A new class of goodness-of-fit tests with applications to the problem of detecting sparse heterogeneous mixtures

Natalia Stepanova and Tatjana Pavlenko

Carleton University and KTH Royal Institute of Technology

August 2014, Linköping

Outline

- Statement of the problem and motivation
- Limit distributions under the null hypothesis
- Confidence bands
- Tabulation of cumulative distribution functions
- Detection of sparse heterogeneous mixtures
- Attainment of the Ingster optimal detection boundary
- Concluding remarks
- References

Statement of the problem

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. r.v.'s with a continuous CDF F on $\mathbb{R}$, and let $\mathbb{F}_n(t) = n^{-1}\sum_{i=1}^{n} \mathbb{I}(X_i \le t)$, $t \in \mathbb{R}$, be the EDF based on $X_1, \ldots, X_n$. Consider the problem of testing the hypothesis of goodness-of-fit

$$H_0:\ F = F_0$$

against either a two-tailed alternative $H_1:\ F \neq F_0$ or an upper-tailed alternative $H_1':\ F > F_0$, using the test statistics

$$T_n(q) = \sup_{0<F_0(t)<1} \frac{\sqrt{n}|\mathbb{F}_n(t) - F_0(t)|}{q(F_0(t))},$$

$$T_n^+(q) = \sup_{0<F_0(t)<1} \frac{\sqrt{n}\,(\mathbb{F}_n(t) - F_0(t))}{q(F_0(t))},$$

where function q belongs to some family of weight functions.

The family of EFKP upper-class functions

Definition 1: Let q be any strictly positive function defined on $(0,1)$ with the property $q(u) = q(1-u)$ for $u \in (0, 1/2)$, which is nondecreasing in a neighborhood of zero and nonincreasing in a neighborhood of one. Such a function will be called an **Erdős–Feller–Kolmogorov–Petrovski (EFKP) upper-class function** of a Brownian bridge $\{B(u), 0 \leq u \leq 1\}$, if there exists a constant $0 \leq b < \infty$ such that

$$\limsup_{u \to 0} |B(u)|/q(u) \overset{\text{a.s.}}{=} b. \tag{1}$$

An EFKP upper-class function q of a Brownian bridge is called a **Chibisov–O'Reilly function** if $b = 0$ in (1).

The family of EFKP upper-class functions (cont-d)

An important example of an EFKP upper-class function with $0 < b < \infty$ in (1) is the function

$$q(u) = \sqrt{u(1-u)\log\log(1/(u(1-u)))}. \tag{2}$$

Such a choice of q stems from Khinchine's local law of the iterated logarithm, which implies, via the representation of a Brownian bridge in terms of a standard Wiener process, that

$$\limsup_{u\to 0} \frac{|B(u)|}{\sqrt{u(1-u)\log\log(1/u(1-u))}} \overset{\text{a.s.}}{=} \sqrt{2}.$$

In applications, we recommend to use the weight function as in (2) since, unlike the Chibisov–O'Reilly function $q(u) = (u(1-u))^{1/2-\nu}$, $0 < \nu < 1/2$, it does not involve any parameter that has to be chosen by the experimenter.

The family of regularly varying functions

Definition 2: Let q be any strictly positive function defined on $(0,1)$ with the property $q(u) = q(1-u)$ for $u \in (0, 1/2)$, which is nondecreasing in a neighborhood of zero and non-increasing in a neighborhood of one. Such a weight function will be called **regularly varying with power** $\tau \in (0, 1/2]$ if for any $b > 0$

$$\lim_{t \to 0} q(bt)/q(t) = b^{\tau}.$$

The so-called **standard deviation proportional (SDP)** weight function

$$q(t) = \sqrt{t(1-t)}$$

is regularly varying with power $\tau = 1/2$, whereas the Chibisov–O'Reilly function $q(t) = (t(1-t))^{1/2-\nu}$, $\nu \in (0, 1/2)$, is regularly varying with power $\tau = 1/2 - \nu$.

Generalizations of the CsCsHM statistics

The two-sided statistic $T_n(q)$ with an EFKP upper-class function q appeared for the first time in the paper of M. Csörgő, S. Csörgő, Horváth, and Mason (1986). Therefore we call the statistics $T_n(q)$ and $T_n^+(q)$ **two-sided** and **one-sided Csörgő-Csörgő-Horváth-Mason (CsCsHM) statistics**, respectively. The following generalizations of the CsCsHM statistics are also of interest. For $0 \le a < b \le 1$, let $I = (a, b)$ and define the statistics

$$T_n(q, I) = \sup_{a < F_0(t) < b} \frac{\sqrt{n}|\mathbb{F}_n(t) - F_0(t)|}{q(F_0(t))},$$

$$T_n^+(q, I) = \sup_{a < F_0(t) < b} \frac{\sqrt{n}(\mathbb{F}_n(t) - F_0(t))}{q(F_0(t))},$$

which, for each n, have the same null distributions as $\sup_{u \in I} \sqrt{n}|\mathbb{U}_n(u) - u|/q(u)$ and $\sup_{u \in I} \sqrt{n}(\mathbb{U}_n(u) - u)/q(u)$.

Connection to the higher criticism approach

The EDF-based tests standardized by the SDP weight function $\delta(t) = \sqrt{t(1-t)}$ have been extensively studied in the literature. If under H_0 the iid observations are $U(0,1)$, a popular statistic of this kind is the **higher criticism statistic**

$$\text{HC}_n = \sup_{0<u<\alpha_0} \frac{\sqrt{n}(\mathbb{U}_n(u) - u)}{\sqrt{u(1-u)}}, \quad 0 < \alpha_0 < 1.$$

The statistic HC_n was introduced by Donoho & Jin (2004) for multiple testing situations where most of the component problems correspond to the null hypothesis and there may be a small fraction of component problems that correspond to non-null hypotheses.

Connection to the higher criticism approach (cont-d)

The test statistic HC_n is derived from the random variable

$$\max_{0<\alpha\le\alpha_0} \frac{\sqrt{n}\,(M_n/n-\alpha)}{\sqrt{\alpha(1-\alpha)}},$$

where M_n is the number of hypotheses among n independently tested hypotheses H_{0i}, $i=1,\ldots,n$, that are rejected at level α, which measures the maximum deviation of the observed proportion of rejections from what one would expect it to be purely by chance as the Type I error level changes from zero to α_0. Two modifications of HC_n due to Donoho & Jin (2004) and Jager & Wellner (2007) are:

$$\mathrm{HC}_n^+ = \sup_{1/n<u<\alpha_0} \frac{\sqrt{n}(\mathbb{U}_n(u)-u)}{\sqrt{u(1-u)}}, \quad \mathrm{HC}_n^* = \sup_{U_{(1)}<u<U_{([\alpha_0 n])}} \frac{\sqrt{n}(\mathbb{U}_n(u)-u)}{\sqrt{u(1-u)}}.$$

Convergence in distribution of the HC statistic

Proposition 1. For any $0 < \alpha_0 < 1$ and any $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbf{P}\left(a_n \sup_{0<u<\alpha_0} \frac{\sqrt{n}|\mathbb{U}_n(u) - u|}{\sqrt{u(1-u)}} - b_n \le x \right) = e^{-e^{-x}},$$

$$\lim_{n\to\infty} \mathbf{P}\left(a_n \sup_{0<u<\alpha_0} \frac{\sqrt{n}(\mathbb{U}_n(u) - u)}{\sqrt{u(1-u)}} - b_n \le x \right) = e^{-\frac{1}{2}e^{-x}},$$

where

$$a_n = \sqrt{2\log\log n}, \quad b_n = 2\log\log n + \frac{1}{2}\log\log\log n - \frac{1}{2}\log(4\pi).$$

Thus, regardless of a particular value of $0 < \alpha_0 < 1$, one always has the same extreme value distribution. Proposition 1 continues to hold for HC_n^+ and HC_n^*.

Motivation

In the sup-norm scenario, when normalizing the process $\sqrt{n}(\mathbb{U}_n(u) - u)$ by $\sqrt{u(1-u)}$, one arrives at the situation where **"all the action takes place in the tails"**. This observation together with the fact that, under the null hypothesis, the statistics HC_n, HC_n^+, and HC_n^* tend to ∞ in probability (see Prop. 1), as well as almost surely (see Ch. 16 in Shorack and Wellner (1986)), motivated us to search for a better weighed analog of the higher criticism statistic, for which the **"action is moved away from the tails"** and whose limit distribution is sensitive to the choice of α_0.

With this in mind, we propose the CsCsHM test statistics as competitors to HC_n and its modifications. Unlike the test procedures based on latter, in order to perform well, the test procedures based on the former do not require a very large sample size of $n = 10^6$ and work well even for $n = 10^2$.

Convergence in distribution of the CsCsHM statistics

The following extension of Theorem 4.2.3 in Csörgő et al. (1986) holds true.

Proposition 2. Let q be an EFKP upper-class function of a Brownian bridge $\{B(u), 0 \le u \le 1\}$. Then, under H_0, for any numbers $0 \le a < b \le 1$, as $n \to \infty$,

$$\sup_{a<F_0(t)<b} \frac{\sqrt{n}|\mathbb{F}_n(t) - F_0(t)|}{q(F_0(t))} \xrightarrow{\mathcal{D}} \sup_{a<u<b} \frac{|B(u)|}{q(u)},$$

$$\sup_{a<F_0(t)<b} \frac{\sqrt{n}(\mathbb{F}_n(t) - F_0(t))}{q(F_0(t))} \xrightarrow{\mathcal{D}} \sup_{a<u<b} \frac{B(u)}{q(u)}.$$

In particular, for the competitor of HC$_n$ we have

$$\sup_{0<F_0(t)<\alpha_0} \frac{\sqrt{n}(\mathbb{F}_n(t) - F_0(t))}{q(F_0(t))} \xrightarrow{\mathcal{D}} \sup_{0<u<\alpha_0} \frac{B(u)}{q(u)}.$$

Test procedures based on the CsCsHM-type statistics

The tests based on $T_n(q)$ and $T_n^+(q)$ are **consistent** against the alternatives $H_1 : F \neq F_0$ and $H_1' : F > F_0$, respectively.

The main advantage of using the family of the CsCsHM test statistics over the higher criticism statistics is the **identification of the proper limit distribution under the null hypothesis**. The above convergence in distribution results suggest the following test procedures of asymptotic level α. Set

$$T(q) := \sup_{0<u<1} |B(u)|/q(u), \quad T^+(q) := \sup_{0<u<1} B(u)/q(u).$$

Then, one would reject H_0 in favor of H_1 when $T_n(q) > t_\alpha(q)$, where $t_\alpha(q)$ is chosen to have $P(T(q) \geq t_\alpha(q)) = \alpha$; and one would reject H_0 in favour of H_1' whenever $T_n^+(q) > t_\alpha^+(q)$, where $t_\alpha^+(q)$ is determined by $P(T^+(q) \geq t_\alpha^+(q)) = \alpha$.

Confidence band based on the CsCsHM statistic

Proposition 2 continues to hold for the statistics

$$\sup_{a<F_0(t)<b} \frac{\sqrt{n}|\mathbb{F}_n(t) - F_0(t)|}{q(\mathbb{F}_n(t))}, \qquad \sup_{a<F_0(t)<b} \frac{\sqrt{n}(\mathbb{F}_n(t) - F_0(t))}{q(\mathbb{F}_n(t))},$$

where $\sqrt{n}|\mathbb{F}_n(t) - F_0(t)|/q(\mathbb{F}_n(t)) = 0$ for $\mathbb{F}_n(t) \in \{0, 1\}$. This result makes it possible to construct an asymptotically correct $100(1-\alpha)\%$ confidence band $[L_n(t), U_n(t)]$ for $F(t)$ on the interval $t \in [X_{(1)}, X_{(n)})$, where

$$L_n(t) = \max\{0, \mathbb{F}_n(t) - \frac{c_\alpha}{\sqrt{n}}\, q(\mathbb{F}_n(t))\},$$

$$U_n(t) = \min\{1, \mathbb{F}_n(t) + \frac{c_\alpha}{\sqrt{n}}\, q(\mathbb{F}_n(t))\},$$

and $c_\alpha = H^{-1}(1-\alpha)$ with $H(t) = \mathbf{P}\left(\sup_{0<u<1} |B(u)|/q(u) \leq t\right)$.

Numerical comparison of confidence bands

The three graphs below depict confidence bands for simulated data. The solid line is the true CDF. The solid lines above and below the middle line are a 95 percent CsCsHM confidence band. The red dashed lines are a 95 percent Kolmogorov–Smirnov confidence band. The blue dotted lines are a 95 percent Eicker–Jaeschke confidence band.

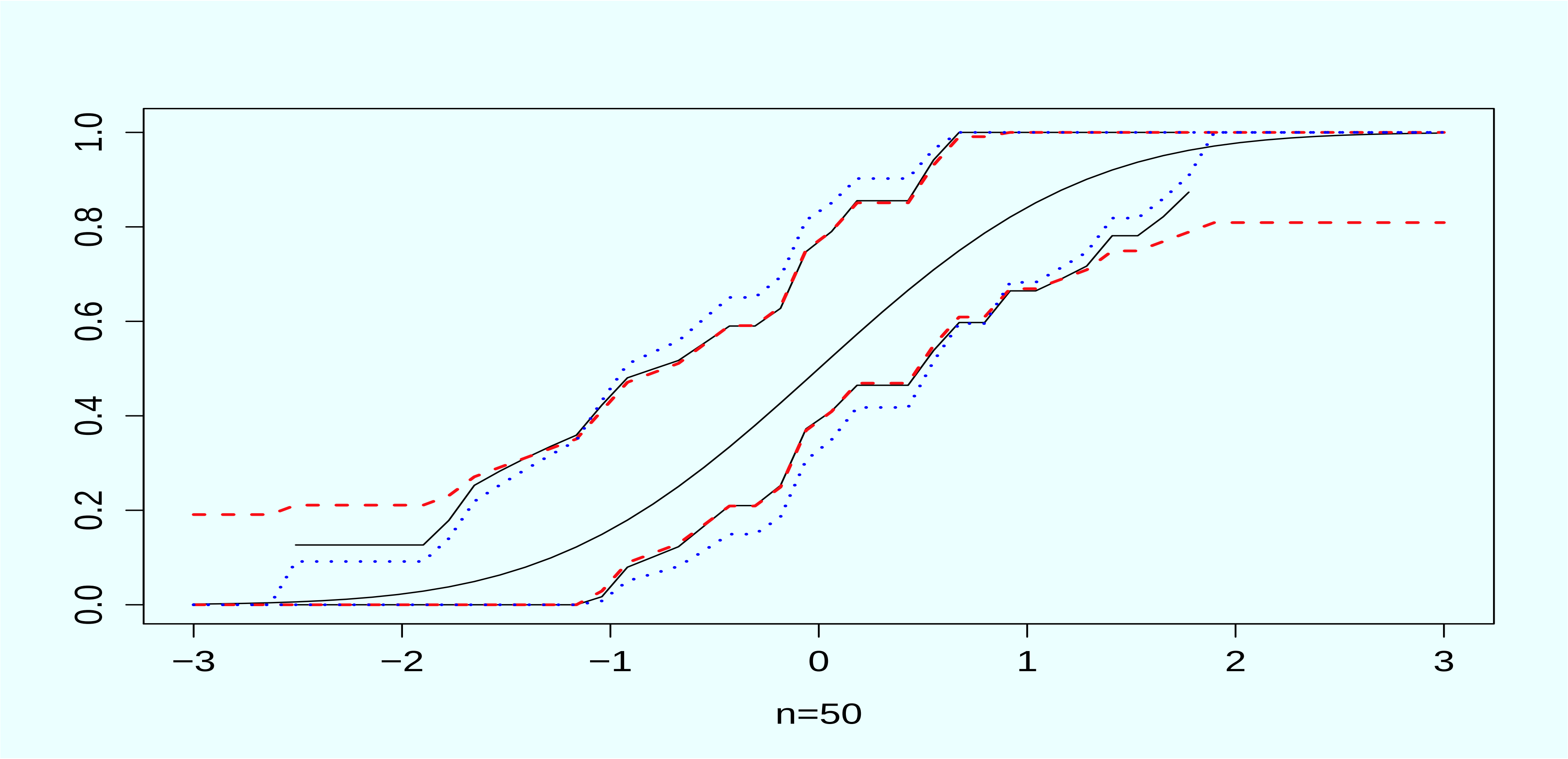
The Kolmogorov–Smirnov confidence band is derived from

$$\mathbf{P}_F\left(\sqrt{n}\sup_{-\infty<t<\infty}|\mathbb{F}_n(t)-F(t)|\leq x\right)\to K(x),$$

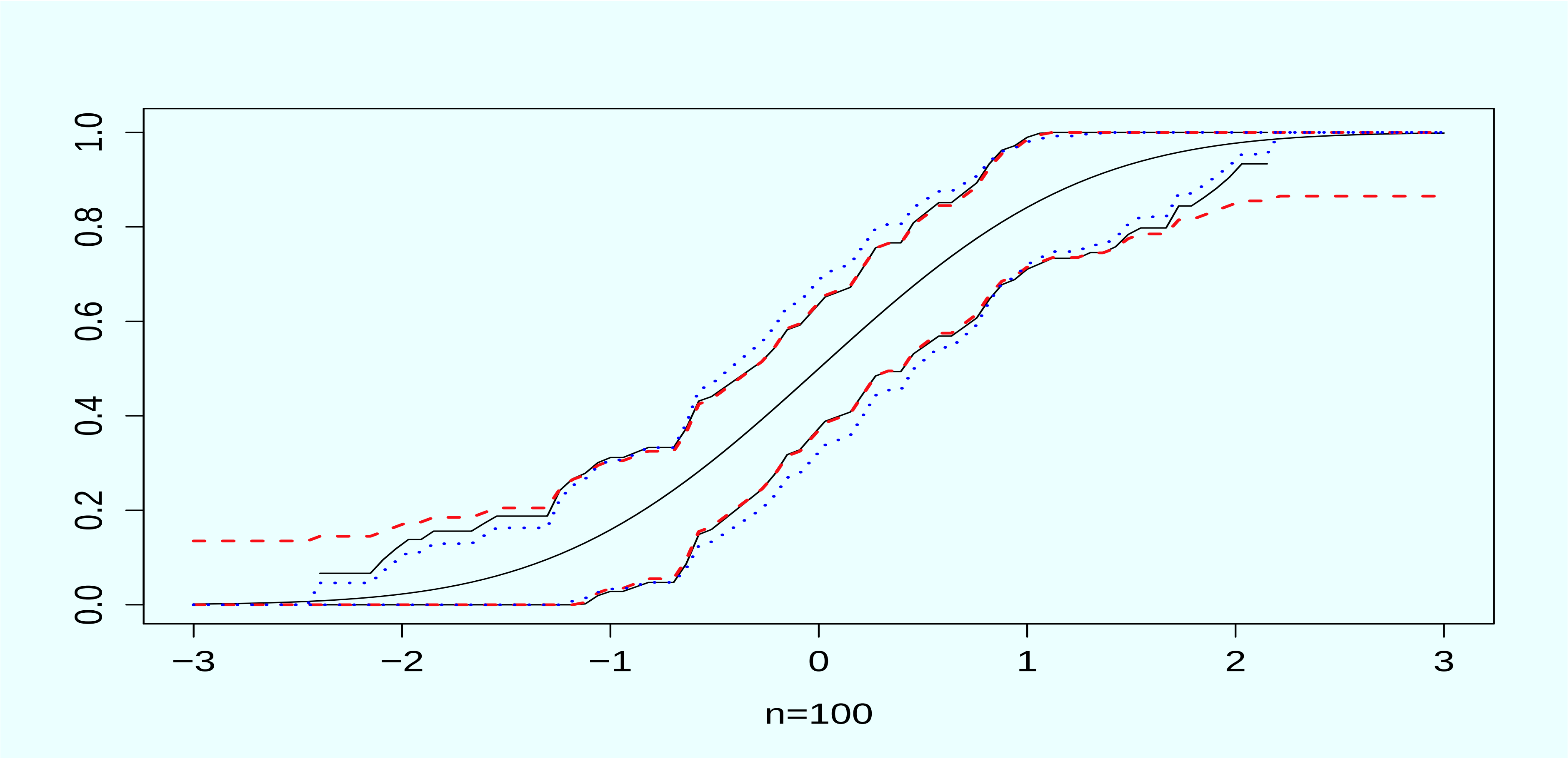
where $K(x)$ is the Kolmogorov CDF. The Eicker–Jaeschke confidence band is obtained from teh relation

$$\lim_{n\to\infty}\mathbf{P}_F\left(a_n\sup_{0<F(t)<1}\frac{\sqrt{n}|\mathbb{F}_n(t)-F(t)|}{\sqrt{\mathbb{F}_n(t)(1-\mathbb{F}_n(t))}}-b_n\leq x\right)=e^{-4e^{-x}}.$$

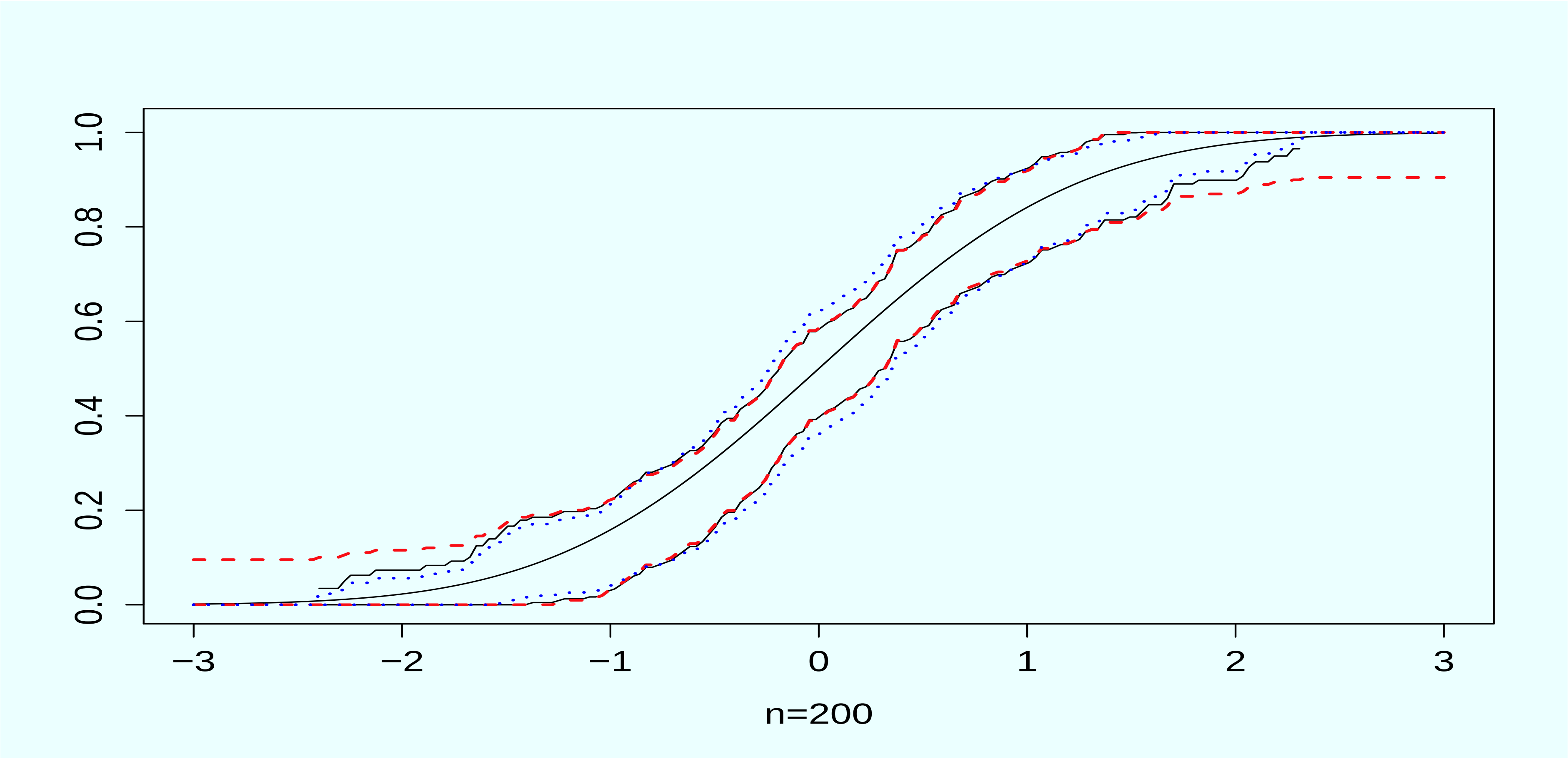
Numerical comparison of confidence bands (cont-d)



Numerical comparison of confidence bands (cont-d)



Numerical comparison of confidence bands (cont-d)



Summary of the numerical comparison of confidence bands

Numerical simulations show that, even for moderate sample sizes, when compared to the Kolmogorov–Smirnov confidence band, the CsCsHM confidence band is of the same length “in the middle” and is shorter on the tails. The new CsCsHM confidence band outperforms the Eicker–Jaeschke confidence band “in the middle” and does a similar job on the tails.

Conjecture

It is known that, as $n \to \infty$ (see Gnedenko et al. (1960)),

$$\sup_{-\infty<x<\infty} \left| \mathbf{P}_F \left(\sqrt{n} \sup_{-\infty<t<\infty} |\mathbb{F}_n(t) - F(t)| \le x \right) - \mathbf{P}_F \left(\sup_{0<F(t)<1} |B(F(t)| \le x \right) \right| = O\left(n^{-1/2}\right), \quad (3)$$

where $\{B(u), 0 \le u \le 1\}$ is a Brownian bridge. That is, under H_0, the CDF of the two-sided Kolmogorov–Smirnov statistic converges to the Kolmogorov CDF $K(x)$, uniformly in $x \in \mathbb{R}$, at the rate of $O(n^{-1/2})$. The results of numerical experiments suggest that the rate of convergence of the CDF's of the CsCsHM statistics $T_n(q)$ and $T_n^+(q)$ to their respective limit CDF's may be as good as that in (3). A theoretical justification of this claim is an open problem.

Tabulation of the the distribution of $\sup_{a<u<b} B(u)/q(u)$

1. Choose a large positive integer n. Generate n independent normal $N(0,1)$ random variables.
2. Choose a large positive integer M. Repeat step 1 M times, and for $m = 1, \ldots, M$, let $X_1^{(m)}, \ldots, X_n^{(m)}$ denote the data obtained on the mth iteration.
3. For each $m = 1, \ldots, M$, calculate the partial sums $S_k^{(m)} = \sum_{i=1}^{k} X_i^{(m)}$, $k = 1, \ldots, n$.
4. For each $m = 1, \ldots, M$, find the value of
$$T_n^{(m)} = \max_{k:k/n \in (a,b)} \frac{S_k^{(m)} - (k/n)S_n^{(m)}}{q(k/n)n^{1/2}}.$$
5. Use $G_{n,M}(x) = M^{-1} \sum_{m=1}^{M} \mathbb{I}\left(T_n^{(m)} \leq x\right)$ to approximate the limit CDF $G(x) = \mathbf{P}\left(\sup_{a<u<b} B(u)/q(u) \leq x\right)$, $x \in \mathbb{R}$.

Detection of sparse heterogeneous mixtures

An important particular case of a goodness-of-fit testing problem in high dimensions is that of detecting sparse heterogeneous mixtures. The latter problem has been extensively studied after the publications of Ingster (1997, 1999). First, consider testing the null hypothesis

$$H_0 : X_1, \ldots, X_n \overset{iid}{\sim} N(0, 1),$$

i.e., the specified CDF F_0 in the hypothesis of goodness-of-fit is the standard normal CDF, against a sequence of alternatives

$$H_{1,n} : X_1, \ldots, X_n \overset{iid}{\sim} (1 - \varepsilon_n)N(0, 1) + \varepsilon_n N(\mu_n, 1),$$

where $\varepsilon_n \sim n^{-\beta}$ for some **sparsity index** $\beta \in (1/2, 1)$ and $\mu_n = \sqrt{2r \log n}$ with $0 < r < 1$. The parameters β and r are assumed unknown, and $n \to \infty$.

Detection of sparse heterogeneous mixtures (cont-d)

In order to apply the previously developed theory to the problem of testing H_0 versus $H_{1,n}$, we transform the initial observations. Namely, let $Y_i = 1 - \Phi(X_i)$ and let $\mathcal{G}(u)$ denote a common CDF of the Y_i's taking values in $[0, 1]$. Then the problem of testing H_0 versus $H_{1,n}$ transforms to testing

$$\mathcal{H}_0 : \mathcal{G}(u) = F_0(u), \quad \text{the uniform } U(0, 1) \text{ CDF}$$

against a sequence of alternatives

$$\mathcal{H}_{1,n} : \mathcal{G}(u) = F_0(u)+\varepsilon_n \left((1 - u) - \Phi \left(\Phi^{-1}(1 - u) - \mu_n\right)\right) > F_0(u).$$

The one-sided CsCsHM test statistic takes the form

$$T_n^+(q) = \sup_{0<u<1} \sqrt{n}(\mathbb{G}_n(u) - u)/q(u),$$

where $\mathbb{G}_n(u) = n^{-1} \sum_{i=1}^n \mathbb{I}(Y_i \leq u)$.

Detection of sparse heterogeneous mixtures (cont-d)

Another model of interest, which was found to be useful in various classification problems (see, e.g., Pavlenko et al. (2012)) has the form:

$$H_0' \quad : \quad X_1, \ldots, X_n \overset{iid}{\sim} \chi^2_\nu(0),$$
$$H_{1,n}' \quad : \quad X_1, \ldots, X_n \overset{iid}{\sim} (1-\varepsilon_n)\chi^2_\nu(0) + \varepsilon_n \chi^2_\nu(\delta_n),$$

where $\chi^2_\nu(\delta)$ denotes the noncentral chi-square distribution with ν degrees of freedom and noncentrality parameter δ, $\varepsilon_n \sim n^{-\beta}$ for some $\beta \in (1/2, 1)$, and $\delta_n = 2r\log n$ for some $0 < r < 1$. For $\nu = 2$ this model connects to the problem of detecting covert communications (see Donoho & Jin (2004)). The parameters β and r are assumed unknown, and $n \to \infty$.

Detection of sparse heterogeneous mixtures (cont-d)

Now, let $S_i = 1 - H_{\nu,0}(X_i)$, where $H_{\nu,\delta}$ is the CDF of a $\chi^2_\nu(\delta)$ distribution, and let $\mathcal{H}(u)$ denote a common CDF of the S_i's. Then the problem of testing H_0' versus $H_{1,n}'$ transforms to testing

$$\mathcal{H}_0' : \mathcal{H}(u) = F_0(u), \quad \text{the uniform } U(0,1) \text{ CDF}$$

against a sequence of alternatives

$$\mathcal{H}_{1,n}' : \mathcal{H}(u) = F_0(u) + \varepsilon_n \left((1-u) - H_{\nu,\delta_n}\left(H_{\nu,0}^{-1}(1-u) \right) \right) > F_0(u).$$

The test statistic becomes

$$T_n^+(q) = \sup_{0<u<1} \sqrt{n}(\mathbb{H}_n(u) - u)/q(u),$$

where $\mathbb{H}_n(u) = n^{-1} \sum_{i=1}^n \mathbb{I}(S_i \le u)$.

Attainment of the Ingster optimal detection boundary

Next two theorems show that if the parameter r exceeds the **detection boundary** $\rho(\beta)$ obtained by Ingster (1997), which is defined by

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

then the test procedure based on the one-sided CsCsHM statistic $T_n^+(q)$ distinguishes between $\mathcal{H}_0$ and $\mathcal{H}_{1,n}$, and between $\mathcal{H}'_0$ and $\mathcal{H}'_{1,n}$. Since $T_n^+(q)$ does not require the knowledge of β and r, following Donoho & Jin (2004), we will call such a test procedure **optimally adaptive.**

Attainment of the Ingster optimal detection boundary

Theorem 1. *Consider the test of asymptotic level* α *that rejects* $\mathcal{H}_0$ *when*

$$T_n^+(q) \geq t_\alpha^+(q),$$

where the critical value $t_\alpha^+(q)$ *is chosen to have* $\mathbf{P}\left(\sup_{0<u<1} B(u)/q(u) \geq t_\alpha^+(q)\right) = \alpha$. *For every alternative* $\mathcal{H}_{1,n}$ *with* r *exceeding the detection boundary* $\rho(\beta)$, *the asymptotic level* α *test based on* $T_n^+(q)$ *has a full power, i.e.,*

$$\mathbf{P}_{\mathcal{H}_{1,n}}(T_n^+(q) \geq t_\alpha^+(q)) \to 1, \quad n \to \infty.$$

In words, when distinguishing between the null and alternative hypotheses, the test procedure based on $T_n^+(q)$ **perform optimally adaptively to unknown sparsity and size of non-null effects**.

Attainment of the Ingster optimal detection boundary

Theorem 2. *Consider the test of asymptotic level α that rejects $\mathcal{H}'_0$ when*

$$T_n^+(q) \geq t_\alpha^+(q),$$

where the critical value $t_\alpha^+(q)$ is as in Theorem 1. For every alternative $\mathcal{H}'_{1,n}$ with r exceeding the detection boundary $\rho(\beta)$, the asymptotic level α test based on $T_n^+(q)$ has a full power, that is,

$$\mathbf{P}_{\mathcal{H}'_{1,n}}(T_n^+(q) \geq t_\alpha^+(q)) \to 1, \quad n \to \infty.$$

$$* * *$$

Theorems 1 and 2 say that if $r > \rho(\beta)$, then asymptotically our test procedure based on $T_n^+(q)$ **distinguishes between $\mathcal{H}_0$ and $\mathcal{H}_{1,n}$, as well as between $\mathcal{H}'_0$ and $\mathcal{H}'_{1,n}$.**

Remark to Theorems 1 and 2

Results similar to Theorems 1 and 2 hold true for the whole class of statistics $T_n^+(q, I)$ indexed by a subinterval $I = (a, b) \subseteq (0, 1)$, in which case the critical region takes the form

$$T_n^+(q, I) \geq t_\alpha^+(q, I),$$

where $t_\alpha^+(q, I)$ is determined by

$$\mathbf{P}\Big(\sup_{a<u<b} B(u)/q(u) \geq t_\alpha^+(q, I)\Big) = \alpha.$$

In particular, this applies to our competitor of HC$_n$ given by

$$T_n^+(q, (0, \alpha_0)) = \sup_{0<F_0(t)<\alpha_0} \frac{\sqrt{n}(\mathbb{F}_n(t) - F_0(t))}{q(F_0(t))}, \quad 0 < \alpha_0 < 1,$$

which is, thus, also optimally adaptive.

Concluding remarks

We study a new family of goodness-of-fit test statistics that have the form of the empirical process in weighted sup-norm metrics with EKFP weight functions q. These statistics, which we call the CsCsHM statistics, may be viewed as competitors to the higher criticism statistic of Donoho & Jin. An immediate advantage of our approach is the identification of the proper limit distributions of the CsCsHM test statistics under the null hypothesis. These limit distributions are easily tabulated.

When compared to the higher criticism statistic HC_n^+, the one-sided CsCsHM test statistic provides a right solution in the sense that it does correctly the job that the former was intended to do without requiring a large sample size like $n = 10^6$ that, in case of the former, only indicated explosion to infinity instead of slow convergence.

Selected references

1. CSÖRGŐ, M., CSÖRGŐ, S., HORVÁTH, L., and MASON, D. (1986). Weighted empirical and quantile processes. *Ann. Probab.* **14**, 31–85.
2. DARLING, D. A. and ERDŐS, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J.* **23**, 143–145.
3. DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32**, 962–994.
4. EICKER, F. (1979). The asymptotic distribution of the suprema of the standardized empirical processes. *Ann. Statist.* **7**, 116–138.

Selected references (cont-d)

5. GNEDENKO, B. V., KOROLYUK, V. S., and SKOROHOD, A. V. (1960). Asymptotic expansions in probability theory. In *Proc. Fourth Berkley Symp. Math. Statist. Prob.* **2**, 153–169, Univ. California Press.

6. INGSTER, YU. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distribution. *Math. Meth. Statist.* **6**, 47–69.

7. INGSTER, YU. I. (2002). Adaptive detection of a signal of growing dimension. I, II. *Math. Meth. Statist.* **10**, 395–421; **11**, 37–68.

8. JAESCHKE, D. (1979). The asymptotic distribution of the supremum of the standardized empirical distribution function on subintervals. *Ann. Statist.* **7**, 108–115.

Selected references (cont-d)

9. JAGER, L. and WELLNER, J. A. (2007). Goodness-of-fit test via phi-divergences. *Ann. Statist.* **35**, 2018–2035.
10. ORASCH, M. and POULIOT, W. (2004). Tabulating weighted sup-norm functionals used in change-point problem. *J. Stat. Comput. Simul.* **74**, 249–276.