Statistica Sinica **32** (2022), 323-344
doi:https://doi.org/10.5705/ss.202020.0048

CLT FOR U-STATISTICS WITH GROWING DIMENSION

Cyrus DiCiccio and Joseph Romano

LinkedIn Corporation and Stanford University

Abstract: We present a general triangular array central limit theorem for U-statistics, where the kernel $h_k(x_1, \ldots, x_k)$ and its dimension k may increase with the sample size. Motivating examples that require such a general result are presented, including a class of Hodges–Lehmann estimators, subsampling estimators, and combining p-values using data splitting. A result for the so-called M-statistic is also presented, which is defined as the median of some kernel computed over all subsets of the data of a given size. The conditions in the theorems are verified in the motivating examples as well.

Key words and phrases: Data splitting, Hodges–Lehmann estimator, hypothesis testing, P-values, subsampling, U-statistics.

1. Introduction

Suppose $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) according to a distribution P. Consider the U-statistic

$$U_n(X_1, \ldots, X_n) = \binom{n}{k}^{-1} \sum h_k(X_{i_1}, \ldots, X_{i_k}) \ , \tag{1.1}$$

where h_k is a symmetric kernel of order $k = k_n$ (which may increase with n), and the sum is taken over all $\binom{n}{k}$ combinations of k observations taken from the sample. We specifically allow the order $k = k_n$ of the kernel h_{k_n} to depend on n, as does the kernel itself. For cleaner notation, we may just write k and h_k rather than k_n and h_{k_n}, respectively. However, we will also allow k to be fixed and $k \to \infty$ as $n \to \infty$.

The asymptotic theory of U-statistics was developed in a landmark paper by Hoeffding (1948). The classical result assumes the kernel is fixed, $n \to \infty$, and P is fixed. This study provides general conditions to show asymptotic normality in a triangular array setup. This generality allows the kernel and its order to vary with the sample size, as necessitated by certain applications, such as the examples presented here. A uniform in P result is given in Romano and Shaikh

Corresponding author: Joseph Romano, Departments of Statistics and Economics, Stanford University, Stanford, CA 94305, USA. E-mail: romano@stanford.edu.

(2012), where the kernel is fixed.

When k is allowed to vary with n, such that $k = k_n \to \infty$ as $n \to \infty$, sufficient conditions for the asymptotic normality of such U-statistics appear in Mentch and Hooker (2016), who consider inference for random forests. Unfortunately, as noted in Song, Chen and Kato (2019), their conditions never hold, because they assume conditions that cannot hold simultaneously (as will be explained later). We provide rigorous sufficient conditions, which are shown to hold in a variety of examples. Alternative conditions appear in Peng, Coleman and Mentch (2019) (Theorem 1) and Zhou, Mentch and Hooker (2019) (Theorems 3 and 4). Another way to show asymptotic normality with an increasing kernel order is to appeal to the Berry–Esseen bounds for U-statistics, as in van Zwet (1984), Chen and Shao (2007), and Song, Chen and Kato (2019). Although these results provide bounds on a normal approximation, they impose higher moment assumptions; see also Song, Chen and Kato (2019) (Remark 2.3) for additional references.

As an alternative to U_n in (1.1), we also consider the median statistic M_n, defined by

$$M_n(X_1, \ldots, X_n) = \text{median}\left\{h_k(X_{i_1}, \ldots, X_{i_k})\right\} \ , \tag{1.2}$$

where the median is taken over all $\binom{n}{k}$ combinations of k observations taken from the sample. (In this case, we may also allow the kernel to be asymmetric.)

The remainder of the paper is organized as follows. Section 2 presents four motivating examples for the results obtained. The main theorems are given in Section 3. The examples are revisited in Section 4, where the conditions are verified. Section 5 concludes the paper. All proofs are deferred to Section 6.

2. Motivating Examples

In this section, we provide examples to motivate the need for a general result.

Example 1. [Maximin Tests] Assume $X_1, \ldots, X_n$ are independent (but not necessarily i.i.d.) normal, with $X_i \sim N(\mu_i, 1)$. The problem is to test the null hypothesis H_0 that all μ_i are equal to zero against the (multi-directional) alternative that not all μ_i are zero. Of course, for this problem, there is no uniformly most powerful (UMP) level-α test, but there is a uniformly most powerful invariant (UMPI) level-α test, which rejects for large values of $\sum_{i=1}^{n} X_i^2$. However, if we believe the indices i for which $\mu_i \neq 0$ are sparse, we can outperform the UMPI test; see Arias-Castro, Candès and Plan (2011). Indeed, we may wish to direct power against alternatives for which there are not too many nonzero μ_i. We can formulate the problem as follows. Fix $\epsilon = \epsilon_n > 0$ and $k = k_n$, and determine the maximin level-α test against alternatives where at least k of the μ_i satisfy

$\mu_i \geq \epsilon$. (Note that we can treat the case where these alternatives may satisfy $|\mu_i| \geq \epsilon$ similarly. However, for expository reasons, we focus on the case of positive alternatives.) We apply standard arguments to determine the maximin test, as in Lehmann and Romano (2005). Intuitively, the least favorable distribution places equal mass at the $\binom{n}{k}$ points in the alternative parameter space, where each point $(\mu_1, \ldots, \mu_n)$ satisfies that exactly k of the components are ϵ, and the rest are zero. For this choice of a least favorable distribution, the conditions of Theorem 8.1.1. in Lehmann and Romano (2005) hold, and the maximin test rejects for large values of the U-statistic given in (1.1), where

$$h_k(X_1, \ldots, X_k) = \exp\left(\epsilon \sum_{i=1}^{k} X_i\right). \tag{2.1}$$

We wish to examine the asymptotic behavior of this test statistic (both for setting critical values and approximating the power) in situations where possibly $k \to \infty$ and/or $\epsilon \to 0$ (as well as letting the data distribution vary at time n).

Example 2. [Class of Hodges–Lehmann Estimators] Suppose $X_1, \ldots, X_n$ are i.i.d. according to a symmetric distribution on the real line. Based on robustness considerations, the classical Hodges–Lehmann estimator is defined as the median of all pairwise averages of the observations. Evidently, the Hodges–Lehmann estimator is an M-statistic (1.2) (with $k = 2$). More generally, consider the statistic (1.2) with

$$h_k(X_1, \ldots, X_k) = k^{-1/2} \sum_{i=1}^{k} X_i \ .$$

Let $\hat{\theta}_{n,k} = k^{-1/2} M_n$. (Note that we could have equivalently defined the kernel with $k^{-1/2}$ replaced by k^{-1}, such that the estimator is just M_n. However, it is convenient for the purpose of applying our results to define the kernel as above, so that it is of order one in probability.) As k varies, one might consider this class of estimators as k ranges from $k = 1$ (the usual sample median) to $k = n$ (the sample mean), with the choice balancing efficiency and robustness considerations. The purpose here is to provide a limit theorem for general k, while allowing k to increase with n.

Example 3. [Subsampling Distribution] Suppose $X_1, \ldots, X_n$ are i.i.d. P, and we are interested in a real-valued parameter $\xi(P)$. Assume $\hat{\xi}_n = \hat{\xi}_n(X_1, \ldots, X_n)$ is an estimator of $\xi(P)$. Fix $1 < k < n$, and let $S_1, \ldots, S_N$ be the $N = \binom{n}{k}$ subsets of size k taken without replacement from the data, ordered in any fashion. For a given hypothesized value of ξ, say ξ_0, let $J_n(t, P)$ be the true c.d.f. of $\tau_n(\hat{\xi}_n - \xi_0)$, evaluated at some generic t. Typically, $\tau_n = \sqrt{n}$. Then, a subsampling estimator of $J_n(t,P)$ is given by

$$U_n(t) = \frac{1}{N}\sum_{i=1}^{N} I\left\{\tau_k(\hat{\xi}_k(S_i) - \xi_0) \le t\right\} \tag{2.2}$$

(The usual subsampling estimator replaces ξ_0 in (2.2) with $\hat{\xi}_n$, though both are relevant, depending on the ultimate goal; see Chapter 2 in Politis, Romano and Wolf (1999).) Evidently, for each t, $U_n(t)$ is a U-statistic of degree k. In order to consistently estimate the true distribution $J_n(t,P)$, it is generally required that $k \to \infty$. Rather than consistency, we would like to determine the limiting distribution of $U_n(t) - J_n(t,P)$, appropriately normalized.

Example 4. [Combining p-values Using Data Splitting] Data splitting involves partitioning a data set into disjoint "splits" or subsamples, which can then be used for various statistical tasks. Typically, one portion of the data is used for some form of selection (such as model fitting, dimension reduction, or choice of tuning parameters), and then a second, independent portion of the data is used for some further purpose, such as estimation or model fitting. In addition, data splitting can be used in prediction to assess a model's performance (where a portion of the data is used to select and/or fit the model, and the remainder is used to assess its performance), or in inference to perform tests of significance after hypotheses or test statistics have been selected. Data splitting has become a useful remedy for data snooping (giving a valid inference after the selection of a hypothesis), estimating nuisance parameters, and avoiding over-fitting in prediction problems. The main complaint about data splitting is that the choice of split is arbitrary (and random). Furthermore, the resulting inference violates the sufficiency principle, which states that inference in i.i.d. problems should be invariant with respect to ordering. Recent methods have proposed combining p-values over multiple splits of the data; see Ruschendorf (1982), Meinshausen, Meier and Bühlmann (2009), Vovk and Wang (2012), and DiCiccio, DiCiccio and Romano (2020). For example, if $\hat{p}_{n,i}$ is a p-value computed over some subsample S_i of the data, then we can combine these p-values by taking their average $\bar{p}_n$ (which is a U-statistic) or their median. Conservative methods that control the probability of a type-1 error at level α compare the average p-value or median p-value with $\alpha/2$. These methods are conservative in nature in that the resulting rejection probability is significantly below the desired nominal level. The proposed method improves upon these methods by exploiting the U-statistic nature of the average of the p-values.

3. Main Results

In this section, the main asymptotic normality theorem is developed for U-statistics with a growing kernel order, as well as for the corresponding M-statistic.

3.1. A general U-statistic CLT under growing kernel order

Suppose $X_1, \ldots, X_n$ are i.i.d. P. Consider the U-statistic given in (1.1), where h_k is assumed to be a symmetric kernel of order $k = k_n$, and the sum is taken over all $\binom{n}{k}$ combinations of k observations taken from the sample. We specifically allow the order $k = k_n$ of the kernel h_{k_n} to depend on n, as does the kernel itself. For cleaner notation, we may write k and h_k rather than k_n and h_{k_n}, respectively, but we allow k to be fixed and $k \to \infty$ as $n \to \infty$. (Note that if h_k is not symmetric in its arguments, it can always be symmetrized by further averaging. Thus, for the purposes of the CLT, we assume h_k is symmetric.)

Define $\theta_k = E(h_k(X_1, \ldots, X_k))$ and

$$\zeta_{1,k} = \text{Var}(h_{1,k}(X)) \ ,$$

where

$$h_{1,k}(x) = E(h_k(x, X_2, \ldots, X_k)) - \theta_k \ .$$

All expectations and variances are computed under the probability distribution P generating the data, noting that $P = P_n$ may also vary with n.

More generally, for $1 \leq c \leq k$, define

$$h_{c,k}(X_1, \ldots, X_c) = E[h_k(X_1, \ldots, X_k)|X_1, \ldots, X_c] - \theta_k$$

and

$$\zeta_{c,k} = \text{Var}(h_{c,k}(X_1, \ldots, X_c)) \ , \tag{3.1}$$

so that $\zeta_{k,k}$ is the variance of the kernel, based on a sample of size k equal to the order of the kernel.

Sufficient conditions for the asymptotic normality of such U-statistics are given in Mentch and Hooker (2016), but their result is not valid because their conditions can never hold simultaneously. In particular, they assume $\zeta_{1,k} \nrightarrow 0$, which fails in our applications. Moreover, they assume that the second moment of the kernel is uniformly bounded, so that $\zeta_{k,k} \leq C < \infty$. However, by Theorem 1 in Hoeffding (1948), it follows that $\zeta_{1,k} \leq \zeta_{k,k}/k \leq C/k \to 0$. Therefore, the conditions $\zeta_{k,k} \leq C$ and $\zeta_{1,k} \nrightarrow 0$ are incompatible, and thus the conditions in their theorem can never apply.

In some of our applications, the kernel is uniformly bounded (such as when it is some p-value), in which case, the $\zeta_{c,k}$ are also uniformly bounded as c, k, and n vary. In such case, $\zeta_{1,k}$ is of order $1/k$ and tends to zero. Nevertheless, the conditions in our theorem can be verified. As shown in Corollary 1, the important condition is that $k\zeta_{1,k} \not\to 0$.

Remark 1. (Simple Consistency). Under weak conditions, U_n is consistent in the sense $U_n - \theta_k \xrightarrow{P} 0$. It suffices to show $\mathrm{Var}(U_n) \to 0$. However, as is well known, $\mathrm{Var}(U_n) \leq k\zeta_{k,k}/n$. Thus, if the $\zeta_{k,k}$ are uniformly bounded (which follows if the kernels are uniformly bounded) and $k/n \to 0$, then consistency follows.

The theorem below applies in a triangular array setup, where n observations are i.i.d. P_n. Then, quantities such as $\zeta_{c,k}$ in (3.1) are computed under P_n. Let

$$\hat{U}_n = \frac{k_n}{n} \sum_{i=1}^{n} h_{1,k}(X_i) \ . \tag{3.2}$$

Theorem 1. *Assume the order $k = k_n$ of the kernel h_k satisfies $k^2/n \to 0$. Furthermore, assume that $\zeta_{k,k}/k\zeta_{1,k}$ is bounded.*

(i) Then,

$$\frac{n\mathrm{Var}(U_n)}{k^2\zeta_{1,k}} \to 1. \tag{3.3}$$

(ii) In addition,

$$\frac{(U_n - \theta_k) - \hat{U}_n}{\sqrt{(k^2/n)\zeta_{1,k}}} \xrightarrow{P} 0 \ , \tag{3.4}$$

and so

$$U_n - \theta_k = O_P\left(\frac{k^2}{n}\zeta_{1,k}\right) .$$

(ii) If, in addition, for all $\delta > 0$,

$$\lim_{n\to\infty} \frac{1}{\zeta_{1,k}} \int_{|h_{1,k}(x)|>\delta\sqrt{n\zeta_{1,k}}} h_{1,k}^2(x) dP_n(x) = 0, \tag{3.5}$$

then

$$\frac{\sqrt{n}\,(U_n(X_1,\ldots,X_n) - \theta_k)}{\sqrt{k^2\zeta_{1,k}}} \xrightarrow{d} N(0,1). \tag{3.6}$$

This result also holds for the "incomplete" U-statistic, which is the average of the kernels computed over B_n randomly and uniformly chosen subsamples of the data, provided that $n/B_n \to 0$.

Corollary 1. *Under the above notation, if $k^2/n \to 0$, the kernel h_k is uniformly bounded (both as k and the data vary), and $k\zeta_{1,k} \nrightarrow 0$, then asymptotic normality* (3.6) *holds.*

Remark 2. In some applications, the condition that $k\zeta_{1,k} \nrightarrow 0$ holds because $k\zeta_{1,k}$ is of strict order one. Of course, if k is fixed, as in the classical case, all that is required for asymptotic normality is $\zeta_{1,k} > 0$.

3.2. Asymptotic normality of the M-statistic

Suppose instead of using U_n as an estimator, where the kernel is averaged over all subsamples of size k of the data, we use the median of the values of the kernel computed on all subsamples of size k, that is, M_n defined in (1.2), which we refer to as an M-statistic. In this section, we do not assume h_k is symmetric, and so the median is taken over all $n!/(n-k)!$ ordered indices $i_1, \ldots, i_k$ taken without replacement from $1, \ldots, n$. (If taking the median over an even number, say $2m$ values, define the median as the mth-order statistic. Note that if $k > 1$, then $\binom{n}{k}k!$ is always even.) We would like to prove a triangular array CLT for M_n when $k = k_n$ varies with n.

Suppose that h_k has a c.d.f. F_k, and that $\tilde{\theta}_k$ satisfies $F_k(\tilde{\theta}_k) = 1/2$.

Define

$$\tilde{h}_k(x_1, \ldots, x_k; t) := \frac{1}{k!} \sum I\left\{h_k(x_{i_1}, \ldots, x_{i_k}) > \tilde{\theta}_k + t\right\}, \tag{3.7}$$

where the average is taken over all permutations of $1, \ldots, k$. In addition, define

$$\tilde{\zeta}_{1,k}(t) = \mathrm{Var}[\tilde{\phi}_{1,k}(X; t)],$$

with

$$\tilde{\phi}_{1,k}(x; t) = E[\tilde{h}_k(x, X_2, \ldots, X_k; t)].$$

We assume that the sequence $\{F_k\}$ is asymptotically equidifferentiable (as $k = k_n \to \infty$) relative to the sequence $\tilde{\theta}_k$; that is, for any $\epsilon_k \to 0$,

$$F_k(\tilde{\theta}_k + \epsilon_k) - F_k(\tilde{\theta}_k) = \epsilon_k F_k'(\tilde{\theta}_k) + o(\epsilon_k) \ . \tag{3.8}$$

We apply (3.8) with the choice $\epsilon_k = \delta_k$ defined by

$$\delta_k = \sqrt{\frac{\tilde{\zeta}_{1,k}(0)k^2}{n}}.$$

Note that $\tilde{\zeta}_{1,k}$ is bounded in k, so that if we assume that $k^2/n \to 0$, then

$\delta_k \to 0$. Then,

$$E\left(\tilde{h}_k\left(X_1,\ldots,X_k;\delta_k\right)\right) = \frac{1}{2} - F_k'(\tilde{\theta}_k)\delta_k + o(\delta_k). \tag{3.9}$$

Finally, assume that $F_k'(\tilde{\theta}_k) \to f(\tilde{\theta})$, which is just some positive constant. (Note, f and $\tilde{\theta}$ need not have meaning separately. However, typically, F_k' tends to some f and $\tilde{\theta}_k \to \tilde{\theta}$.)

Theorem 2. *Under the above setup, assume further that* $k^2/n \to 0$, $k\zeta_{1,k}(0) \nrightarrow 0$ *and, for any fixed* t,

$$\frac{\tilde{\zeta}_{1,k}(\delta_k t)}{\tilde{\zeta}_{1,k}(0)} \to 1 \tag{3.10}$$

as $n \to \infty$. *Then,*

$$\sqrt{\frac{n}{\tilde{\zeta}_{1,k}(0)k^2}}\left(M_n - \tilde{\theta}_k\right) \xrightarrow{d} N\left(0, \frac{1}{f^2(\tilde{\theta})}\right).$$

4. Examples, Revisited

Example 5. [Example 1, revisited.] Consider U_n given by (1.1), with h_k given by (2.1). We verify the conditions for asymptotic normality under H_0, though power can be studied similarly. Letting Z denote a standard normal variable,

$$E(U_n) = E\left[\exp(\epsilon\sqrt{k}Z)\right] = \exp\left(\frac{\epsilon^2 k}{2}\right).$$

In addition,

$$\begin{aligned} E[h_k(X_1,X_2,\ldots,X_k)|X_1] &= \exp(\epsilon X_1)E\left\{\exp[\epsilon(X_2+\cdots+X_k)]\right\} \\ &= \exp\left[\epsilon X_1 + \frac{\epsilon^2(k-1)}{2}\right]. \end{aligned}$$

Then, $\zeta_{1,k}$, the variance of this last quantity, is given by

$$\begin{aligned} \zeta_{1,k} &= \exp[\epsilon^2(k-1)]Var[\exp(\epsilon X_1)] \\ &= \exp[\epsilon^2(k-1)]\left[E\exp(2\epsilon X_1) - (E\exp(\epsilon X_1))^2\right] \\ &= \exp[\epsilon^2(k-1)][\exp(2\epsilon^2) - \exp(\epsilon^2)] = \exp(\epsilon^2 k)[\exp(\epsilon^2) - 1]. \end{aligned}$$

Similarly,

$$\zeta_{k,k} = Var\left\{\exp[\epsilon(X_1+\cdots+X_k)]\right\} = E\left[\exp(2\epsilon\sqrt{k}Z)\right] - \left\{E\left[\exp(\epsilon\sqrt{k}Z)\right]\right\}^2$$

$$= \exp(2\epsilon^2 k) - \exp(\epsilon^2 k) = \exp(\epsilon^2 k)[\exp(\epsilon^2 k) - 1].$$

We need to verify that the ratio $\zeta_{k,k}/(k\zeta_{1,k})$ is bounded. However,

$$\frac{\zeta_{k,k}}{k\zeta_{1,k}} = \frac{\exp(\epsilon^2 k) - 1}{k\,[\exp(\epsilon^2) - 1]}\ . \tag{4.1}$$

Of course, if $k = 1$, then the ratio (4.1) is always one, so the condition holds. Certainly, if both $k > 1$ and $\epsilon > 0$ are fixed, then the ratio (4.1) is fixed. In addition, if k is fixed, but $\epsilon = \epsilon_k \to 0$, then by L'Hospital's rule, the ratio tends to one, and so the condition holds. If $k \to \infty$, but $\epsilon^2 k \to 0$ (so that also $\epsilon^2 \to 0$), then by Taylor approximation to the numerator and denominator, it is easy to see that the condition holds, as again, the ratio tends to one. Actually, one just needs $\epsilon^2 k$ to remain bounded. Indeed, the numerator in (4.1) is then bounded, and the denominator is easily seen to be bounded below by $k\epsilon^2$. If $k\epsilon^2 \to 0$, then we have already treated that case, but if it is bounded away from zero and ∞, then the ratio (4.1) is bounded. Hence, it is only required that $\epsilon^2 k$ is bounded from above (unless $k = 1$, in which case, the condition holds regardless). Conversely, it is easy to check that if $k > 1$ and $k\epsilon^2 \to \infty$, then the ratio (4.1) is not bounded. Note that if we are trying to detect an alternative in which k of the μ_i are equal to ϵ, and the rest are zero, then such alternatives are contiguous to the null. Finally, asymptotic normality holds as long as ϵ_k stays bounded from above (and so it can tend to zero).

Example 6. [Example 2, revisited.] Consider the generalized Hodges-Lehmann estimators $\hat{\theta}_{n,k} = k^{-1/2} M_n$, where M_n is defined by (1.2) with

$$h_k(X_1, \ldots, X_k) = k^{-1/2} \sum_{i=1}^{k} X_i.$$

To illustrate the theorem, assume $X_1, \ldots, X_n$ are i.i.d. normal with mean zero and variance one. Therefore, assume $k > 1$. We have $\tilde{\theta}_k = 0$ and

$$\tilde{h}_k(x_1, \ldots, x_k; t) = I\left(\frac{x_1 + \cdots + x_k}{\sqrt{k}} > t\right)\ .$$

Note that (3.9) holds with $f(\tilde{\theta}) = \phi(0) = 1/\sqrt{2\pi}$, where $\phi(\cdot)$ is the standard normal density. Then,

$$\tilde{\phi}_{1,k}(x, t) = P\left\{\frac{x + X_2 + \cdots + X_k}{\sqrt{k}} > t\right\}$$

$$= 1 - \Phi\left(t\sqrt{\frac{k}{k-1}} - \frac{x}{\sqrt{k-1}}\right),$$

$$\tilde{\zeta}_{1,k}(t) = Var\left[\Phi\left(t\sqrt{\frac{k}{k-1}} - \frac{X}{\sqrt{k-1}}\right)\right]$$

and

$$\tilde{\zeta}_{1,k}(0) = Var\left[\Phi\left(\frac{X}{\sqrt{k-1}}\right)\right] := \tau_k^2,$$

where $X \sim N(0,1)$. Note that by the Taylor approximation, for large k,

$$\tau_k^2 = Var\left[\frac{X}{\sqrt{k-1}}\phi(0)\right] = \frac{1}{2\pi(k-1)} + o\biggl(\frac{1}{k}\biggr), \tag{4.2}$$

and thus

$$\lim_{k\to\infty} k\tau_k^2 = \frac{1}{2\pi}\ .$$

Here, and in other places, the approximation (4.2) can be justified because

$$\tau_k^2 = Var\left(\Phi(0) + \frac{X}{\sqrt{k-1}}\phi(0) + R_k\right) = \frac{1}{2\pi(k-1)} + o\biggl(\frac{1}{k}\biggr),$$

where

$$R_k = \frac{X^2}{2(k-1)}\phi'(Y_k),$$

and Y_k is some intermediate point between *zero* and $X/\sqrt{k-1}$. Because ϕ' (as well as higher derivatives) is uniformly bounded, it follows that $E(R_k^2) = O(1/k^2)$. Then, by Cauchy–Shwartz, $Cov((X/\sqrt{k-1}), R_k) = o(1/k)$, and so (4.2) follows. (In fact, by continuing the Taylor Series for another term and noting that $Cov(X, X^2) = 0$, we find that the error is not just $o(1/k)$, but $O(1/k^2)$.)

Note that, because $\tilde{\zeta}_{1,k}(0) = O(1/k)$, $\delta_k = O(\sqrt{k/n})$. Similarly, by the Taylor approximation,

$$\begin{aligned}\tilde{\zeta}_{1,k}(\delta_k t) &= Var\left[\Phi\left(\delta_k t\sqrt{\frac{k}{k-1}} - \frac{X}{\sqrt{k-1}}\right)\right]\\ &= Var\left[\Phi\left(\frac{X}{\sqrt{k-1}}\right)\right] + O\biggl(\frac{\delta_k^2}{k}\biggr).\end{aligned}$$

If k is fixed, $\delta_k^2/k \to 0$ and (3.10) holds easily. If $k \to \infty$, then $k\tilde{\zeta}_{1,k}(\delta_k t) \to 1/(2\pi)$, because $kO(\delta_k^2/k) = o(1)$. Therefore, condition (3.10) holds. Hence,

$$\sqrt{\frac{n}{\tau_k^2 k^2}} M_n \xrightarrow{d} N(0, 2\pi) \ ,$$

or, equivalently,

$$\sqrt{\frac{n}{k\tau_k^2}} \hat{\theta}_{n,k} \xrightarrow{d} N(0, 2\pi) \ .$$

Therefore, when k is fixed,

$$\sqrt{n}\hat{\theta}_{n,k} \xrightarrow{d} N(0, 2\pi k \tau_k^2) \ . \tag{4.3}$$

When $k \to \infty$ and $k^2/n \to 0$, because $k\tau_k^2 \to 1/(2\pi)$, we have

$$\sqrt{n}\hat{\theta}_{n,k} \xrightarrow{d} N(0, 1) \ .$$

In this case, $\hat{\theta}_{n,k}$ is asymptotically efficient.

Example 7. [Example 3, revisited.] Consider the subsampling estimator $U_n(\cdot)$ defined in (2.2). Fix t, and note that $U_n = U_n(t)$ has expectation $\theta_k = J_k(t, P)$, where $J_k(t, P)$ is the true sampling distribution of $\tau_k(\hat{\xi}_k - \xi_0)$ based on a sample of size k. Typical subsampling arguments, as in Chapter 2 of Politis, Romano and Wolf (1999), show that $U_n(t) - J_n(t, P) \xrightarrow{p} 0$. A more detailed result would be to find the order of the error in the difference, or even its limiting distribution. To this end, we can simply write

$$U_n(t) - J_n(t, P) = [U_n(t) - \theta_k] - [J_n(t, P) - J_k(t, P)] \ .$$

The bias term $[J_n(t, P) - J_k(t, P)]$ is nonrandom and can be analyzed separately (e.g., using Edgeworth expansions). The U-statistic theory applies to the first term $[U_n(t) - \theta_k]$, the analysis of which we now illustrate using Corollary 1. We specialize as follows. Assume the X_i are i.i.d. $N(\xi, 1)$ and $\hat{\xi}_n = n^{-1}\sum_i X_i$. Take $\xi_0 = 0$ and $\tau_n = \sqrt{n}$. The kernel, $h_k(x_1, \ldots, x_k) = I\{k^{-1/2}\sum_{i=1}^k x_i \leq t\}$, is clearly bounded. Then,

$$\begin{aligned} h_{1,k}(x) &= P\left\{k^{-1/2}(X_1 + \cdots + X_k) \leq t | X_1 = x\right\} - \Phi(t) & (4.4)\\ &= \Phi\left(t\sqrt{\frac{k}{k-1}} - \frac{x}{\sqrt{k-1}}\right) - \Phi(t), & (4.5) \end{aligned}$$

and

$$\zeta_{1,k} = Var\left[\Phi\left(t\sqrt{\frac{k}{k-1}} - \frac{X}{\sqrt{k-1}}\right)\right] .$$

As $k \to \infty$, by the Taylor approximation,

$$\zeta_{1,k} = Var\left[\phi(t)\frac{X}{\sqrt{k-1}}\right] + o\bigg(\frac{1}{k}\bigg) = \frac{\phi^2(t)}{k} + o\bigg(\frac{1}{k}\bigg).$$

Note that the condition $k\zeta_{1,k} \nrightarrow 0$ holds easily because $k\zeta_{1,k} \to \phi^2(t) > 0$. Therefore, we conclude that if $k^2/n \to 0$ and $k \to \infty$, then

$$\sqrt{\frac{n}{k}}[U_n(t) - \Phi(t)] \xrightarrow{d} N(0, \phi^2(t)) \ .$$

Example 8. [Example 4, revisited.] Consider the average p-value, $\bar{p}_n$, computed on subsamples of size k of the data. We show how to use the basic results to derive its limiting distribution using a relatively simple example. We further derive the limiting distribution of $\bar{p}_n$ under contiguous alternatives, and compute the limiting local power function. Although the methodology is offered in a simplified setting, it shows the potential for such an approach more broadly. Specifically, we consider the context of testing for a single mean. Obviously, these methods are not needed here. However, this model admits simple expressions of asymptotic power, which facilitates comparisons of the methods. Moreover, it specifically shows that conservative methods are much too conservative, and result in tests with very low power.

Let $X_1, \ldots, X_n$ be i.i.d. real-valued with unknown mean μ. The problem is to test the null hypothesis H_0 that the mean is zero versus the alternative that the mean is greater than zero. In order to study the power of the tests that combine splits of the data, we further assume the underlying distribution is $N(\mu, 1)$.

Let $\bar{X}_{n,k,i}$ be the average of the ith subsample of size k,. In addition, let $\hat{p}_{n,k,i}$ denote the p-value based on this subsample; that is, $\hat{p}_{n,k,i} = 1 - \Phi(\sqrt{k}\bar{X}_{n,k,i})$. The limiting power of the UMP level-α test against the contiguous alternatives $h/\sqrt{n}$ is

$$1 - \Phi(z_{1-\alpha} - h)$$

when using the full data, and

$$1 - \Phi(z_{1-\alpha} - \sqrt{\tau}h) \tag{4.6}$$

when using a single subsample (or split) of size k satisfying $k/n = \tau$. Assume $k/n \to \tau \in (0,1)$, the fraction in the sample used for testing. Assume the number of splits or subsamples $N = \binom{n}{k}$, so that all possible splits are used. For $r \in (0,1)$, consider the conservative procedure (or family of procedures) that rejects H_0 if the proportion of p-values (computed over all splits) $\leq \alpha r$ is $\geq r$. (In the case

$r = 1/2$, the procedure requires that at least half of the p-values are $\leq \alpha/2$; equivalently, twice the median p-value must be $\leq \alpha$.) As shown in DiCiccio, DiCiccio and Romano (2020), this procedure is level-α. This is the exact or finite-sample version of an asymptotic approach first suggested in Meinshausen, Meier and Bühlmann (2009). They did not present any analytical expressions for power. DiCiccio, DiCiccio and Romano (2020) obtained the limiting power of this procedure for testing $H_0 : \mu = 0$ against contiguous alternatives $h/\sqrt{n}$, given by

$$1 - \Phi\left[\frac{1}{\sqrt{\tau}}(z_{1-r\alpha} - z_{1-r}\sqrt{1-\tau}) - h\right] . \tag{4.7}$$

Note that (4.7) shows that, even asymptotically, the approach is conservative; that is, when $h = 0$, the limiting rejection probability is below α. It further implies that the limiting power for small positive h can be less than α, and loss of power results. By comparison, the limiting power against $h/\sqrt{n}$ of a single-split sample test by taking one sample of size k is given by (4.6). Even with $\tau < 1$, the test based on a single subsample of size k has better limiting power for small h than that of the conservative tests, which combine p-values computed on many subsamples of size k. On the other hand, for sufficiently large h, (4.7) is larger than (4.6). In this case, the many-split sample test is an improvement over the single sample test, even though it conservatively controls the type-1 error. However, the power is only larger for values of the local parameter where the power is already near one.

By deriving the limiting distribution of the average (or median) p-value, we can construct an asymptotically level α with greatly improved power. Indeed, the distribution of $\bar{p}_n$ is concentrated near 1/2 under H_0, and so an appropriate critical value (sequence) is near 1/2 as well. In contrast, the conservative procedure uses a critical value of $\alpha/2$ (based on either the mean or median p-value). Furthermore, and perhaps surprisingly, tests exploiting the U-statistic structure achieve the optimal limiting local power function of the UMP level-α test. The challenge is to derive the appropriate limiting distribution, so that a better, or less conservative, critical value may be used.

Define the average p-value taken over all subsamples of size k as

$$U_n(X_1, \ldots, X_n) = \bar{p}_n = \frac{1}{N}\sum_{i=1}^{N} \hat{p}_{n,k,i} = \frac{1}{N}\sum_{i=1}^{N}[1 - \Phi(\sqrt{k}\bar{X}_{n,k,i})] ,$$

where $N = \binom{n}{k}$. Evidently, $\bar{p}_n$ is a U-statistic of the form (1.1).

Theorem 3. *Let $X_1, \ldots, X_n$ be* i.i.d. *according to a normal distribution with* *mean μ and variance one.*

(i) If k is fixed and $\mu = 0$, then

$$\sqrt{\frac{n}{k}}\left(\bar{p}_n - \frac{1}{2}\right) \xrightarrow{d} N(0, k\zeta_{1,k}) \ , \tag{4.8}$$

where

$$\zeta_{1,k} = Var\left[\Phi\left(\frac{X}{\sqrt{2k-1}}\right)\right] \ , \tag{4.9}$$

and $X \sim N(0,1)$ and $\Phi(\cdot)$ is the standard normal c.d.f.

(ii) If $k \to \infty$ and $k/\sqrt{n} \to 0$, then $k\zeta_{1,k} \to 1/(4\pi)$. Moreover, under $H_0 : \mu = 0$,

$$\sqrt{\frac{n}{k}}\left(\bar{p}_n - \frac{1}{2}\right) \xrightarrow{d} N\left(0, \frac{1}{4\pi}\right).$$

Consider the one-sided test that rejects H_0 if $\sqrt{n}(\bar{p}_n - 1/2) < z_\alpha k\sqrt{\zeta_{1,k}}$. Its limiting power against contiguous alternatives $h/\sqrt{n}$ is

$$P(N(h,1) > z_{1-\alpha}) = 1 - \Phi(z_{1-\alpha} - h) \ ,$$

which is the same as the UMP level-α test. The same is true if $k\sqrt{\zeta_{1,k}}$ is replaced by $\sqrt{k/4\pi}$ in the construction of the critical value of the test.

Remark 3. If k is fixed, the average of the p-values computed over all splits of the data remains asymptotically normal; however, the overall test is less powerful asymptotically than the UMP test against local alternatives. A justification of this is implicit in the proof of Theorem 3.

Despite testing on small portions of the data, using the average p-value has the same limiting local power as the UMP test. Using the asymptotic normality of the p-value, the test rejects when the average p-value falls below the threshold $1/2 + z_\alpha\sqrt{k/(4\pi n)}$. In contrast, the conservative method rejects when the average or median p-value is below $\alpha/2$, which can be substantially smaller than the above threshold.

An asymptotically level-α test can also be performed based on the median of the p-values by viewing the median p-value $\tilde{p}_n$ as a median statistic M_n of the form (1.2). This method also achieves the optimal local limiting power function.

Theorem 4. *Suppose $X_1, \ldots, X_n$ are i.i.d. according to a normal distribution with mean μ and variance one. Suppose $k \to \infty$ such that $k/\sqrt{n} \to 0$. Then,* *under a sequence of local alternatives* $h/\sqrt{n}$,

$$\sqrt{\frac{2\pi n}{k}}\left(\tilde{p}_n - \frac{1}{2}\right) \xrightarrow{d} N(h,1) ,$$

where $\tilde{p}_n$ *is the median p-value computed over all splits. Consider the test that rejects* H_0 *if* $\tilde{p}_n < 1/2 + z_\alpha\sqrt{k/(2\pi n)}$. *Then, the limiting power of the one-sided test of* $H_0 : \mu = 0$ *against* $h/\sqrt{n}$ *is*

$$1 - \Phi(z_{1-\alpha} - h) .$$

Note that the asymptotically level-α test rejects if the median is less than $1/2 + z_\alpha\sqrt{k/(2\pi n)}$, which can be substantially larger than $\alpha/2$. For example, if $\alpha = 0.1$, $n = 100$, and $k = 10$, $1/2 + z_\alpha\sqrt{k/n} \approx 0.34$, whereas $\alpha/2 = 0.05$. The asymptotic local power of this test based on the median p-value using an appropriate (not conservative) critical value achieves that of the optimal UMP test.

5. Conclusion and Further Questions

In this paper, we have considered a U-statistic sequence where the kernel size grows with the sample size. We developed conditions under which asymptotic normality results. At the same time, we considered the corresponding M-statistic, defined as the median of the kernel computed over subsamples of the data. Other quantiles can be considered by similar arguments. Using four examples, we demonstrated the utility of our results, and verified the conditions, showing how to verify the conditions and calculate relevant quantities (e.g., asymptotic variances) in more complex problems. The problem was largely motivated by that of combining p-values obtained by data splitting, as well as providing sufficient conditions that can be verified. The simplified example suggests that the statistical approach may be quite promising. Our results will allow further development of this area, where only conservative procedures are in use.

6. Proofs

Proof of Theorem 1. To prove (i), follow, for example, the argument in van der Vaart (1998). Thus, it suffices to show that $\mathrm{Var}(U_n)/\mathrm{Var}(\hat{U}_n) \to 1$, where $\hat{U}_n$ is defined in (3.2). Indeed, Theorem 11.2 of van der Vaart (1998) applies not only for fixed k, but also when $k = k_n \to \infty$. As is well known (and argued in the proof of Theorem 12.3 of van der Vaart (1998)),

$$\mathrm{Var}(U_n) = \sum_{c=1}^{k} \binom{n}{k}^{-1} \binom{k}{c} \binom{n-k}{k-c} \zeta_{c,k} \ , \tag{6.1}$$

where

$$\zeta_{c,k} = \mathrm{Cov}\left[h_k(X_1, \ldots, X_c, X_{c+1}, \ldots, X_k), h_k(X_1, \ldots, X_c, X_{k+1}, \ldots, X_{2k-c}) \right] \ , \tag{6.2}$$

the covariance between the kernel based on two data sets with exactly c variables in common. By conditioning on $X_1, \ldots, X_c$, it is readily seen that (3.1) and (6.2) agree. First, note that the $c = 1$ term in (6.1) divided by $\mathrm{Var}(\hat{U}_n) = k^2 \zeta_{1,k}/n$ tends to one; that is,

$$\frac{(k/\binom{n}{k})\binom{n-k}{k-1}\zeta_{1,k}}{(k^2/n)\zeta_{1,k}} = \frac{(n-k)!(n-k)!}{(n-1)!(n-2k+1)!} \to 1.$$

The last limit uses $k^2/n \to 0$, and can be seen by applying Stirling's formula, taking logs, and using a Taylor expansion. What remains is to show that the sum from $c = 2$ to $c = k$ in (6.1) divided by $k^2 \zeta_{1,k}/n$ tends to zero. However,

$$\begin{aligned}
&\frac{\sum_{c=2}^{k} \binom{n}{k}^{-1} \binom{k}{c} \binom{n-k}{k-c} \zeta_{c,k}}{(k^2/n)\zeta_{1,k}} \\
&\leq \frac{\sum_{c=2}^{k} (1/c!) \left[k!/(k-c)! \right]^2 ((n-k)!/n!)((n-k)!/(n-2k+c)!)\zeta_{c,k}}{(k^2/n)\zeta_{1,n}} \\
&\leq \frac{\sum_{c=2}^{k} (k^{2c}/c!)(1/(n-k+1)^c)\zeta_{c,k}}{(k^2/n)\zeta_{1,k}} \leq \sum_{c=2}^{k} \frac{1}{c!} \epsilon_n^{c-1} \frac{\zeta_{c,k}}{\zeta_{1,k}},
\end{aligned} \tag{6.3}$$

where

$$\epsilon_n = \frac{k^2}{n-k+1} \ .$$

Using the inequality $\zeta_{c,k} \leq c\zeta_{k,k}/k$ (see Hoeffding (1948)) gives that (6.3) is bounded above by

$$\frac{\zeta_{k,k}}{k\zeta_{1,k}} \sum_{c=2}^{k} \frac{1}{(c-1)!} \epsilon_n^{c-1} \leq \frac{\zeta_{k,k}}{k\zeta_{1,k}} \sum_{j=1}^{k-1} \epsilon_n^j = \frac{\zeta_{k,k}}{k\zeta_{1,k}} \cdot \frac{\epsilon_n - \epsilon_n^k}{1 - \epsilon_n} \ . \tag{6.4}$$

The second factor in the last expression for (6.4) tends to zero because $\epsilon_n \to 0$. Thus, as long as $\zeta_{k,k}/k\zeta_{1,k}$ stays bounded, the result follows.

To prove (ii), note that expression (3.4) has mean zero and variance given by one minus the left-hand side of (3.3). Apply Chebychev. The rest of the proof is then trivial.

Proof of Corollary 1. Because the h_k are uniformly bounded, so are the $\zeta_{k,k}$. Hence, the condition in Theorem 1 $\zeta_{k,k}/k\zeta_{1,k}$ is bounded, because $k\zeta_{1,k} \nrightarrow 0$. Moreover, the Lindeberg condition (3.5) necessarily holds because $n\zeta_{1,k} = (n/k) \cdot k\zeta_{1,k} \to \infty$, so that the region of integration in the integral is empty for large n.

Proof of Theorem 2. For any fixed t,

$$P\left\{\sqrt{\frac{n}{\tilde{\zeta}_{1,k}(0)k^2}}\left(M_n - \tilde{\theta}_k\right) \le t\right\} = P\left\{M_n \le \tilde{\theta}_k + \delta_k t\right\}$$

$$= P\left\{\binom{n}{k}^{-1}\sum \tilde{h}_k\left(X_{i_1}, \ldots, X_{i_k}; \delta_k t\right) \le \frac{1}{2}\right\} = P\{Z_n \le x_k\},$$

where

$$Z_n = \sqrt{\frac{n}{\tilde{\zeta}_{1,k}(0)k_n^2}}\binom{n}{k}^{-1}\sum\left[\tilde{h}_k(X_{i_1}, \ldots, X_{i_k}; \delta_k t) - E\tilde{h}_k(X_{i_1}, \ldots, X_{i_k}; \delta_k t)\right]$$

and

$$x_k = \frac{1}{\delta_k}\left\{\frac{1}{2} - E[\tilde{h}_k(X_{i_1}, \ldots, X_{i_k}; \delta_k t]\right\}.$$

(The above follows by definition of the median, and then substracting $E\tilde{h}_k(X_1, \ldots, X_k; \delta_k t)$ from both sides and dividing by δ_k.) We claim $Z_n \xrightarrow{d} N(0,1)$. To establish this, consider the U-statistic $U_n = U_n(t)$, with symmetric kernel $\tilde{h}_k(\cdot; t)$ defined by

$$U_n(t) = \binom{n}{k}^{-1}\sum \tilde{h}_k(X_{i_1}, \ldots, X_{i_k}; \delta_k t)\ .$$

By Corollary 1,

$$\sqrt{\frac{n}{k^2\tilde{\zeta}_{1,k}(\delta_k t)}}[U_n(t) - E(U_n(t))] \xrightarrow{d} N(0,1)\ . \tag{6.5}$$

The left side of (6.5) and Z_n only differ in that $\zeta_{1,k}(0)$ in Z_n is replaced by $\zeta_{1,k}(\delta_k t)$. However, the assumption that $\tilde{\zeta}_{1,k}(\delta_k t)/\tilde{\zeta}_k(0) \to 1$ together with Slutsky's theorem proves that $Z_n \xrightarrow{d} N(0,1)$. In addition, (3.9) and the assumption $F_k'(\tilde{\theta}_k) \to f(\tilde{\theta})$ imply that $x_k \to f(\tilde{\theta})t$. Therefore, by Slutsky's theorem,

$$P\{Z_n \le x_k\} \to \Phi]f(\tilde{\theta})t],$$

as required.

Proof of Theorem 3. We first apply Theorem 1 in the case where the order of the kernel k is fixed. Define the kernel

$$h_k(X_1, \ldots, X_k) = 1 - \Phi(\sqrt{k}\bar{X}_k) \ ,$$

which is the p-value of a test of H_0 computed on a subsample of size k, and $\bar{X}_k = \sum_{i=1}^{k} X_i/k$. For this choice of kernel,

$$h_{1,k}(x) = 1 - E\left(\Phi(\sqrt{k}\bar{X}_k)|X_1 = x\right) = 1 - E\Phi\left(\frac{x}{\sqrt{k}} + Y\right),$$

where $Y \sim N(0, (k-1)/k)$. Thus, we can simplify

$$h_{1,k}(x) = 1 - E\left[I\left\{Z < \frac{x}{\sqrt{k}} + Y\right\}\right] \ ,$$

where $Z \sim N(0,1)$ and Z is independent of Y. Therefore,

$$h_{1,k}(x) = 1 - \Phi\left(\frac{x}{\sqrt{2k-1}}\right) ,$$

and $\zeta_{1,k}$ is given in (4.9). By Theorem 1, it follows that, under H_0,

$$\sqrt{n}\left(\bar{p}_n - \frac{1}{2}\right) = \frac{k}{\sqrt{n}} \sum_{i=1}^{n} \left[h_{1,k}(X_i) - \frac{1}{2}\right] + o_P(1), \tag{6.6}$$

and so (4.8) follows. To calculate the limiting distribution under the sequence of alternatives when the mean is $h/\sqrt{n}$, note that by contiguity, the approximation (6.6) holds as well; that is, the term that goes to zero in probability under $h = 0$ does so under general h as well. The linear term does not have mean $1/2$, but we can use a Taylor expansion argument (and noting that the moments in the error term are bounded) to calculate

$$E_h[h_{1,k}(X)] = 1 - E\left[\Phi\left(\frac{Z + h/\sqrt{n}}{\sqrt{2k-1}}\right)\right] ,$$

where $Z \sim N(0,1)$. Then,

$$E_h[h_{1,k}(X)] = \frac{1}{2} - \frac{h/\sqrt{n}}{\sqrt{2k-1}} E\left[\phi\left(\frac{Z}{\sqrt{2k-1}}\right)\right] + O\left(\frac{1}{n}\right).$$

However, using that the moment-generating function of Z^2 is $(1-2t)^{-1/2}$, one can calculate

$$E\left[\phi\left(\frac{Z}{\sqrt{2k-1}}\right)\right] = \frac{1}{\sqrt{2\pi}} \cdot \left(1 - \frac{1}{2k}\right)^{1/2} ,$$

and thus

$$E_h[h_{1,k}(X)] = \frac{1}{2} - \frac{h}{\sqrt{4\pi kn}} + O\left(\frac{1}{n}\right) .$$

Furthermore, under $\mu = h/\sqrt{n}$,

$$Var_h[h_{1,k}(X)] = Var\left[\Phi\left(\frac{Z}{\sqrt{2k-1}} + \frac{h/\sqrt{n}}{\sqrt{2k-1}}\right)\right] = \zeta_{1,k} + o(n^{-1/2}) .$$

By (6.6) and these calculations, it follows that, under $h/\sqrt{n}$,

$$\sqrt{n}\left(\bar{p}_n - \frac{1}{2}\right) \xrightarrow{d} N\left(-\sqrt{\frac{k}{4\pi}}h, k^2\zeta_{1,k}\right) .$$

It now follows that the test that rejects if $\sqrt{n}(\bar{p}_n - 1/2) < z_\alpha k\sqrt{\zeta_{1,k}}$ has limiting power or rejection probably under $h/\sqrt{n}$ given by

$$P_h\left\{\sqrt{n}\left(\bar{p}_n - \frac{1}{2}\right) < z_\alpha k\sqrt{\zeta_{1,k}}\right\} = 1 - \Phi\left(z_{1-\alpha} - \frac{h}{\sqrt{4\pi k\zeta_{1,k}}}\right) .$$

We now show $k\zeta_{1,k} \to (4\pi)^{-1}$ as $k \to \infty$. However,

$$k\zeta_{1,k} = kVar\left[\Phi\left(\frac{Z}{\sqrt{2k-1}}\right)\right] = kVar\left[\Phi(0) + \frac{Z}{\sqrt{2k-1}}\phi(0) + r_k\right] ,$$

where the error term can be ignored because it has a variance of order $1/k^2$. Hence,

$$k\zeta_{1,k} = k\frac{1}{2\pi(2k-1)} + o(1) \to \frac{1}{4}\pi.$$

Thus, as $k \to \infty$, the limiting power tends $1 - \Phi(z_{1-\alpha} - h)$, as in the case of the UMP test.

For the case in which $k \to \infty$ at the same time as $n \to \infty$, we can apply Theorem 1, along with the same calculations for fixed k.

Proof of Theorem 4. Here, we follow the notation of Theorem 2, with

$$h_k(X_1, \ldots, X_k; t) = I\left\{1 - \Phi(\sqrt{k}\bar{X}_k) > \tilde{\theta}_k + t\right\} .$$

Then, $\tilde{\theta}_k$ is the median of the distribution of h_k under $h/\sqrt{n}$, or the median of the distribution of $1-\Phi(Z+h\sqrt{k/n})$ when Z is standard normal. Thus, a trivial calculation gives $\tilde{\theta}_k = 1-\Phi(h\sqrt{k/n})$. Then,

$$\tilde{\phi}_{1,k}(x;t) = E[h_k(x, X_2, \ldots, X_k); t],$$

and

$$\tilde{\zeta}_{1,k}(t) = \mathrm{Var}[\tilde{\phi}_{1,k}(X;t)].$$

Now,

$$\begin{aligned}
\tilde{\phi}_{1,k}(x;t) &= P_h\left\{1-\Phi(\sqrt{k}\bar{X}_k) > 1-\Phi\left(h\sqrt{\frac{k}{n}}\right)+t\right\} \\
&= P\left\{\Phi\left(Y+\frac{x}{\sqrt{k}}\right) < \Phi\left(h\sqrt{\frac{k}{n}}\right)-t\right\} \\
&= P\left\{Y+\frac{x}{\sqrt{k}} < \Phi^{-1}\left[\Phi\left(h\sqrt{\frac{k}{n}}\right)-t\right]\right\},
\end{aligned}$$

where Y follows a normal distribution with mean $(k-1)h/\sqrt{nk}$ and variance $(k-1)/k$. Hence,

$$\tilde{\phi}_{1,k}(x;t) = \Phi\left[\frac{\Phi^{-1}[\Phi(h\sqrt{k/n})-t] - x/\sqrt{k} - (k-1)h/\sqrt{kn}}{\sqrt{(k-1)/k}}\right].$$

Assume the null hypothesis $h=0$, in which case $\tilde{\theta}_k = 1/2$. In this case,

$$\begin{aligned}
\tilde{\phi}_{1,k}(x;0) &= 1-\Phi\left(\sqrt{\frac{k}{k-1}}\frac{x}{\sqrt{k}}\right) \\
&= \frac{1}{2} - \sqrt{\frac{k}{k-1}}\frac{x}{\sqrt{k}}\phi(0) + o\left(\frac{1}{k}\right).
\end{aligned}$$

Therefore,

$$\frac{\tilde{\zeta}_{1,k}(0)}{(\phi(0))^2/k} \to 1$$

as $k\to\infty$. Similarly, one can show that

$$\zeta_{1,k}(t) = \frac{\phi^2(z_{\frac{1}{2}-t})}{k} + o\left(\frac{1}{k}\right);$$

thus, the conditions of Theorem 2 are met.

Therefore, we have that, under the null hypothesis,

$$\sqrt{n}\frac{\tilde{M}_n - 1/2}{\sqrt{k(\phi(0))^2}} \xrightarrow{d} N(0,1).$$

Under the sequence of local alternatives, $\mu = h/\sqrt{n}$, the median $\tilde{\theta}_k$ is given by

$$\tilde{\theta}_k = 1 - \Phi\left(h\sqrt{\frac{k}{n}}\right) = \frac{1}{2} - \phi(0)h\sqrt{\frac{k}{n}} + o_p\left(\frac{1}{\sqrt{n}}\right).$$

By similar arguments, the limiting local power of the test based on the median p-value is

$$1 - \Phi\left(z_{1-\alpha} - h\right).$$

Acknowledgments

We thank David Ritzwoller for his helpful comments and suggestions.

References

Arias-Castro, E., Candès, E. J. and Plan, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *The Annals of Statistics* **39**, 2533–2556.

Chen, L. H. and Shao, Q. M. (2007). Probability inequalities for sums of bounded random variables. *Bernoulli* **13**, 581–599.

DiCiccio, C., DiCiccio, T. and Romano, J. (2020). Exact tests via multiple data splitting. *Statistics and Probability Letters* **166**, 108865.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* **19**, 293–325.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses.* 3rd Edition. Springer, New York.

Meinshausen, N., Meier, L. and Bühlmann, P. (2009). p-values for high-dimensional regression. *Journal of the American Statistical Association* **104**, 1671–1681.

Mentch, L. and Hooker, G. (2016). Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. *Journal of Machine Learning Research* **17**, 1–41.

Peng, W., Coleman, T. and Mentch, L. (2019). Asymptotic distributions and rates of convergence for random forests via generalized U-statistics. *arXiv preprint arXiv:1905.10651.*

Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling.* Springer, New York.

Romano, J. P. and Shaikh, A. M. (2012). On the uniform asymptotic validity of subsampling and the bootstrap. *The Annals of Statistics* **40**, 2798–2822.

Ruschendorf, L. (1982). Random variables with maximum sums. *Advances in Applied Probability* **14**, 623–632.

Song, Y., Chen, X. and Kato, K. (2019). Approximating high-dimensional infinite-order U-statistics: Statistical and computational guarantees. *Electronic Journal of Statistics* **13**, 4797–4848.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

van Zwet, W. (1984). A berry-esseen bound for symmetric statistics. *Probability Theory and Related Fields* **66**, 425–440.

Vovk, V. and Wang, R. (2012). Combining p-values via averaging. *ArXiv e-prints.*

Zhou, Z., Mentch, L. and Hooker, G. (2019). Asymptotic normality and variance estimation for supervised ensembles. *arXiv preprint arXiv:1912.01089.*

Cyrus DiCiccio

LinkedIn Corporation, 1000 W Maude Ave., Sunnyvale, CA 94085, USA.

E-mail: cdiciccio@linkedin.com

Joseph Romano

Departments of Statistics and Economics, Stanford University, Sequoia Hall, Stanford, CA 94305, USA.

E-mail: romano@stanford.edu

(Received February 2020; accepted June 2020)