-values. For example, as $n \to \infty$, HC could have the long-tail problem due to the possibility of getting very small input p-values under the null. To address this issue, the MHC was created with $\mathcal{R} = \{1 \le i \le n/2, P_{(i)} \ge 1/n\}$ [13], [22], [23]. The significant influence of restricting $P_{(i)} \ge 1/n$ under finite n is also demonstrated in Section VI.

The gGOF family covers a lot of classic test statistics. The simple one-sided Kolmogorov-Smirnov (denoted KS^+) test statistic directly measures the difference between $P_{(i)}$ and i/n (c.f. [31], page 447). Under the roof of the gGOF family in (11), the KS^+ statistic corresponds to the contrast f function:

$$f_{KS^+}(x, y) = x - y. \tag{13}$$

Because smaller input p-values are more likely to indicate the alternative, the absolute difference $i/n - P_{(i)}$ should be

reweighed with regard to $P_{(i)}$ or i/n. Such rescaling leads to the Higher Criticism (HC) statistics:

$$\begin{aligned} f_{HC^{2004}}(x,y) &= \sqrt{n}\frac{x-y}{\sqrt{y(1-y)}}; \\ f_{HC^{2008}}(x,y) &= \sqrt{n}\frac{x-y}{\sqrt{x(1-x)}}. \end{aligned} \tag{14}$$

Jager and Wellner introduced a collection of ϕ-divergence statistics [18]; each one of them can be written by a contrast function indexed by a parameter s:

$$\begin{aligned} f_s^{\phi}(x,y) &= \frac{1}{s(1-s)}(1-x^s y^{1-s}-(1-x)^s(1-y)^{1-s}), \\ &\quad s\neq 0,1, \\ f_1^{\phi}(x,y) &= x\log(\frac{x}{y})+(1-x)\log(\frac{1-x}{1-y}), \\ f_0^{\phi}(x,y) &= y\log(\frac{y}{x})+(1-y)\log(\frac{1-y}{1-x}). \end{aligned} \tag{15}$$

At certain s values (e.g., $s=2$ or -1) these statistics are two-sided in the sense that switching the values of $x=i/n$ and $y=P_{(i)}$ does not change the statistics. However, as mentioned above, because smaller input p-values indicates signals, we consider the one-sided version of ϕ-divergence statistics. A simple adjustment of the f function could be:

$$f_s(x,y)=\begin{cases} \sqrt{2nf_s^{\phi}(x,y)} & y\le x, \\ -\sqrt{2nf_s^{\phi}(x,y)} & y>x. \end{cases} \tag{16}$$

Now for all s, $f_s(x,y)$ is guaranteed decreasing in y. Such one-sided ϕ-divergence statistics cover the HC exactly: $f_2=f_{HC^{2004}}$ and $f_{-1}=f_{HC^{2008}}$. Also, $s=1$ corresponds to the one-sided Berk-Jones statistic (i.e., the R_n^+ statistic in equation (1.8) of [17], or the BJ_n^+ in (1.9) of [13], or the T_{BJ} in (2) of [23]); $s=0$ gives the one-sided reverse Berk-Jones statistic (i.e. the $\tilde{R}_n$ in (6) of [32] or the equation (1) and discussion in [18]).

V. Analytical Calculation Methods for gGOF Distributions

This section presents our analytical calculation methods for the distributions of gGOF test statistics under both the null and alternative hypotheses. We first summarize the general idea of the calculation. Then guided by that, specific approaches for calculating the exact or approximated distributions will be developed under specific settings and assumptions.

A. General Calculation Strategy

Consider the general hypothesis models in (3). For any given continuous CDFs F_0, F_1, we define a monotone transformation function in domain $[0,1]$:

$$D(x)=\begin{cases} x & \text{under } H_0, \\ 1-F_1(F_0^{-1}(1-x)) & \text{under } H_1. \end{cases} \tag{17}$$

Note that for any input p-value P_i, $D(P_i)\sim$ Uniform$[0,1]$ under either H_0 or H_1.

Secondly, consider the gGOF statistic $S_{n,\mathcal{R}}$ in (12). For each fixed x, we define the inverse of the contrast function $f(x,y)$:

$$g(x,\cdot)=f^{-1}(x,\cdot). \tag{18}$$

For example, the g functions for the HC statistics defined in (14) at a fixed b are

$$\begin{aligned} g_{HC^{2004}}(x,b) &= \tfrac{1}{1+b^2/n}[x+\tfrac{b^2/n-(b/\sqrt{n})\sqrt{b^2/n+4x(1-x)}}{2}]; \\ g_{HC^{2008}}(x,b) &= x-(b/\sqrt{n})\sqrt{x(1-x)}. \end{aligned} \tag{19}$$

Note that the g function corresponds to the "rejection curve" as specified in [33]. In general, if the closed form of a g function is not available (e.g., for a ϕ-divergence statistic with arbitrary s), it can always be numerically obtained since $f(x,y)$ is strictly decreasing in y.

Now under either H_0 or H_1, the CDF of $S_{n,\mathcal{R}}$ is

$$\begin{aligned} P(S_{n,\mathcal{R}}\le b) &= P(\sup_{\mathcal{R}} f(\frac{i}{n},P_{(i)})\le b) \\ &= P(\bigcap_{\mathcal{R}}\{P_{(i)}>g(\frac{i}{n},b)\}) \\ &= P\{D(P_{(i)})>D(g(\tfrac{i}{n},b)), \text{all } i,P_{(i)} \text{ in } \mathcal{R}\}. \end{aligned} \tag{20}$$

For both exact and approximate calculations of the distributions, we take advantage of the fact that under either H_0 or H_1, $U_{(i)}:=D(P_{(i)})$ is the i^{th} order statistic of Uniform[0, 1], and we study the joint distribution of $U_{(i)}$ under the restriction $\mathcal{R}$ in different ways.

To simplify the presentation, we list below the notations to be referred later on.

(N1) Based on equations (17) and (18), define

$$u_k := D(g(\frac{k}{n},b)\vee\alpha_0),$$

where $\alpha_0\ge 0$ is the lower bound constant for truncating $P_{(i)}$ in (12).

(N2) Let $F_{B(\alpha,\beta)}(x)$ and $\bar{F}_{B(\alpha,\beta)}(x)=1-F_{B(\alpha,\beta)}(x)$ denote the CDF and survival function of Beta(α,β) distribution.

(N3) Let $F_{\Gamma(\alpha)}(x)$ and $\bar{F}_{\Gamma(\alpha)}(x)$ denote the CDF and survival function of Gamma$(\alpha,1)$ distribution, respectively, where the shape parameter is α and the scale parameter is 1.

(N4) Based on the notation *(N3)*, define

$$h_k(x) := xF_{\Gamma(k-1)}(kx)-F_{\Gamma(k)}(kx).$$

(N5) Let $f_{P(\lambda)}(x)$ denote the probability mass function of Poisson(λ) distribution.

B. Exact Calculations

In this section we provide calculation methods for the exact distributions of any gGOF statistics in (11) under either H_0 or H_1 in (3). Accordingly, the test p-value and the statistical power of gGOF can be calculated in an exact manner. Three main theorems are provided, each concerns a specific truncation domain $\mathcal{R}$. The first theorem is for truncation based on the index i only. For example, the initial HC was defined with $\mathcal{R}=\{1\le i\le n/2\}$ [13].

Theorem 5.1: Consider any gGOF statistic in (11) with $\mathcal{R}=\{k_0\le i\le k_1\}$ for given $1\le k_0\le k_1\le n$. Let $m=n-k_1+1$. Follow notations (N1) and (N2), and define

$$\begin{aligned} a_{k_1} &= \tfrac{n!}{(n-k_1+1)!}\bar{F}_{B(1,m)}(u_{k_1}), \text{ and for } k=k_1-1,...,1, \\ a_k &= \frac{n!}{(n-k+1)!}\bar{F}_{B(k_1-k+1,m)}(u_{k_1})-\sum_{j=1}^{k_1-k}\frac{u_{k+j-1}^j}{j!}a_{k+j}. \end{aligned}$$

Under either H_0 or H_1, we have

$$P(S_{n,\mathcal{R}} \le b) = \bar{F}_{B(k_1,m)}(u_{k_1}) - \sum_{i=k_0}^{k_1-1} \frac{u_i^i}{i!} a_{i+1}.$$

It should be noted that for calculating p-values of goodness-of-fit type statistics, recursive methods are a classic way [34], [35], [36], [37], [38], [39], [40], [29]. The limitation is that these methods do not allow truncation of the input p-values (i.e., they require $\mathcal{R} = \{1 \le i \le n\}$) and their computational complexity is $O(n^3)$, which is pretty high. Denuit's method [21], [22] allows $\mathcal{R} = \{k_0 \le i \le k_1\}$ and at the same time reduces the complexity to $O(n^2)$. Our result in Theorem 5.1 further reduces the complexity to $O((k_1 - k_0)^2)$, which is significant especially when $(k_1 - k_0) = o(n)$.

Going beyond the scope of Denuit's method, the next theorem concerns a non-trivial scenario where the truncation domain is based on the magnitude of input p-values (rather than their indices), i.e., $\mathcal{R} = \{\alpha_0 \le P_{(i)} \le \alpha_1\}$. This is an important scenario, for example, the HC statistic defined in (10) and the MHC are related to such kind of truncation.

Theorem 5.2: Consider any gGOF statistic in (11) with $\mathcal{R} = \{\alpha_0 \le P_{(i)} \le \alpha_1\}$ for given $0 \le \alpha_0 < \alpha_1 \le 1$. Follow notations (N1) and (N2), and define

$$\begin{aligned}
&\beta_0 = D(\alpha_0), \quad \beta_1 = D(\alpha_1), \quad c_{ij} = \tfrac{\beta_0^{i-1}(1-\beta_1)^{n-j+1}}{(i-1)!(n-j+1)!}, \\
&a_j(k) = \frac{n!}{(j-k)!}\beta_1^{j-k}\bar{F}_{B(j-k,1)}(\frac{u_{j-1}}{\beta_1}) - \sum_{l=1}^{j-k} \frac{u_{k+l-1}^l}{l!} a_j(k+l), \\
&a_j(j) = 0, \quad 1 \le i \le n, i < j \le n+1, k = 1, ..., j-1.
\end{aligned}$$

Under either H_0 or H_1, we have

$$P(S_{n,\mathcal{R}} \le b) = \sum_{i=1}^{n} \sum_{j=i+1}^{n+1} c_{ij} a_j(i).$$

Comparing Theorems 5.1 and 5.2, it is clear that the truncation imposed on $P_{(i)}$ requires much more complicated computation than the truncation imposed on i. The complexity of the formula in Theorem 5.2 is $O(n^3)$ (or more precisely $O(n^3/6)$). Next, the following theorem provides the exact calculation under the most general $\mathcal{R}$ defined in (12), where the truncation is for both the index and the input p-values.

Theorem 5.3: Consider any gGOF statistic in (11) with general $\mathcal{R} = \{\alpha_0 \le P_{(i)} \le \alpha_1\} \cap \{k_0 \le i \le k_1\}$ for given $1 \le k_0 \le k_1 \le n$ and $0 \le \alpha_0 < \alpha_1 \le 1$. Follow notations (N1) and (N2) and those in Theorem 5.2. For $1 \le i \le k_1, \tilde{i} < j \le n+1, k = 1, ..., \tilde{j} - 1$, define $\tilde{i} = i \vee k_0$, $\tilde{j} = j \wedge (k_1 + 1)$, $\tilde{\beta}_0 = \beta_0 I_{\{i<k_0\}}$,

$$\begin{aligned}
a_j(k) \quad = \quad & \frac{n!}{(j-k)!}\beta_1^{(j-k)}\bar{F}_{B(\tilde{j}-k,j-\tilde{j}+1)}(\frac{u_{\tilde{j}-1}}{\beta_1}) \\
& - \sum_{l=1}^{\tilde{j}-k} \frac{u_{k+l-1}^l}{l!} a_j(k+l),
\end{aligned}$$

and $a_j(\tilde{j}) = 0$. Under either H_0 or H_1, we have

$$\begin{aligned}
&P(S_{n,\mathcal{R}} \le b) = \\
&\sum_{i=1}^{k_1} \sum_{j=\tilde{i}+1}^{n+1} c_{ij} \Big[\frac{n!(\beta_1 - \tilde{\beta}_0)^{j-i}}{(j-i)!} \bar{F}_{B(\tilde{j}-i,j-\tilde{j}+1)}(\frac{u_{\tilde{j}-1} - \tilde{\beta}_0}{\beta_1 - \tilde{\beta}_0}) \\
&\qquad - \sum_{k=\tilde{i}}^{\tilde{j}-1} \frac{(u_k - \tilde{\beta}_0)^{k-i+1}}{(k-i+1)!} a_j(k+1) \Big].
\end{aligned}$$

The complexity of the formula in Theorem 5.3 is $O(nk_1^2)$. Adding truncation on index i actually simplifies the computation comparing with Theorem 5.2. As discussed above, too small input p-values under H_0 is a concern for the performance of some gGOF statistics (e.g., causing long-tail problem for HC). Thus, the truncation on input p-values could be on the lower bound α_0 only, which can also significantly reduce the computational complexity. Corollary 5.1 below addresses such special case of Theorem 5.3 with $\alpha_1 = 1$, where the formula complexity reduces to $O(k_1^2)$.

Corollary 5.1: Consider any gGOF statistic in (11) with $\mathcal{R} = \{\alpha_0 \le P_{(i)}\} \cap \{k_0 \le i \le k_1\}$ for given $1 \le k_0 \le k_1 \le n$ and $\alpha_0 > 0$. Follow notations (N1) and (N2) and those in Theorem 5.1, 5.2. Define $c_i = \frac{\beta_0^{i-1}}{(i-1)!}, 1 \le i \le k_1$. Under either H_0 or H_1, we have

$$\begin{aligned}
&P(S_{n,\mathcal{R}} \le b) = \\
&\sum_{i=1}^{k_1} c_i \Big[\frac{n!(1-\tilde{\beta}_0)^{n+1-i}}{(n+1-i)!} \bar{F}_{B(k_1+1-i,m)}(\frac{u_{k_1} - \tilde{\beta}_0}{1-\tilde{\beta}_0}) \\
&\qquad - \sum_{k=\tilde{i}}^{k_1-1} \frac{(u_k - \tilde{\beta}_0)^{k+1-i}}{(k+1-i)!} a_{k+1} \Big].
\end{aligned}$$

A special case of Corollary 5.1 is the MHC in (9) with $\mathcal{R} = \{1 \le i \le n/2, P_{(i)} \ge 1/n\}$[13], [23]. As shown in Fig. 3, LS approximation [23] is good only for the right-tail of the distribution under H_0. Corollary 5.1 gives the perfect distributions under both H_0 and H_1.

Obviously, Theorem 5.3 addresses the most general truncation and covers other theorems and corollary. Based on this general formula, the formula in Theorem 5.1 is obtained by fixing $i = 1$, $j = n + 1$, $\alpha_0 = 0$, and $\alpha_1 = 1$. The formula of Theorem 5.2 is covered by letting $k_0 = 1, k_1 = n$. The formula of Corollary 5.1 is covered by fixing $j = n + 1$ and $\alpha_1 = 1$. However, we still separate these formulas and their implementations in order to simplify the computation whenever possible.

C. Asymptotic Calculations

In this section we study approximation approaches for calculating the distributions of the gGOF statistics based on appropriate asymptotics. The purpose is to 1) further simplify computation, and 2) reveal more insights to understand the gGOF performance. Approaches are backed by asymptotics but hold good accuracy under small or moderate n.

Two strategies are considered here. First, we follow the basic idea of the exact calculation described above, except applying distribution approximation. This strategy maintains

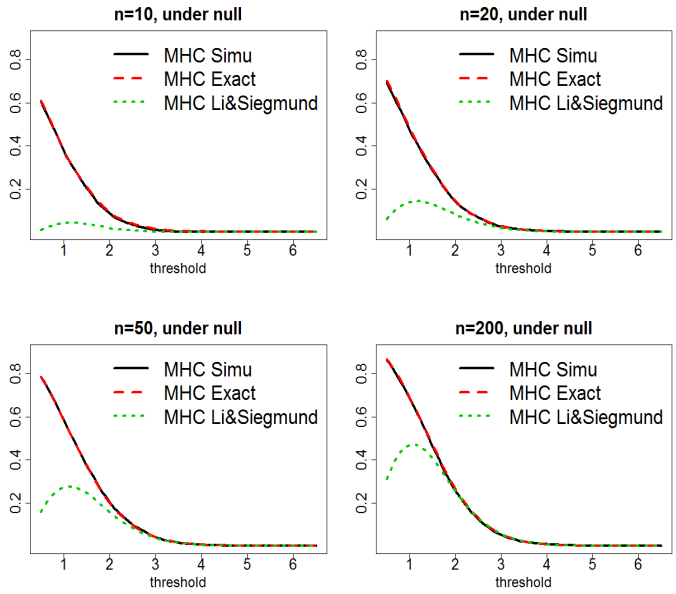


Fig. 3. Comparisons among different methods for calculating the MHC p-values over threshold b. Simu: curves obtained by simulation; Exact: by our Corollary 5.1; Li&Siegmund: by [23].

the generality of the results and provides the inspiring technique of the gamma approximation. The second strategy is to follow the LS style asymptotics [23] and produce non-iterative one-step formulas. The cost for such simplified calculation is the requirement of stronger assumptions. For the simplicity of presentation, the theorems below focus on the case of $\mathcal{R} = \{k_0 \le i \le k_1\}$. The results can be extended to a more general $\mathcal{R}$ for truncation on $P_{(i)}$.

By the first approximation strategy Theorem 5.4 below gives a formula based on the approximation by the joint gamma distribution.

Theorem 5.4: Consider any gGOF statistic in (11) with $\mathcal{R} = \{k_0 \le i \le k_1\}$. Follow the notations (N1) and (N3), and define

$$d_k = (n+1)D(g(\tfrac{k}{n}, b)), \quad k = k_0, ..., k_1,$$
$$c_k = \bar{F}_{\Gamma(k)}(d_{k_1}) - \sum_{j=1}^{k-1} \frac{d_{k_1-k+j}^j}{j!} c_{k-j}, \quad k = k_1, ..., 2, \text{and}$$
$$c_1 = \bar{F}_{\Gamma(1)}(d_{k_1}).$$

Under either H_0 or H_1, we have

$$P(S_{n,\mathcal{R}} \le b) = (1 + o(1)) \left(\bar{F}_{\Gamma(k_1)}(d_{k_1}) - \sum_{k=k_0}^{k_1-1} \frac{d_k^k}{d!} c_{k_1-k} \right).$$

Theorem 5.4 demands the same computational complexity as Theorem 5.1 does. However, it evidences that gamma approximation is a good choice under general settings of the gGOF statistics and hypotheses, since the formula is pretty accurate under finite n (see Section VI for numeric results). This result inspired us to apply gamma approximation for distribution calculation with further simplified formula.

Now we consider the second approximation strategy. Under stronger assumptions, in particular if $D(g(\frac{k}{n}, b))$ in (17) is a linear or near-linear function of k, we can provide a one-step formula for the distribution calculation. Starting with the exact linear case, Proposition 5.1 gives such a one-step formula that guarantees the same accuracy as Theorem 5.4 because both are based on the same gamma approximation.

Proposition 5.1: Consider a gGOF statistic in (11) with $\mathcal{R} = \{1 \le i \le k_1\}$ and $D(g(\frac{k}{n}, b)) = a + \lambda k$, for some $\lambda \ge 0$. Following notations (N3) and (N4), under either H_0 or H_1, we have

$$P(S_{n,\mathcal{R}} \le b) = (1 + o(1))e^{-a} \left(1 - \lambda + h_{k_1}(\lambda)\right).$$

One example that satisfies the linearity of $D(g(\frac{k}{n}, b))$ is the simple Kolmogorov-Smirnov (KS^+) statistic in (13) under H_0, where $a = -(n+1)b$ and $\lambda = \frac{n+1}{n}$. The following corollary summarizes this case.

Corollary 5.2: Consider the simple Kolmogorov-Smirnov statistic KS^+ in (13) with $\mathcal{R} = \{1 \le i \le k_1\}$. Following notations (N3) and (N4), for $b \le \frac{1}{n}$, we have that under H_0,

$$P(KS^+ \le b) = (1 + o(1))e^{(n+1)b} \left(-\frac{1}{n} + h_{k_1}(\frac{n+1}{n}) \right).$$

In general, the requirement of linear $D(g(\frac{k}{n}, b))$ is often too stringent. However, if $D(g(\frac{k}{n}, b))$ is close to linear, we can still simplify the calculation of Theorem 5.4. In particular, Theorem 5.5 below provides a sufficient condition on $D(g(\frac{k}{n}, b))$, under which the LS style asymptotics [23] can be extended to the gGOF family under general hypotheses. Again, we apply the gamma approximation, rather than the beta approximation used in the original LS paper. Gamma distribution has a simpler density function for easier technique proof in the generalization to the gGOF family. See the proof of Theorem 5.5 in Supplement for details.

Theorem 5.5: Consider any gGOF statistic in (11) with $\mathcal{R} = \{k_0 \le i \le k_1\}$. Follow notations (N1)–(N5), and define $d_k = (n+1)D(g(\frac{k}{n}, b))$, $d_k' = (n+1)\frac{d}{dx}D(g(\frac{k}{n}, b))$, and $k^* = \min\{k_1 - k, \sqrt{n}\}$. Assume $D(g(x, b))$ satisfies

1) $D(g(x, b)) < 1$ is increasing and convex in x for $\frac{k_0}{n} \le x \le \frac{k_1}{n}$;
2) $\frac{d}{dx}D(g(x, b)) < 1$; and
3) $D(g(k/n, b)) < \frac{k}{n+1}$, for $k > 1$ and large n.

Under either H_0 or H_1 in (3), we have

$$P(S_{n,\mathcal{R}} \ge b) = (1+o(1)) \sum_{k=k_0}^{k1} \left(1 - \frac{d_k'}{n} + h_{k^*}(\frac{d_k'}{n}) \right) f_{P(d_k)}(k).$$

This sufficient condition on $D(g(\frac{k}{n}, b))$ can be partially satisfied by HC^{2004} under H_0, for which $D(g(\frac{k}{n}, b)) = g(x, b)$ is given in (19). The result is officially stated in Corollary 5.3 below, which basically says that the condition is satisfied on the right tail when b is in the order of $O(\sqrt{n})$.

Corollary 5.3: Consider HC^{2004} statistic in (14) with $\mathcal{R} = \{k_0 \le i \le k_1\}$. Let $b_0 = \frac{b}{\sqrt{n}}$ be a positive constant $> 2x - 1$, $\frac{k_0}{n} < x < \frac{k_1}{n}$. Define

$$g(x, b_0) = \frac{1}{1 + b_0^2}[x + (b_0^2 - b_0\sqrt{b_0^2 + 4x(1-x)})/2],$$
$$g'(x, b_0) = \frac{1}{1 + b_0^2}[1 - \frac{b_0(1-2x)}{\sqrt{b_0^2 + 4x(1-x)}}].$$

Following the notation (N2), under H_0, we have

$$P(HC^{2004} \geq b) = (1+o(1)) \cdot \sum_{k=k_0}^{k1} \left(1 - g'(\frac{k}{n}, b_0) + h_{k^*}(g'(\frac{k}{n}, b_0))\right) f_{P(g(\frac{k}{n}, b_0)n)}(k).$$

The formula of Corollary 5.3 is different from that given in Li and Sigmund [23]. However, both formulae require the threshold $b = O(\sqrt{n})$. Thus, in theory both methods do not get the whole distribution. Also, the accuracy depends on the linear approximation of the $D(g(\frac{k}{n}, b))$ function, which is often hardly true under general H_1. Thus this type of calculation has a natural limitation for being utilized to calculate statistical power.

VI. Study of Statistical Power

In this section we first evidence the accuracy of our analytical methods by comparing the calculations with the Monte-Carlo simulations under various settings of H_0 and H_1. Then, by calculation we compare the finite-n performance of those optimal tests over various signal patterns. Unless specified otherwise, results reported below were based on truncation domain $\mathcal{R} = \{1 \leq i \leq n/2\}$ and the number of simulations was set at 5,000.

A. Calculation Accuracy

Our calculation methods can handle *i.i.d.* input statistics of arbitrary continuous distributions. In this section we evaluate how accurate our calculation methods are for constructing the distribution curves of HC^{2004} statistic, as an example of the gGOF, under various H_0 and H_1.

First, we calculate the null distribution of HC^{2004} statistic under general H_0 in (7). Fig. 4 shows the right-tail probability of the HC statistic over varying threshold b. Comparing with simulation (black solid curves), the exact calculation by Theorem 5.1 (cyan dashed curves) has a perfect match. The approximation by Theorem 5.4 is fairly accurate over the whole distribution too. The calculations by Li and Siegmund [23] (blue dotted curves) and by Corollary 5.3 (green dashed curves) can provide good approximation for the right tail, and thus can be used for calculating small test p-values at large threshold. Li and Siegmund's formula has a limitation for the left tail of the distribution; the formula of Corollary 5.3 provides a correction of a sort, which is preferred at small n but is more conservative at large n.

Now we assess the accuracy of calculating the alternative distribution of HC^{2004} statistic. Assume the input statistics were from a mixture model of either

$$H_1 : T_i \overset{\text{i.i.d.}}{\sim} (1-\epsilon)N(0,1) + \epsilon N(1,1),$$

or

$$H_1 : T_i \overset{\text{i.i.d.}}{\sim} (1-\epsilon)N(0,1) + \epsilon t_\nu.$$

The input p-values for the gGOF were obtained by $P_i = 1 - \Phi(T_i)$, i.e., under $H_0 : T_i \overset{\text{i.i.d.}}{\sim} N(0,1)$. These two alternatives can be roughly interpreted as that ϵ proportion of "signals" have either different means (i.e., $N(1,1)$) or different variances (i.e., the Student's t with degrees of freedom ν) when comparing with the "noise" (i.e., $N(0,1)$). Accordingly, Fig. 5 demonstrates the right-tail probability of HC^{2004} statistic (row 1: $\mu = 1$, $\epsilon = 0.1$; row 2: $\nu = 5$, $\epsilon = 0.5$). In both cases the exact calculation (by Theorem 5.1, cyan dashed curves) is perfect and the approximation (by Theorem 5.4, dot-dashed) is close to simulation (solid curves), with its accuracy increasing together with n.

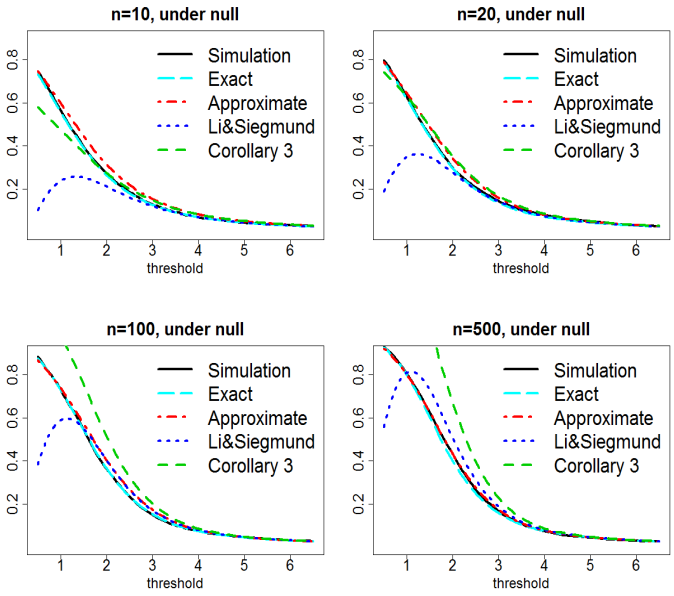


Fig. 4. Comparison among different calculations for the null distribution of HC^{2004}. Simulation: curves obtained by simulations; Exact: by Theorem 5.1; Approximate: by Theorem 5.4; Li&Siegmund: by [23]; Corollary 3: by Corollary 5.3.

Besides the normal distributions, we also assessed four non-normal settings studied in the initial paper of the HC [13]. The first setting regards a chi-squared model:

$$H_0 : T_i \overset{\text{i.i.d.}}{\sim} \chi^2_\nu(0), \text{ vs. } H_1 : T_i \overset{\text{i.i.d.}}{\sim} (1-\epsilon)\chi^2_\nu(0) + \epsilon\chi^2_\nu(\delta),$$

where ν is the degree of freedom, δ is the non-centrality parameter. The second setting is a Student's t mixture model:

$$H_0 : T_i \overset{\text{i.i.d.}}{\sim} t_\nu(0), \text{ vs. } H_1 : T_i \overset{\text{i.i.d.}}{\sim} (1-\epsilon)t_\nu(0) + \epsilon t_\nu(\delta).$$

The third setting is a chi-squared-exponential mixture model:

$$H_0 : T_i \overset{\text{i.i.d.}}{\sim} \exp(\nu), \text{ vs. } H_1 : T_i \overset{\text{i.i.d.}}{\sim} (1-\epsilon)\exp(\nu) + \epsilon\chi^2_\nu(\delta).$$

The fourth setting concerns a generalized normal distribution (also known as the power exponential distribution) model:

$$H_0 : T_i \overset{\text{i.i.d.}}{\sim} GN_p(0,\sigma), \quad \text{vs.}$$
$$H_1 : T_i \overset{\text{i.i.d.}}{\sim} (1-\epsilon)GN_p(0,\sigma) + \epsilon GN_p(\mu,\sigma),$$

where the probability density function of $GN_p(\mu,\sigma)$ is

$$p(x) = \frac{1}{C_p}\exp(-\frac{|x-\mu|^p}{p\sigma^p}), \quad C_p = 2p^{1/p}\Gamma(1+1/p)\sigma.$$

Notice that $GN_1(\mu,\sigma)$ is the Laplace distribution and $GN_2(\mu,\sigma)$ is $N(\mu,\sigma^2)$. Each row of Fig. 6 illustrates the

Fig. 5. The alternative distribution of HC^{2004} statistic under H_0 : $T_i \overset{\text{i.i.d.}}{\sim} N(0,1)$ vs. H_1 : $T_i \sim 0.9N(0,1) + 0.1N(1,1)$ (row 1), or H_1 : $T_i \sim 0.5N(0,1) + 0.5t_5$ (row 2). Column 1: $n = 10$; column 2: $n = 100$. Simulation: curves obtained by simulations; Exact: by Theorem 5.1; Approximate: by Theorem 5.4.

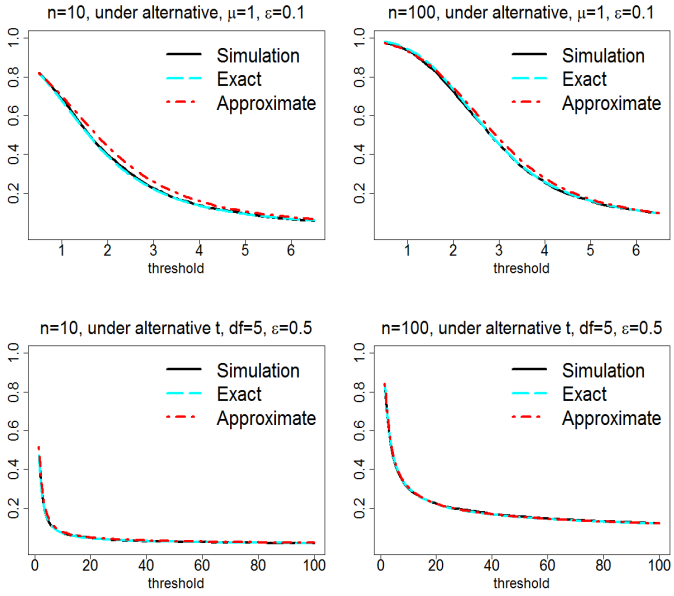
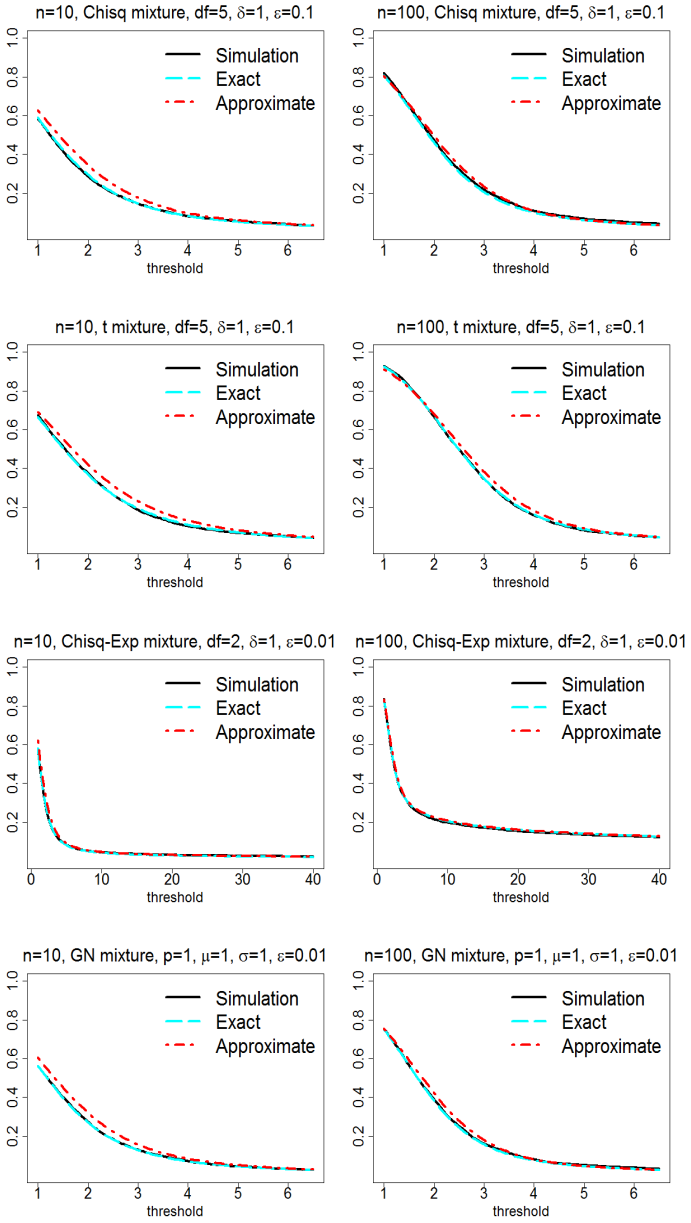


Fig. 6. The alternative distributions of HC^{2004} statistic under four non-Gaussian settings for H_0 and H_1. Column 1: $n = 10$; column 2: $n = 100$. Simulation: curves obtained by simulations; Exact: by Theorem 5.1; Approximate: by Theorem 5.4.



alternative distribution of HC^{2004} under each of the four settings for $n = 10$ (left column) or 100 (right column). Again, the exact calculation is perfect and the approximation is fairly accurate especially when n is large.

For the one-step calculation formula given by Proposition 5.1, the boundary is assumed linear: $D(g(\frac{k}{n}, b)) = a + \lambda k \geq 0$ in (20). One example is the KS^+ statistic in (13) under H_0. Fig. 7 demonstrates the accuracy of the calculation based on either a fixed slope $\lambda = 0.5$ or a fixed intercept $a = 0.5$. Here $k_0 = 1$, $k_1 = n = 50$. It shows that this gamma-approximation-based one-step formula performs well if the linearity of $D(g(\frac{i}{n}, b))$ is satisfied. As the boundary $a + \lambda k$ increases, the probabilities from both calculation and simulation decrease as expected.

B. *Performance of Optimal Tests Under Finite n*

As discussed in Section II, the asymptotic optimal methods for weak-sparse signals possess the same asymptotic property. It is of interest to know the performance of those statistics under finite n. Here we focus on the ϕ-divergence statistics defined in (16), which are asymptotically optimal for any $s \in [-1, 2]$ [18]. As discussed in Section II, the values of $s = 2, 1, 0, -1$ correspond to HC^{2004}, the Berk-Jones statistic, the reverse Berk-Jones statistic, and HC^{2008}, respectively. As shown below, these s values represent a spectrum of statistics with a trend of performance changes.

First, we show the accuracy of test p-value calculations in a similar way as that given in Li and Siegmund [23]. Specifically, for each gGOF statistic the thresholds at the significance levels of 10%, 5% and 1% were obtained through calculation (by Theorem 5.1). Then at these thresholds the empirical type I error rates were acquired through simulations (10,000 repetitions). As shown in Table I, the close match of the given significance levels and the obtained empirical type I error rates evidences that the calculations for the test p-values of these statistics are accurate. Not surprisingly, the accuracy by the approximated calculation of Li and Siegmund [23] requires relatively large n, whereas the calculation by Theorem 5.1 is exact and shall be perfectly accurate at any n.

Now through power calculation (again by Theorem 5.1), we can systematically compare the power of any gGOF statistics. To be consistent with literature, here we focus on the classic Gaussian mixture model in (4). With the type I error rate controlled at 5%, Fig. 8 provides the statistical power of

Fig. 7. Left-tail probability in (20) with a hypothetical linear boundary function $D(g(\frac{k}{n}, b)) = a + \lambda k$. Simulation: curves obtained by simulations; Approximation: by Proposition 5.1. Left panel: fix $\lambda = 0.5$ and vary a; right panel: fix $a = 0.5$ and vary λ.

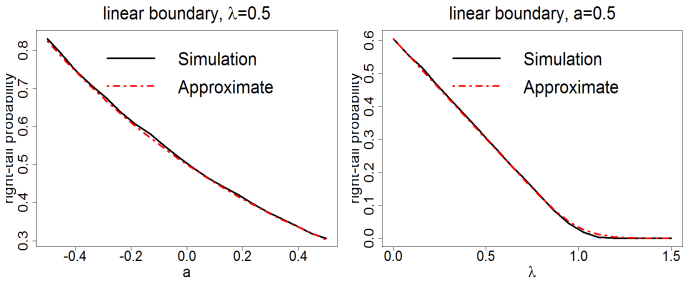


TABLE I
EMPIRICAL TYPE I ERROR RATES AT THE CALCULATED THRESHOLDS FOR THE SIGNIFICANCE LEVELS OF 10%, 5% AND 1%. HC^{2004}: $s = 2$; B-J: $s = 1$, REVERSE B-J: $s = 0$, AND HC^{2008}: $s = -1$.

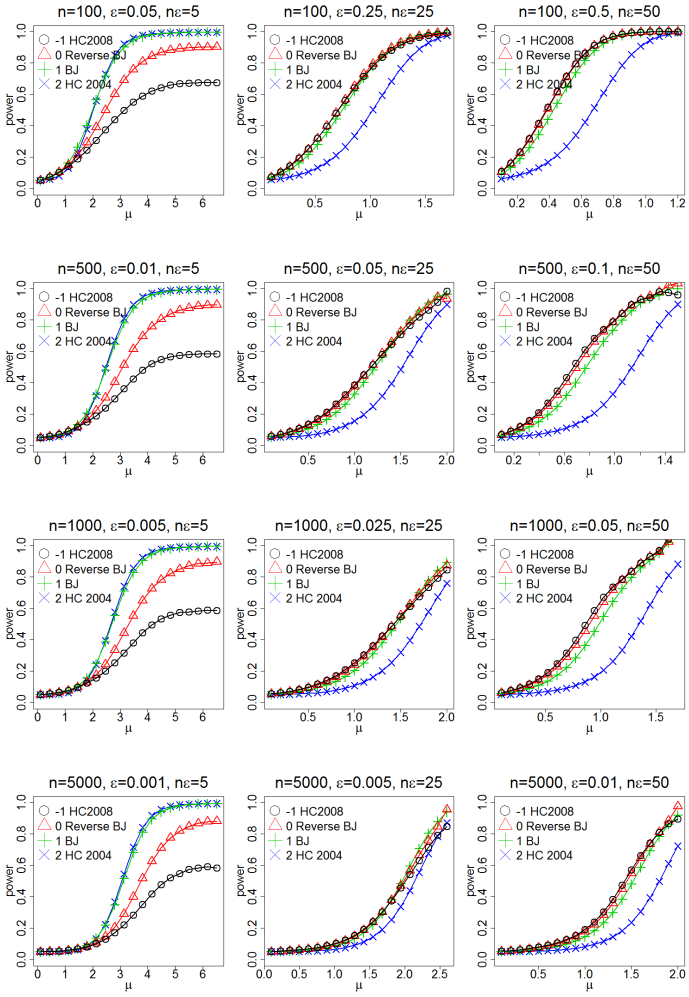
s	n	10%		5%		1%	
		Threshold	Emp. Err.	Threshold	Emp. Err.	Threshold	Emp. Err.
2	10	3.357	0.992	4.648	0.049	10.088	0.010
	50	3.507	0.102	4.714	0.050	10.102	0.011
	100	3.539	0.103	4.723	0.049	10.102	0.009
1	10	2.181	0.101	2.504	0.050	3.110	0.011
	50	2.408	0.098	2.716	0.048	3.300	0.010
	100	2.478	0.104	2.780	0.049	3.354	0.009
0	10	1.750	0.100	1.974	0.049	2.390	0.011
	50	2.040	0.101	2.301	0.047	2.803	0.011
	100	2.136	0.101	2.402	0.051	2.915	0.010
-1	10	1.618	0.098	1.838	0.051	2.227	0.009
	50	1.909	0.099	2.165	0.049	2.662	0.009
	100	2.010	0.107	2.271	0.052	2.777	0.010

HC^{2004}, B-J, reverse B-J, and HC^{2008} at various signal patterns represented by parameters (n, μ, ϵ). Figure S1 in the supplementary document provides further comparisons under the same settings except that the type I error rate is controlled at 0.5%. From both figures there are a few interesting observations on the relative behaviors of these statistics.

First, it seems that at finite n the average number $n\epsilon$ of signals is more relevant than the proportion ϵ of the signals. To see this point, note that columns 1–3 of the figure panels correspond to fixed signal numbers $n\epsilon = 5, 25, 50$, respectively; each column demonstrates a pattern of comparative performance among these four statistics. Meanwhile, the panels on the diagonal of the figure correspond to a fixed signal proportion $\epsilon = 0.05$, where the relative performances of the four statistics changed significantly over increased n. Similar observations can also be seen at fixed $\epsilon = 0.01$ or 0.005 but different n.

Second, considering the signal sparsity in terms of signal numbers, within the ϕ-divergence family, a bigger s value is related to better performance for sparser signals (e.g., HC^{2004} with $s = 2$ has the highest power in the first column), whereas a smaller s value is related to better performance for denser signals (e.g., HC^{2008} with $s = -1$ has the highest power in the third column). Note that the "reverse" versions of the

Fig. 8. Comparison of statistical power. HC^{2004}: $s = 2$; B-J: $s = 1$, reverse B-J: $s = 0$, and HC^{2008}: $s = -1$. Rows 1–4: $n = 100, 500, 1000, 5000$. Columns 1–3: $n\epsilon = 5, 25, 50$. Type I error rate: 5%.



statistics (i.e., $s = 0$ or -1 for reverse B-J or reverse HC) involve rescaling by i/n rather than by $P_{(i)}$ (comparing with the "original" B-J and HC). Therefore, they are less sensitive to signals with small p-values. Furthermore, their power could be "saturated" in detecting very sparse signals – the power stops increasing no matter how strong the signals are (see the first column of the figures). To understand this phenomenon, following (14) consider for example $HC_n^{2008} = \sup_i f_i$, where $f_i = \sqrt{n}\frac{i/n - P_{(i)}}{\sqrt{i/n(1-i/n)}}$. When signals are sparse and strong, the signals correspond to very small $P_{(i)}$'s for $i \ll n$, so that their $f_i \approx \sqrt{i}$ regardless the signal strength μ. For example, when the number of signals is $n\epsilon = 5$, their largest f_i is not far from $\sqrt{5} = 2.236$. In general, when n is large, $n\epsilon$ is small, and the type I error rate is stringently controlled, the power could be capped, even at a rather low level.

Third, regarding the range $[-1, 2]$ that corresponds to the asymptotic optimal ϕ-divergence tests, with $s = 1$ in the middle the B-J is not the best for sparser or denser signals. However, it is pretty robust over various μ, n and ϵ. This observation is consistent with the results of [23], [41].

Fig. 9. Statistical power along the ARW detection boundary (at type I error rate 5%).

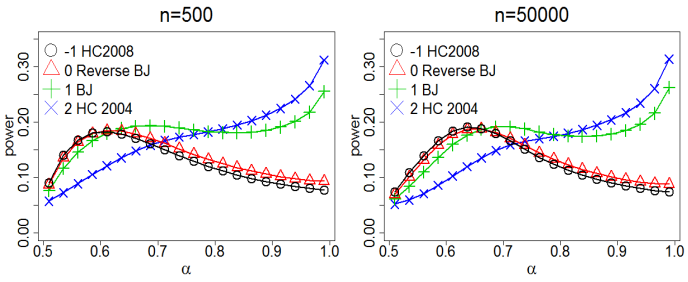
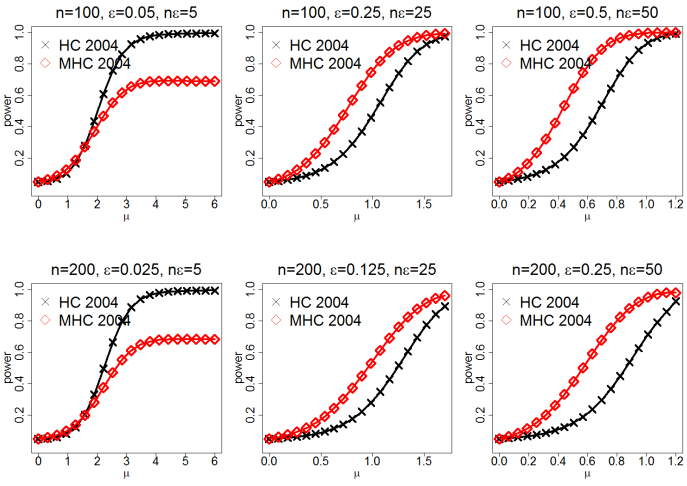


Fig. 10. Power comparison for the HC statistic with $\mathcal{R} = \{1 \le i \le n/2\}$ and the MHC statistic with $\mathcal{R} = \{1 \le i \le n/2, P_{(i)} \ge 1/n\}$. Type I error rate: 5%.



It is also of interest to compare the performance of these optimal methods along the asymptotic detection boundary given in (8) under finite n. As discussed in Section II, when $r < \rho^\star(\alpha)$, signals are too weak to be reliably detectable by any statistics. Whenever $r > \rho^\star(\alpha)$, all of these four optimal tests have asymptotically full power as $n \to \infty$. Thus, the area right above the detection boundary is a challenging scenario, and is where the optimal methods are prominent. Fig. 9 shows the statistical power of the four optimal methods over the sparsity parameter $\alpha \in (1/2, 1)$ and $r = \rho^\star(\alpha)$. It shows that the statistical power of these methods are in fact significantly different even for very large n. In consistence with Fig. 8, HC^{2008} and reverse B-J have similar power curve; they are more powerful for denser signals (i.e., at a smaller α corresponding to bigger $\epsilon = n^{-\alpha}$). HC^{2004} is more powerful for very sparse signals (i.e., at a larger α). B-J is not always the most powerful statistic, but again shows a more robust performance over all α values.

Last but not least, the truncation domain $\mathcal{R}$ in (12) is important to the performance of test statistics. In particular, as discussed in Introduction, the truncation based on $P_{(i)}$ could have extra benefit over the truncation based on i only [13], [23]. Here we compare HC^{2004} under $\mathcal{R} = \{1 \le i \le n/2\}$ with the MHC under $\mathcal{R} = \{1 \le i \le n/2, P_{(i)} \ge 1/n\}$. Fig. 10 shows that the MHC performs poorly when the number of signals is small, whereas it improves the performance when the number of signals increases. One reason is because $1/n$ is fairly large at finite n. By excluding input p-values less than $1/n$, MHC could easily miss those signal-representing input p-values, especially when there are just a few strong true signals. However, when signals are dense, the MHC is more powerful because (A) with high chance some signals (especially the weaker ones) will have input p-values larger than $1/n$, and (B) removing input p-values less than $1/n$ corrects the long-tail problem of the HC [13]. Thus, in practice when n is not too big, the original HC is still a better choice for relatively sparser and stronger signals, whereas MHC is better for denser and weaker signals.

VII. Data Analysis Framework for Application

In this section, we provide a framework for applying the gGOF tests and our analytical calculations in data analysis. The input p-values are obtained based on the generalized linear models (GLMs), a tool for broad applications. Specifically, with an appropriate link function, a GLM can be defined as

$$link(E(Y_k|\mathbf{X}_k, \mathbf{Z}_k)) = \mathbf{X}_k'\beta + \mathbf{Z}_k'\gamma, \tag{21}$$

where for the kth subject, $k = 1, ..., N$, Y_k denotes the response observation, $\mathbf{X}_k = (X_{k1}, ..., X_{kn})$ denotes a vector of the n targeting covariates, from which we want to test whether any signals exist. The vector $\mathbf{Z}_k = (Z_{k1}, ..., Z_{km})$ denotes m covariates controlling their effects to the response. The null hypothesis is that none of the targeting covariates are associated with the response, and therefore there are no signals:

$$H_0 : \beta_i = 0, i = 1, ..., n.$$

The nonzero β_i's under the alternative hypothesis represent signals. Many statistics can be used to test this null hypothesis. One classic example is a marginal test with statistics [42], [43]:

$$M_i = \sum_{k=1}^{N} X_{ki}(Y_k - \tilde{Y}_k), i = 1, ..., n,$$

where $\tilde{Y}_k$ is the fitted outcome value under H_0, which can be obtained by the least squares or the iteratively re-weighted least squares estimation. It can be shown that under H_0 the vector of the marginal statistics

$$\mathbf{M} = (M_1, ..., M_n) \xrightarrow{D} N(\mathbf{0}, \mathbf{\Sigma}),$$

as $N \to \infty$. The covariance matrix $\mathbf{\Sigma}$ can be estimated by

$$\hat{\mathbf{\Sigma}} = \mathbf{X}'\mathbf{W}\mathbf{X} - \mathbf{X}'\mathbf{W}\mathbf{Z}(\mathbf{Z}'\mathbf{W}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{W}\mathbf{X},$$

where matrices $\mathbf{X} = (X_{ki})$, $\mathbf{Z} = (Z_{ki})$, and $\mathbf{W}$ is the covariance matrix of Y. In the case of multiple regression model for quantitative traits, $\mathbf{W} = \hat{\sigma}^2\mathbf{I}$, where $\hat{\sigma}^2$ is the least squares estimate of the residual variance. In the case of logistic regression model for binary traits, $\mathbf{W} = diag\{\tilde{Y}_k(1-\tilde{Y}_k), k = 1, ..., N\}$.

We can de-correlate $\mathbf{M}$ to obtain the input statistics for the gGOF:

$$(T_1, ..., T_n) = \hat{\mathbf{\Sigma}}^{-\frac{1}{2}}\mathbf{M} \xrightarrow{D} N(\mathbf{0}, \mathbf{I}_{n\times n}),$$

and thus the input p-values asymptotically follow the null hypothesis in (7):

$$P_i = 2(1 - \Phi(|T_i|)) \overset{i.i.d.}{\rightarrow} \text{Uniform}[0, 1].$$

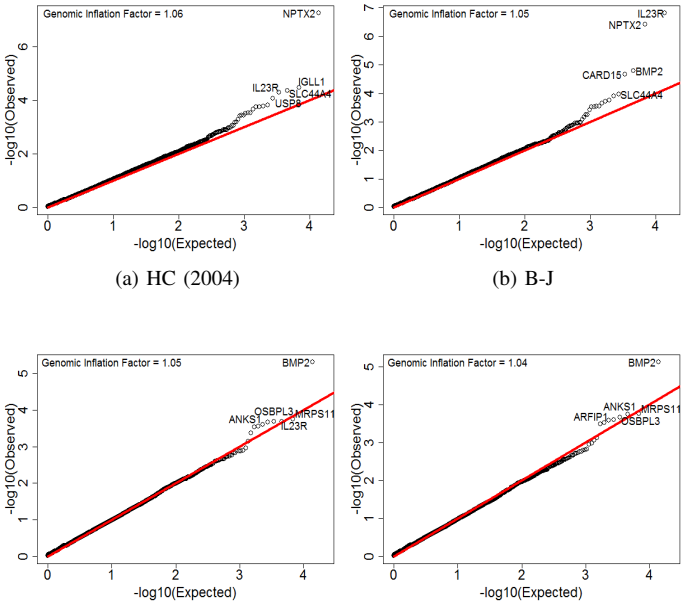
For any gGOF statistic, its test p-value can be calculated by the methods given in this paper for measuring how significant the data implies the existence of signals. It should be noted that the input statistics are not required to follow normal distribution; the calculation methods only requires that the input p-values are *i.i.d.* Uniform$[0, 1]$ under the null. That is, other input statistics following t or chi-squared distribution can be used as long as they are independent.

VIII. A GWAS of Crohn's Disease

Here we illustrate one application of the gGOF tests in detecting genetic signals of complex diseases. The GWAS tends to screen as many markers as possible, while true disease markers often have a relatively small number, and their genetic effects are often moderate to small [44]. Therefore, it is appealing to apply optimal tests for weak-sparse signals to detect novel disease genes. Here we focus on the gene-based SNP-set test. That is, the input p-values from SNPs within each gene form a gGOF statistic to test how significant the gene is associated. The same idea can be extended to SNP-set tests based on other meaningful genome segments, e.g., in pathway-based association studies [45]. Following the notations in (21), Y_k denotes the value of a phenotype (e.g., a quantitative trait or a binary diseases status), $\mathbf{X}_k$ denotes the genotype vector of the n SNPs in the gene to be tested, and $\mathbf{Z}_k$ denotes a vector of m environmental and/or other independent genetic factors as controlling covariates.

We applied the gene-based analysis framework to a GWAS data of Crohn's disease from NIDDK-IBDGC (National Institute of Diabetes, Digestive and Kidney Diseases - Inflammatory Bowel Disease Genetics Consortium) [46]. It contains 1,145 individuals from non-Jewish population (572 Crohn's disease cases and 573 controls). After typical quality control for the genotype data, 308,330 somatic SNPs were grouped into 15,857 genes according to their physical locations. As a special case of the GLM in (21) the logistic regression model was applied to search genes associated with Crohn's disease susceptibility. The controlling covariates $\mathbf{Z}_k = (1, Z_{k1}, Z_{k2})$ contain the intercept and the first two principal components of the genotype data, which serve the purpose of controlling potential population structure [47]. In case that a gene contains only one SNP, no gGOF tests were needed because the single input p-value represents the significance of that gene.

We examined four gGOF statistics: HC^{2004}, B-J, reverse B-J, and HC^{2008}. Fig. 11 gives the QQ plots of the gene-based test p-values calculated by Theorem 5.1. The genomic inflation factors (i.e., the ratios of empirical median of -log(p-values) vs. the expected median under H_0 [48]) are all close to 1, indicating that the genome-wide type I errors were well controlled. Among the four statistics, B-J seemed having higher power because it yielded more genes significantly above the diagonal line of the H_0-expected test p-values. Among the top ranked genes, many of them are relevant to Crohn's disease. In particular, *IL23R* and *CARD15* (also known as *NOD2*) are well-known Crohn's disease genes [49], [50], [46]. Gene *NPTX2* was top ranked by both HC^{2004} and B-J. It hasn't been reported previously through association studies, but could be a putative disease gene because it encodes a neuronal petraxin, which is related to C-reactive protein [51], an indicator for Crohn's disease activity level [52]. Furthermore, *NPTX2* has an important paralog gene *APCS* (www.genecards.org), which is related to arthritis, a disease highly correlated with Crohn's disease [53]. Gene *SLC44A4* is also related to the pathophysiology of Crohn's disease. Defects in this gene can cause sialidosis [51], a lysosomal storage disease due to a deficiency of sialidase, an enzyme important for various cells to defend against infection [54]. Gene *BMP2* was identified by B-J, reverse B-J, and HC^{2008}. This gene could also be relevant because it is associated with digestive phenotypes, especially colon cancer [55], [56]. Certainly, further studies are needed to validate those top ranked genes.



(a) HC (2004) (b) B-J

(c) reverse B-J (d) reverse HC (2008)

Fig. 11. The QQ plots of the calculated gene association test p-values of four weak-sparse-optimal tests: HC (2004), B-J, reverse B-J, and reverse HC (2008).

IX. Discussion

This paper proposed a generic gGOF family and provided techniques to calculate their null and alternative distributions in the finite n case. The methodology gave a foundation for applying and comparing existing important test statistics, especially those optimal statistics for detecting weak-and-sparse signals. It also gave a framework for developing new

test statistics to address specific signal patterns in scientific and engineering fields of signal detection.

The gGOF is a broad family of supremum based one-sided test statistics, which only requires that the function $f(\frac{i}{n}, y)$ in (11) is strictly decreasing in y at each i and n. For a fixed n in analyzing a given data, we can write $f(\frac{i}{n}, y) = f_i(y)$, i.e., $f_i(y)$ is strictly decreasing in y at each i. Following this essence, it is easy to see that the gGOF covers the one-sided "exact Berk-Jones" statistic M_n^+ (cf. equation (1.9) in [17], or equations (3.1) and (3.2) in [22]), because it is equivalent to a gGOF statistic with $f_i(P_{(i)}) = \bar{F}_{B(i,n-i+1)}(P_{(i)})$. Therefore, the distributions of M_n^+ under both H_0 and H_1 follow (20), and our methods can be used to calculate its p-value and power.

Our study on the relative performance of the gGOF statistics in finite case is related to a few papers [57], [33], [22]. These papers compared the HC and the exact B-J statistics from the multi-hypothesis testing perspective. In particular, using the "local levels" (i.e., the local significance) of $P_{(i)}$ over each $i = 1, ..., n$, it is shown that the exact B-J gives each $P_{(i)}$ equal local significance level, whereas the HC weights heavier on smaller $P_{(i)}$ under finite n. This interesting observation is consistent with our results that HC is more powerful to detect very sparse signals (where signals most likely correspond to the smallest $P_{(i)}$'s), whereas the B-J (as an approximation of the exact B-J [17]) is more powerful for denser signals (where signals could show in $P_{(i)}$ with larger i). Since the gGOF covers any statistics in the form of $\sup_i f_i(P_{(i)})$, the local-level study could be extended to the gGOF statistics. However, in this paper we focused on the global hypothesis testing problem, for which we directly calculate the statistical power rather than using indirect criteria such as local levels.

The problem of global hypothesis testing in finite case is also quite relevant to signal detection from an engineering viewpoint. For example, it has been shown that ordering observations provides a new perspective on small sample signal detection [58], [59], [60], [61]. In these applications, although the goal is mainly to achieve energy-efficient signal detection in sensor networks, their results are interesting in finite sample analysis. In particular, the design on censoring the ordered transmissions [58], [59] can be considered as an engineering realization of p-value truncations. In this sense, the "one-bit detection" strategy [60], [61] is equivalent to the minimal p-value method (because the maximal-magnitude based decision statistic corresponds to the minimal two-sided p-value). Meanwhile, considering some statistics in the gGOF family (e.g., the HC and the B-J) could have higher power than the minimal p-value method for certain weak-and-rare signals [13], we hypothesize that gGOF based new strategies could be further developed to improve the performance of censor networks. Certainly it requires that signal transmission cost is not forbiddingly high, so that it is affordable to look at large statistics from a few sensors before making decision.

The study in this paper can be further improved. In particular, in real data analysis input statistics are often correlated. It would be nice to incorporate such correlation into the calculation of test p-values and statistical power. For this purpose, we will report the results we have gotten in a separate paper.

APPENDIX A

The supplementary document provides proof of the theorems and supportive lemmas, and supplementary figures.

REFERENCES

[1] E. Arias-Castro, D. L. Donoho, and X. Huo, "Near-optimal detection of geometric objects by fast multiscale methods," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2402–2425, 2005.

[2] N. Kundargi, Y. Liu, and A. Tewfik, "A framework for inference using goodness of fit tests based on ensemble of phi-divergences," *IEEE Transactions on Signal Processing*, vol. 61, no. 4, pp. 945–955, 2013.

[3] S. Paris, D. Mary, and A. Ferrari, "Detection tests using sparse models, with application to hyperspectral data," *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1481–1494, 2013.

[4] S. Sulis, D. Mary, and L. Bigot, "A study of periodograms standardized using training datasets and application to exoplanet detection," *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 2136–2150, 2017.

[5] R. Bacher, C. Meillier, F. Chatelain, and O. Michel, "Robust control of varying weak hyperspectral target detection with sparse nonnegative representation," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3538–3550, 2017.

[6] J. G. Ligo, G. V. Moustakides, and V. V. Veeravalli, "Sparse gaussian mixture detection: Low complexity, high performance tests via quantization," in *Information Theory (ISIT), 2017 IEEE International Symposium on*, pp. 1277–1281, IEEE, 2017.

[7] L. Zhang, A. A. Ding, F. Durvaux, F.-X. Standaert, and Y. Fei, "Towards sound and optimal leakage detection procedure.," *IACR Cryptology ePrint Archive*, vol. 2017, p. 287, 2017.

[8] R. F. R. Suleiman, D. Mary, and A. Ferrari, "Dimension reduction for hypothesis testing in worst-case scenarios," *IEEE Transactions on Signal Processing*, vol. 62, no. 22, pp. 5973–5986, 2014.

[9] R. F. R. Suleiman, D. Mary, and A. Ferrari, "Minimax sparse detection based on one-class classifiers," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pp. 5553–5557, IEEE, 2013.

[10] J. Haupt, R. M. Castro, and R. Nowak, "Distilled sensing: Adaptive sampling for sparse detection and estimation," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 6222–6235, 2011.

[11] S. Paris, R. F. R. Suleiman, D. Mary, and A. Ferrari, "Constrained likelihood ratios for detecting sparse signals in highly noisy 3d data," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pp. 3947–3951, IEEE, 2013.

[12] J. Jin, J.-L. Starck, D. L. Donoho, N. Aghanim, and O. Forni, "Cosmological non-gaussian signature detection: Comparing performance of different statistical tests," *EURASIP Journal on Advances in Signal Processing*, vol. 2005, no. 15, p. 297184, 2005.

[13] D. L. Donoho and J. Jin, "Higher criticism for detecting sparse heterogeneous mixtures," *The Annals of Statistics*, vol. 32, no. 3, pp. 962–994, 2004.

[14] D. L. Donoho and J. Jin, "Higher criticism thresholding: Optimal feature selection when useful features are rare and weak," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, pp. 14790–14795, Sep 30 2008.

[15] Z. Wu, Y. Sun, S. He, J. Cho, H. Zhao, and J. Jin, "Detection boundary and Higher Criticism approach for sparse and weak genetic effects.," *The Annals of Applied Statistics*, vol. 8, no. 2, pp. 824–851, 2014.

[16] I. Barnett, R. Mukherjee, and X. Lin, "The generalized higher criticism for testing snp-set effects in genetic association studies," *Journal of the American Statistical Association*, vol. 112, no. 517, pp. 64–76, 2017.

[17] R. H. Berk and D. H. Jones, "Goodness-of-fit test statistics that dominate the kolmogorov statistics," *Probability Theory and Related Fields*, vol. 47, no. 1, pp. 47–59, 1979.

[18] L. Jager and J. A. Wellner, "Goodness-of-fit tests via phi-divergences," *The Annals of Statistics*, vol. 35, no. 5, pp. 2018–2053, 2007.

[19] E. Arias-Castro, E. J. Candès, and Y. Plan, "Global testing under sparse alternatives: Anova, multiple comparisons and the higher criticism," *The Annals of Statistics*, vol. 39, no. 5, pp. 2533–2556, 2011.

[20] T. T. Cai and Y. Wu, "Optimal detection of sparse mixtures against a given null distribution," *IEEE Transactions on Information Theory*, vol. 60, no. 4, pp. 2217–2232, 2014.

[21] M. Denuit, C. Lefèvre, and P. Picard, "Polynomial structures in order statistics distributions," *Journal of Statistical Planning and Inference*, vol. 113, no. 1, pp. 151–178, 2003.

[22] A. Moscovich, B. Nadler, and C. Spiegelman, "On the exact berk-jones statistics and their p-value calculation," *Electronic Journal of Statistics*, vol. 10, no. 2, pp. 2329–2354, 2016.

[23] J. Li and D. Siegmund, "Higher criticism: p-values and criticism," *The Annals of Statistics*, vol. 43, no. 3, pp. 1323–1350, 2015.

[24] J. Hoh, A. Wille, and J. Ott, "Trimming, weighting, and grouping SNPs in human case-control association studies," *Genome Research*, vol. 11, no. 12, pp. 2115–2119, 2001.

[25] Y. I. Ingster, "Some problems of hypothesis testing leading to infinitely divisible distributions," *Mathematical Methods of Statistics*, vol. 6, no. 1, pp. 47–69, 1997.

[26] Y. I. Ingster, "Minimax detection of a signal for i^n-balls," *Mathematical Methods of Statistics*, vol. 7, no. 4, pp. 401–428, 1998.

[27] J. Tukey, "The higher criticism." Course Notes, Statistics 411, Princeton University, 1976.

[28] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes," *The Annals of Mathematical Statistics*, pp. 193–212, 1952.

[29] I. J. Barnett and X. Lin, "Analytical p-value calculation for the higher criticism test in finite-d problems," *Biometrika*, vol. 101, no. 4, pp. 964–970, 2014.

[30] D. L. Donoho and J. Jin, "Higher criticism for large-scale inference: especially for rare and weak effects," *Statistical Science*, vol. 30, no. 1, pp. 1–25 DOI: 10.1214/14–STS506, 2015.

[31] J. Shao, *Mathematical Statistics*. Springer Verlag, 2010.

[32] L. Jager and J. A. Wellner, "A new goodness of fit test: the reversed berk-jones statistic," 2004.

[33] V. Gontscharuk, S. Landwehr, H. Finner, *et al.*, "Goodness of fit tests in terms of local levels with special emphasis on higher criticism tests," *Bernoulli*, vol. 22, no. 3, pp. 1331–1363, 2016.

[34] M. Noé, "The calculation of distributions of two-sided kolmogorov-smirnov type statistics," *The Annals of Mathematical Statistics*, pp. 58–64, 1972.

[35] M. Noé and G. Vandewiele, "The calculation of distributions of kolmogorov-smirnov type statistics including a table of significance points for a particular case," *The Annals of Mathematical Statistics*, vol. 39, no. 1, pp. 233–241, 1968.

[36] V. Kotel'Nikova and E. Chmaladze, "On computing the probability of an empirical process not crossing a curvilinear boundary," *Theory of Probability & Its Applications*, vol. 27, no. 3, pp. 640–648, 1983.

[37] G. R. Shorack and J. A. Wellner, *Empirical processes with applications to statistics*. SIAM, 2009.

[38] G. Steck, "The smirnov two sample tests as rank tests," *The Annals of Mathematical Statistics*, pp. 1449–1466, 1969.

[39] M. Breth, "On a recurrence of steck," *Journal of Applied Probability*, pp. 823–825, 1976.

[40] H. Ruben, "On the evaluation of steck's determinant for rectangle probabilities of uniform order statistics," *Communications in Statistics-Theory and Methods*, vol. 5, no. 6, pp. 535–543, 1976.

[41] G. Walther *et al.*, "The average likelihood ratio for large-scale multiple testing and detecting sparse mixtures," in *From Probability to Statistics and Back: High-Dimensional Models and Processes–A Festschrift in Honor of Jon A. Wellner*, pp. 317–326, Institute of Mathematical Statistics, 2013.

[42] P. McCullagh and J. A. Nelder, *Generalized Linear Models*. Florida: CRC Press LLC, 2nd ed., 1989.

[43] D. J. Schaid, C. M. Rowland, D. E. Tines, R. M. Jacobson, and G. A. Poland, "Score tests for association between traits and haplotypes when linkage phase is ambiguous," *The American Journal of Human Genetics*, vol. 70, no. 2, pp. 425–434, 2002.

[44] D. B. Goldstein, "Common genetic variation and human traits," *New England Journal of Medicine*, vol. 360, no. 17, pp. 1696–1698, 2009.

[45] L. Luo, G. Peng, Y. Zhu, H. Dong, C. I. Amos, and M. Xiong, "Genome-wide gene and pathway analysis," *European Journal of Human Genetics*, vol. 18, no. 9, pp. 1045–1053, 2010.

[46] R. Duerr, K. Taylor, S. Brant, J. Rioux, M. Silverberg, M. Daly, A. Steinhart, C. Abraham, M. Regueiro, A. Griffiths, *et al.*, "A genome–wide association study identifies IL23R as an inflammatory bowel disease gene," *Science Signalling*, vol. 314, no. 5804, p. 1461, 2006.

[47] A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich, "Principal components analysis corrects for stratification in genome-wide association studies," *Nature Genetics*, vol. 38, no. 8, pp. 904–909, 2006.

[48] J. Yang, M. N. Weedon, S. Purcell, G. Lettre, K. Estrada, C. J. Willer, A. V. Smith, E. Ingelsson, J. R. O'Connell, M. Mangino, R. Magi, P. A. Madden, A. C. Heath, D. R. Nyholt, N. G. Martin, G. W. Montgomery, T. M. Frayling, J. N. Hirschhorn, M. I. McCarthy, M. E. Goddard, P. M. Visscher, and the GIANT Consortium, "Genomic inflation factors under polygenic inheritance," *European Journal of Human Genetics*, vol. 19, no. 7, pp. 807–812, 2011.

[49] J.-P. Hugot, M. Chamaillard, H. Zouali, S. Lesage, J.-P. Cézard, J. Belaiche, S. Almer, C. Tysk, C. A. O'Morain, M. Gassull, *et al.*, "Association of nod2 leucine-rich repeat variants with susceptibility to crohn's disease," *Nature*, vol. 411, no. 6837, pp. 599–603, 2001.

[50] Y. Ogura, D. K. Bonen, N. Inohara, D. L. Nicolae, F. F. Chen, R. Ramos, H. Britton, T. Moran, R. Karaliuskas, R. H. Duerr, *et al.*, "A frameshift mutation in nod2 associated with susceptibility to crohn's disease," *Nature*, vol. 411, no. 6837, pp. 603–606, 2001.

[51] D. Maglott, J. Ostell, K. D. Pruitt, and T. Tatusova, "Entrez gene: Gene-centered information at NCBI," *Nucleic Acids Research*, vol. 39, no. suppl 1, pp. D52–D57, 2011.

[52] P. Chamouard, Z. Richert, N. Meyer, G. Rahmi, and R. Baumann, "Diagnostic value of c-reactive protein for predicting activity level of crohn's disease," *Clinical Gastroenterology and Hepatology*, vol. 4, no. 7, pp. 882–887, 2006.

[53] G. Trikudanathan, P. G. Venkatesh, and U. Navaneethan, "Diagnosis and therapeutic management of extra-intestinal manifestations of inflammatory bowel disease," *Drugs*, vol. 72, no. 18, pp. 2333–2349, 2012.

[54] W. D. James, T. Berger, and D. Elston, *Andrew's diseases of the skin: clinical dermatology*. Elsevier Health Sciences, 2011.

[55] S. Yuvaraj, S. H. Al-Lahham, R. Somasundaram, P. A. Figaroa, M. P. Peppelenbosch, and N. A. Bos, "E. coli-produced BMP-2 as a chemo-preventive strategy for colon cancer: a proof-of-concept study," *Gastroenterology Research and Practice*, vol. 2012, 2012.

[56] M. L. Slattery, A. Lundgreen, J. S. Herrick, S. Kadlubar, B. J. Caan, J. D. Potter, and R. K. Wolff, "Genetic variation in bone morphogenetic protein and colon and rectal cancer," *International Journal of Cancer*, vol. 130, no. 3, pp. 653–664, 2012.

[57] D. Mary and A. Ferrari, "A non-asymptotic standardization of binomial counts in higher criticism," in *2014 IEEE International Symposium on Information Theory*, pp. 561–565, IEEE, 2014.

[58] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: A low-communication-rate scheme for distributed detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 2, pp. 554–568, 1996.

[59] R. S. Blum and B. M. Sadler, "Energy efficient signal detection in sensor networks using ordered transmissions," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3229–3235, 2008.

[60] P. Braca, S. Marano, and V. Matta, "Asymptotically consistent one-bit detection in large sensor networks," in *2011 19th European Signal Processing Conference*, pp. 1035–1039, IEEE, 2011.

[61] P. Braca, S. Marano, and V. Matta, "Single-transmission distributed detection via order statistics," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 2042–2048, 2011.