The Annals of Statistics
2015, Vol. 43, No. 3, 1323–1350
DOI: 10.1214/15-AOS1312

HIGHER CRITICISM: p-VALUES AND CRITICISM

BY JIAN LI AND DAVID SIEGMUND

Stanford University

This paper compares the higher criticism statistic (Donoho and Jin [*Ann. Statist.* **32** (2004) 962–994]), a modification of the higher criticism statistic also suggested by Donoho and Jin, and two statistics of the Berk–Jones [*Z. Wahrsch. Verw. Gebiete* **47** (1979) 47–59] type. New approximations to the significance levels of the statistics are derived, and their accuracy is studied by simulations. By numerical examples it is shown that over a broad range of sample sizes the Berk–Jones statistics have a better power function than the higher criticism statistics to detect sparse mixtures. The applications suggested by Meinshausen and Rice [*Ann. Statist.* **34** (2006) 373–393], to find lower confidence bounds for the number of false hypotheses, and by Jeng, Cai and Li [*Biometrika* **100** (2013) 157–172], to detect copy number variants, are also studied.

1. Introduction. Donoho and Jin (2004) consider the problem of deciding whether a large number, n, of independently tested null hypotheses are all true, or whether some of them are not true. They discuss in detail a suggestion of Tukey, called "higher criticism," and they prove a number of asymptotic consistency results. Suppose that $p_{(1)}, \dots, p_{(n)}$ are ordered p-values for each of the individual hypotheses, which under the global null hypotheses that all the individual null hypotheses are true, would be distributed as the order statistics of a uniform sample on [0, 1]. The test statistics of Donoho and Jin are the higher criticism (HC) statistic

$$T_{\mathrm{HC}} = n^{1/2} \max_{k_0 \le k \le k_1} (k/n - p_{(k)}) / \big[p_{(k)}(1 - p_{(k)})\big]^{1/2}, \tag{1}$$

or a modified higher criticism statistic, which we will denote by T_{MHC}. This statistic is (1) modified by the constraint that the kth term is included in the statistic only if $p_{(k)} \ge 1/n$. The recommended values for the k_i are $k_0 = 1$ and $k_1 = n/2$. These statistics reject the global null hypothesis if there is an excess of small p-values.

Donoho and Jin find values t_n such that $\mathbb{P}_0\{T_{\mathrm{HC}} \ge t_n\} \to 0$, where $\mathbb{P}_0$ denotes probability under the global null hypothesis, while under certain "borderline" alternative configurations involving the true number of nonnull hypotheses and a measure of their departures from null, $\mathbb{P}\{T_{\mathrm{HC}} \ge t_n\} \to 1$. Thus the test of the global null is consistent, provided that there is a certain minimal amount of separation between the global null hypothesis and its negation. This is a pure hypothesis testing

Received January 2014; revised October 2014.
MSC2010 subject classifications. Primary 62F10; secondary 62G20.
Key words and phrases. Global p-value, sparse mixture.

problem, in the sense that the minimal amount of separation is inadequate to allow one to identify with confidence which null hypotheses are false, although one can be confident that they exist.

In studying this and related goodness-of-fit statistics based on deviations of the empirical distribution function, an approximate p-value based on a classical result of Darling and Erdős (1956) and adapted by Jaeschke (1979) is often cited. However, since this approximation is often very poor (see below), in practice p-values are often obtained by simulation.

An alternative for relatively small sample sizes is the numerical recursion of Noé (1972), which Owen (1995) admirably exploited in finding confidence bands for a distribution function. Eiger, Nadler and Spiegelman (2013) give a substantially more efficient algorithm, which requires $O(n^2)$ operations rather than the $O(n^3)$ required by Noé's.

The goals of this paper are: (i) to give approximations for the p-value of the higher criticism type statistics that are reasonably accurate, even in the situation of small p-values and large samples, where numerical methods become onerous; (ii) to compare the power of a small number of different statistics that have the same basic properties of consistency outlined above, but may perform differently in practice; (iii) to illustrate application of our results by a discussion of two papers that have developed related ideas for specific scientific problems [viz. Meinshausen and Rice (2006); Jeng, Cai and Li (2013)].

The higher criticism statistic is suggested by standardizing $p_{(k)}$ as if it were asymptotically normally distributed even for small k. As we shall see, this can exact a considerable price on the power of the higher criticism statistic, except when the number of false null hypotheses is very small; this problem becomes very severe when the significance level is small.

We have found particularly appealing an alternative class of statistics suggested by Berk and Jones (1979) as goodness of fit statistics, defined by

$$\begin{aligned} T_{\mathrm{BJ}} = \max_{k_0 \le k \le k_1} (2n)^{1/2} \{ &(k/n) \log(k/np_{(k)}) \\ &+ (1-k/n) \log[(1-k/n)/(1-p_{(k)})] \}^{1/2}. \end{aligned} \tag{2}$$

For use in the context of higher criticism, we are interested in a one-sided version of (2) where each term is modified by the condition that $p_{(k)} < k/n$, which we henceforth assume. As explained below, a slightly modified version designed to focus on the small order statistics is

$$\begin{aligned} T_{\mathrm{MBJ}} = \max_{k_0 \le k \le k_1} &(2n)^{1/2} I\{p_{(k)} < k/n\} \\ &\times [(k/n) \log(k/np_{(k)}) - (k/n - p_{(k)})]^{1/2}, \end{aligned} \tag{3}$$

where the indicator function insures focus on an excess of small p-values.

While Berk and Jones (1979) suggest their statistic on the basis of consideration of Bahadur efficiency, our preferred motivation is in terms of Poisson variation. For example, suppose we observe a Poisson process on [0, 1] and use the log likelihood ratio statistic to test the hypothesis that the intensity is equal to one throughout the interval against the alternative that there is a change-point at t, and the intensity on $[0, t]$ is greater than one. Given that there are n events in the Poisson process, the generalized likelihood ratio statistic observed at the (ordered) times of the events is (2), where now $p_{(k)}$ denotes the time of the kth event. Alternatively, consider the empirical distribution function $F_n(x)$ for a sample of size n from the distribution $F(x)$. For small x, $F_n(x)$ behaves for large n like a nonhomogeneous Poisson process, having log likelihood $F_n(x)\log[F(x)] - F(x)$, which is maximized with respect to $F(x)$ at $F_n(x)\log[F_n(x)] - F_n(x)$. To compare the lower tail of the empirical distribution with the uniform distribution, we consider the likelihood ratio statistic $\max_x\{F_n(x)\log[F_n(x)/x] - [F_n(x) - x]\}$. At the order statistics this becomes (3). When symmetrized by consideration of both upper and lower tails, we get back to (2).

Walther (2013) gives a similar argument in favor of the Berk–Jones statistics. (He then compares the Berk–Jones statistics with a completely different class of statistics, the "average likelihood ratio" statistics, which seem to have excellent power, but do not appear to be useful when estimation of the number of nonnull distributions [Meinshausen and Rice (2006)] or identification of them is also desirable.) In the context of goodness of fit, Jager and Wellner (2007) provide asymptotic theory for a large class of statistics, including those considered here, but they do not consider the behavior of these statistics for finite sample sizes. Our methods apply to many of these statistics, and in Section 5 we discuss briefly the one that Jager and Wellner single out as perhaps a good compromise between statistics that behave well in the tails and those focusing on the center of the distribution.

The organization of the paper is as follows. In Section 2, we give expressions for approximate p-values, a heuristic argument in support of the approximations and some Monte Carlo results demonstrating their accuracy. In Section 3 we discuss comparative power. In Section 4, we revisit within our broader framework the research of Meinshausen and Rice (2006), who discuss lower confidence bounds on the number of false null hypotheses, also in a borderline case where one cannot say exactly which null hypotheses are false. We also re-examine briefly the interesting potential application suggested by Jeng, Cai and Li (2013). Proofs are given in the Appendix.

2. Approximations.

2.1. *Analytic approximations*. We begin with the specific case of (1), for which the calculations are more explicit, and then suggest the minor modifications required for the other statistics. The Appendix contains rigorous and detailed proofs for the case of T_{HC} and T_{MBJ}. The same argument with some technical augmentation works for T_{BJ}. The modified higher criticism statistic requires still slightly different arguments, which lead to slightly different approximations, given below and in the Appendix.

Let $U_{(k)}$, $k = 1, \dots, n$ denote the order statistics for a sample of size n from a uniform [0, 1] distribution. For $1 \le k_0 < k_1 < n$, let

$$Z_n = \max_{k_0 \le k \le k_1} n^{1/2}[k/n - U_{(k)}]/\big[U_{(k)}(1 - U_{(k)})\big]^{1/2}.$$

Let $C(x) = C(x, \xi) = \{x + [\xi^2 - \xi(\xi^2 + 4(1-x)x)^{1/2}]/2\}/(1+\xi^2)$, and observe that $Z_n \ge b$ if and only if $U_{(k)} \le C(k/n, b/n^{1/2})$ for some $k_0 \le k \le k_1$. Hence the problem of approximating $\mathbb{P}\{Z_n \ge b\}$ reduces to computing the sum over $k_0 \le k \le k_1$ of the probabilities of the disjoint events

$$\mathbb{P}\big\{U_{(k)} \le C(k/n), U_{(k+j)} > C\big((k+j)/n\big) \text{ for all } 1 \le j \le k_1 - k\big\}. \tag{4}$$

The distribution of $U_{(k)}$ is Beta with parameters k and $n - k + 1$.

For our approximation we assume that $n \to \infty$. From the joint distribution of $U_{(i)}, i = k, \dots, k_1$, it is easy to show by calculation that the joint conditional distribution of $n[U_{(k+j)} - U_{(k)}]$ given that $U_{(k)} \sim C(k/n)$ converges to the joint distribution of Γ_j, $j = 1, \dots, k_1 - k$, where Γ_j is the jth partial sum of independent, identically distributed exponential random variables scale parameter $\lambda = (1 - k/n)/(1 - C(k/n))$.

Let $C'(x) = 1/(1+\xi^2) - \xi(1-2x)/\{(1+\xi^2)[\xi^2 + 4x(1-x)]^{1/2}\}$ denote the derivative of $C(x) = C(x, \xi)$ with respect to x. Given $U_{(k)} = C(k/n, b/n^{1/2}) - y/n$, the conditional probability that $U_{(i)} > C(i/n, b/n^{1/2})$ for all $k < i \le k_1$ converges to

$$\begin{aligned} &\mathbb{P}\big\{\Gamma_j > jC'\big(k/n, b/n^{1/2}\big) + y \text{ for all } 1 \le j \le k_1 - k\big\} \\ &\qquad \sim \big[1 - \lambda C'(k/n)\big]\exp(-\lambda y), \end{aligned} \tag{5}$$

provided $\lambda C'(k/n) < 1$, which will be the case if C is convex, as we assume throughout.

Let $c = C(k/n)$. Direct analysis of the probability density function of $U_{(k)}$ shows that

$$\mathbb{P}\{U_{(k)} \in c - dy/n\} \sim f(c; k, n-k+1)\exp\big[-(k/n - c)y/c(1-c)\big]\,dy/n,$$

where f denotes the Beta probability density function with the indicated parameters. Integrating asymptotically over $[0, C(k/n, b/n^{1/2})]$ leads to the approximation for the term indexed by k,

$$f(c; k, n+1-k)(c/k)\big[1 - (1-k/n)c'/(1-c)\big], \tag{6}$$

where $c = C(k/n, b/n^{1/2})$, $c' = C'(k/n, b/n^{1/2})$ and $f(x; \alpha, \beta)$ denotes the Beta density with parameters α, β. Our final approximation results from summing (6) over k.

Approximations for the other statistics involve obvious modifications. For the Berk–Jones and modified Berk–Jones statistics, the curve $C(x) = C(x, \xi)$ must be found numerically, while implicit differentiation shows that $C'(x)$ is an explicit function of C and x. For example, for the modified Berk–Jones statistic, for which $c = C(x) < x$ is the solution of $x \log(x/c) - (x - c) = \xi$, by differentiation, we obtain $C'(x) = \log[x/C(x)]/[1 - x/C(x)]$. For the modified higher criticism statistic we neglect terms where $1/n \geq C(k/n, b/n^{1/2})$, and for other terms we modify the asymptotic value of the integral over $[0, C(k/n, b/n^{1/2})]$ by subtracting from it the asymptotic value of the integral over $[0, 1/n]$.

REMARK. The preceding argument is patterned after that of Woodroofe (1976), although the decomposition of a union of events uses the last event that occurs rather than the first. (Woodroofe's proof would be simplified by this change; here it appears to be necessary.) The Appendix gives a more detailed argument, which seems to be unavoidably complex due to the fact that the events indexed by small values of the subscript usually dominate the overall probability, especially in the case of the higher criticism statistic. Other techniques have been used to solve superficially similar problems, but we were unable to use these. In particular, we were unable to adapt the technique developed recently by Yakir and colleagues to solve a variety of difficult problems. See Yakir (2013) and references cited there for many examples.

2.2. *Comparison with simulations*. Table 1 contains approximate p-values evaluated by summing the terms of (6) and comparison with simulations for four different statistics: (i) the original higher criticism statistic (HC), (ii) the modification (MHC) obtained by requiring that $p_{(k)} \geq 1/n$, (iii) the (one-sided) Berk–Jones (BJ) statistic (2) and (iv) the modification suggested in (3) (MBJ).

In all cases $k_0 = 1$ and $k_1 = n/2$. The number of repetitions of the Monte Carlo experiment is 100,000, except in the rows where $n = 30{,}000$, where it is 10,000.

Our approximations appear to be very good, although slightly conservative, as one might conjecture from the derivations, which involve approximating a convex curve by a sequence of successive tangents.

As might be anticipated, the significance thresholds for the higher criticism statistic increase very rapidly for decreasing significance levels. As we will see below, this is the price that the statistic pays to be able to detect very rare false null hypotheses. But at very small significance levels, which are appropriate for the application in Section 4.2, the threshold becomes prohibitively large, unless one takes $k_0 > 1$, which calls into question the advantage of HC for rare false null hypotheses.

2.3. *Other approximations*. Diverse scientists writing about various aspects of the problems considered in this paper and other related (often goodness of fit) problems mention approximations based on the double exponential extreme value

TABLE 1
p-values

Statistic	Threshold	n	Approximate p-value	Simulation
HC	4.83	400	0.05	0.048
HC	10.0	400	0.01	0.01
HC	10.0	1000	0.01	0.010
HC	10.0	5000	0.01	0.010
HC	10.0	30,000	0.01	0.010
HC	31.0	1000	0.001	0.0009
MHC	3.13	400	0.05	0.053
MHC	3.91	400	0.01	0.010
MHC	3.94	1000	0.01	0.0101
MHC	3.98	5000	0.01	0.0098
MHC	4.00	30,000	0.01	0.010
MHC	4.97	1000	0.001	0.0010
BJ	2.90	400	0.05	0.048
BJ	3.45	400	0.01	0.010
BJ	3.50	1000	0.01	0.0095
BJ	3.57	5000	0.01	0.0098
BJ	3.63	30,000	0.01	0.0096
BJ	4.14	1000	0.001	0.0009
MBJ	2.80	400	0.05	0.046
MBJ	3.35	400	0.01	0.0094
MBJ	3.40	1000	0.01	0.0094
MBJ	3.48	5000	0.01	0.0098
MBJ	3.56	30,000	0.01	0.0090
MBJ	4.04	1000	0.001	0.0009

distribution and attributed to Jaeschke (1979), who adapted the original result of Darling and Erdős (1956). An intermediate step in deriving this particular approximation involves the relation of the uniform empirical process to a Brownian bridge, a step that makes the approximation suspect, since the standard empirical distribution at small (or large) values of its argument exhibits Poisson, not Gaussian, variability.

Let $U(t)$ denote a stationary Ornstein–Uhlenbeck process with covariance function $\exp(-|t|)$. Let $B_0(t)$ denote a Brownian bridge process on $[0,1]$ and $W(t)$ a standard Brownian motion process on $[0,\infty)$. Let $0<\tau_0<\tau_1<1$ and put $T_0 = 0.5\log[\tau_1(1-\tau_0)/\tau_0(1-\tau_1)]$. A Gaussian approximation that plays a role in the derivation of the Darling–Erdős result is given by

$$\mathbb{P}\Big\{\max_{\tau_0\le t\le\tau_1} B_0(t)/\big[t(1-t)\big]^{1/2}\ge b\Big\} = \mathbb{P}\Big\{\max_{0\le t\le T_0} U(t)\ge b\Big\} \sim T_0 b\varphi(b), \tag{7}$$

as $b\to\infty$, provided T_0 is bounded or diverges slowly enough that the RHS of (7) tends to 0. Here φ is the standard normal probability density function. For maxima over longer intervals, say $[0, T]$, where T is so large that $Tb\varphi(b)$ tends to a positive limit, which we find it convenient to specify as $\exp(-x)$, it is easily shown by consideration of T/T_0 approximately independent excursions of length T_0 that (7) implies $\mathbb{P}\{\max_{0 \le t \le T} U(t) < b\} \to \exp(-\exp(-x))$. If we put $T = \log(N)/2$, use the relationship between the stationary Ornstein–Uhlenbeck process and the Wiener process that $U[\log(t)/2] = W(t)/t^{1/2}, 1 \le t < \infty$, and choose one particular asymptotic inversion of the limiting relationship defined by $0.5\log(N)b\varphi(b) \to \exp(-x)$ to obtain b as a function of N, we obtain the classical Darling–Erdős approximation. Most authors concede that the approximation is very slow to converge. Meinshausen and Rice (2006) suggest that the approximation only be applied in the case of the modified higher criticism statistic, and in this case it is not unreasonable. For example, for the thresholds in rows 7–12 of Table 1, this approximation would give 0.036, 0.008, 0.008, 0.008, 0.008 and 0.001.

For the same reasons that the Darling–Erdős approximation seems to be reasonable only in special cases, in particular for the modified higher criticism statistic, the approximation (7) is roughly correct for the Berk–Jones statistic and slightly less so for the modified Berk–Jones statistic, but not otherwise. For example, for the thresholds in rows 13–18 of Table 1, this approximation gives the p-values of 0.054, 0.01, 0.01, 0.01, 0.01, 0.001. In fact, a different inversion of b as a function of N gives the approximation suggested by Wellner and Koltchinskii (2003) and studied numerically by Walther (2013), which, as he shows, performs reasonably well for the Berk–Jones statistic, although not quite as well as direct application of (7) in the cases we have tested.

It seems fair to say that there is enough latitude in performing this inversion that one can frequently choose an approximation that seems to apply to a particular problem. It also seems clear from the results in Table 1 that no single approximation of this type can be applied successfully to all four statistics, which while asymptotically equivalent in the sense of Donoho and Jin (2004), require quite different thresholds to control their false positive probabilities.

3. Power. In this section we consider the power under the commonly used alternative model that the data arise from a mixture of a null (often normal) distribution and a shifted version of the null distribution.

3.1. *Analytic approximations.* By a straightforward transformation the evaluation of the power, that is, the probability under a mixture model of rejecting the global null, can be reduced to a problem of the same structure as calculation of the global significance level, with one important difference. Suppose that X_i are independent samples with the distribution function $F(x) = (1-p)F_0(x) + pF_0(x-\delta)$, where we consider in detail the case $F_0(x) = \Phi(x)$, the distribution function of the standard normal distribution and $\delta > 0$. The p-values are $p_j = 1 - \Phi(X_j)$. We obtain independent random samples from the uniform [0, 1] distribution by the following transformation:

$$
\begin{aligned}
U_i &= 1 - F(X_i) \\
&= 1 - F\big[\Phi^{-1}(1 - p_i)\big] \\
&= (1 - p)p_i + p\big\{1 - \Phi\big[\Phi^{-1}(1 - p_i) - \delta\big]\big\}.
\end{aligned}
\tag{8}
$$

Notice that U_i is an increasing function of p_i, so the order statistics of U_i correspond directly to those of p_i via equation (8). Let $d(i/n)$ denote the transformed boundary $(1 - p)C(i/n) + p\{1 - \Phi[\Phi^{-1}(1 - C(i/n)) - \delta]\}$. The global null is rejected if and only if $U_{(k)} \le d(k/n)$ for some $k_0 \le k \le k_1$.

This new curve d is not globally convex, so the argument of Section 2.1 fails here. However, the curve is concave near 0, and becomes convex after some point j_0. Besides, $\{U_{(i)}\}_{i<j_0}$ and $\{U_{(i)}\}_{i>j_0}$ are conditionally independent given $U_{(j_0)}$. Therefore we have

$$
\begin{aligned}
&\mathbb{P}\big\{U_k \le d(k/n) \text{ for some } k_0 \le k \le k_1\big\} \\
&\quad = \mathbb{P}\big\{U_{j_0} \le d(j_0/n)\big\} + \int_{d(j_0/n)}^{1} f_{U_{(j_0)}}(x)\big[g_1(x) + g_2(x) - g_1(x)g_2(x)\big]\,dx,
\end{aligned}
$$

where $g_1(x) = \mathbb{P}\{U_{(k)} \le d(k/n) \text{ for some } k_0 \le k < j_0 | U_{(j_0)} = x\}$ and $g_2(x) = \mathbb{P}\{U_{(k)} \le d(k/n) \text{ for some } j_0 < k \le k_1 | U_{(j_0)} = x\}$. Then we approximate g_2 with the results in Section 2.1, and compute g_1 by Noé's recursion [Noé (1972)]. We observe fairly small j_0 for (p, δ) that gives moderate values of power, leading to fast and accurate implementation of Noé's method, which encounters computational difficulties for large n if used by itself.

To illustrate the approximation above, we consider the higher criticism and modified Berk–Jones statistics for $n = 1000$, $\delta = 2.5$ and $p = 0.02$, also for $\delta = 4$ and $p = 0.005$. The significance level is 0.01. The values for the power we obtained are 0.68 and 0.89 for higher criticism, and, respectively, 0.90 and 0.87 for the modified Berk–Jones statistic. Simulations with 10,000 repetitions gave exactly the same values to two significant figures. It may be interesting to note that the terms contributing substantially to the power have indices considerably smaller than np, the expected number of nonnull distributions. For the higher criticism statistic, most of its power concentrates on the first-order statistic, which is the reason why it often performs poorly when the number of nonnull distributions is not very small.

We have used Noé's method because it is easy to understand and apply. In a recent manuscript Eiger, Nadler and Spiegelman (2013) describe an alternative, which after a number of numerical refinements to improve its accuracy appears to be substantially faster [$O(n^2)$ instead of $O(n^3)$ operations] and hence suitable for larger sample sizes.

3.2. *Power comparison*. Now we compare by simulation the power of the four statistics discussed above under the mixture model. While the analytic method may offer computational advantages for very large n, simulation has a number of compensating advantages for estimating power, when the probabilities of interest are not small and a general idea of their magnitude usually suffices, so very large sample sizes are rarely required. One particular advantage in the case of interest here is that statistical tests can be one-sided or two-sided, and not only the number of nonnull distributions, but also their noncentrality parameters can be variable at essentially no increase in computational effort.

In Table 2 we compare by simulation the power of the four statistics. The listed thresholds correspond to examples in Table 1. The alternative model is a mixture of N(0, 1) and N(δ, 1) distributions. The mixing parameter is p; the δ's are independent and have a N(μ, 0.1) distribution. The p-values are two-sided. The number of repetitions of the simulation experiment was 10,000, except for the last four scenarios, where it was reduced to 1000. If instead of simulating a binomial(p) number of nonnull distributions, we take a deterministic $k = np$ number of nonnull distributions, the power typically increases by roughly 10%, except in some cases when p is very small. The general picture that emerges is that at conventional levels of significance, for very small p the HC statistic can have about 5–8% more power than the two Berk–Jones statistics, which in turn have considerably more power than the modified higher criticism statistic. For larger p, the two Berk–Jones statistics and the modified higher criticism statistic can have about equal power and substantially more power than the original higher criticism statistic. At very small levels of significance, for example, the level used in Section 4.2 where the higher criticism statistic enters into a multiple comparisons analysis, the HC statistic with the recommended $k_0 = 1$ can have very little power, even for very small p.

4. Applications.

4.1. *Example*: *Confidence bounds for the proportion of false null hypotheses*. Consider a mixture model where $N\lambda$ of the hypotheses are false. Meinshausen and Rice (2006) give a lower confidence bound for λ, which is based on a functional of the process $[F_n(t) - t]/\delta(t)$, where $F_n(t)$ is the empirical distribution function of the p-values, and $\delta(t)$ is a suitable function chosen by the user. They suggest the choice $\delta(t) = [t(1-t)]^{1/2}$, which is closely related to the higher criticism statistic, and they observe that the choice $\delta(t) = t$ is similarly related to the Benjamini and Hochberg (1995) false discovery rate criterion. A similar lower confidence bound can be obtained from the (modified) Berk–Jones statistic, which in view of the power calculations of the preceding section, one might expect to behave comparably or perhaps even better than the higher criticism statistic. This subsection will compare lower confidence bounds for the Gaussian mixture model.

TABLE 2
Power

Statistic	n	Threshold	μ	p	Power
HC	400	4.83	4.0	0.01	0.91
BJ		2.90			0.87
MHC		3.13			0.51
MBJ		2.80			0.88
HC	400	4.83	1.5	0.1	0.54
BJ		2.90			0.76
MHC		3.13			0.73
MBJ		2.80			0.79
HC	1000	10.0	1.5	0.08	0.19
BJ		3.50			0.82
MHC		3.94			0.81
MBJ		3.40			0.81
HC	1000	10.0	4.0	0.005	0.85
BJ		3.50			0.81
MHC		3.94			0.43
MBJ		3.40			0.83
HC	1000	31.0	5.0	0.002	0.67
BJ		4.14			0.60
MHC		4.97			0.04
MBJ		4.04			0.62
HC	1000	31.0	2.0	0.05	0.11
BJ		4.14			0.79
MHC		4.97			0.78
MBJ		4.04			0.80
HC	5000	10.0	4.0	0.001	0.71
BJ		3.57			0.65
MHC		3.98			0.32
MBJ		3.48			0.66
HC	5000	10.0	3.0	0.003	0.53
BJ		3.57			0.62
MHC		3.98			0.55
MBJ		3.48			0.63
HC	5000	10.0	1.0	0.08	0.05
BJ		3.57			0.81
MHC		3.98			0.74
MBJ		3.48			0.78
HC	30,000	10.0	3.0	0.001	0.51
BJ		3.63			0.66
MHC		4.00			0.65
MBJ		3.56			0.68
HC	30,000	10.0	2.0	0.005	0.17
BJ		3.63			0.68
MHC		4.00			0.68
MBJ		3.56			0.69

We only consider the modified version, as described in Section 1, so in this section we suppress the word modified in the description.

Assume that (a) $n\gamma_{n,\alpha}$ is increasing in n, and that (b) under the global null hypothesis $\mathbb{P}_0(\sup_{t\in\{s\in(0,1)|F_n(s)\geq s\}}[F_n(t)(\log F_n(t) - \log t) - (F_n(t) - t)] > \gamma_{n,\alpha}) \leq \alpha$ for all n.

Define

$$I_{\mathrm{BJ}} = \left\{\lambda \Big| \sup_{t:F_n(t)-\lambda\geq(1-\lambda)t} (F_n - \lambda)\log\frac{F_n - \lambda}{(1-\lambda)t} - \left[F_n - \lambda - (1-\lambda)t\right] > \gamma_{n,\alpha}\right\}$$

and $\hat{\lambda}_{\mathrm{BJ}} = \sup I_{\mathrm{BJ}}$. Then $\hat{\lambda}_{\mathrm{BJ}}$ can be shown to be a lower confidence bound for λ at confidence level $1-\alpha$.

The proof of this result is similar to the argument given by Meinshausen and Rice (2006) and hence is omitted. To compute the required probability, we suggest using the approximation obtained above. If $\gamma_{n,\alpha}$ is the (approximate) $(1-\alpha)$ level quantile of the quantity in condition (b), the required monotonicity condition (a) is satisfied numerically.

The simulation study reported below compares these two lower confidence bounds. The underlying observations, X_i, $i = 1, 2, \dots, N$, are independently and normally distributed. A fraction λN have a mean of μ while the others have a mean of 0. The lower confidence bounds, $\hat{\lambda}_{\mathrm{MHC}}$ is calculated according to the prescription of Meinshausen and Rice (2006), and that for $\hat{\lambda}_{\mathrm{MBJ}}$ is calculated according to the prescription in the preceding paragraph. The confidence level is 95%. The "bounding sequence" of Rice and Meinshausen, $\beta_{n,\alpha}$ and our corresponding $\gamma_{n,\alpha}$ are determined by the approximations given above. Multiple configurations of model parameters, N, λ and μ, are considered, and the simulations are repeated 10^5 times for each configuration.

Numerical results for a sample size of $N = 400$ are provided in Tables 3 and 4. When the signal is weak, both methods can give a lower bound of 0, and hence the sum of columns three and four can be less than one. Two criteria are considered: (a) the larger of the two lower confidence bounds, and (b) the relative squared distances of the bounds from the true parameter. According to the table, $\hat{\lambda}_{\mathrm{HC}}$ has an advantage over $\hat{\lambda}_{\mathrm{BJ}}$ when the values of μ are small. As μ increases, $\hat{\lambda}_{\mathrm{BJ}}$ first becomes more precise in probability, and then it lies closer to λ than $\hat{\lambda}_{\mathrm{HC}}$ does in (relative) l_2 distance. It should be noted that even when $\mathbb{P}(\hat{\lambda}_{\mathrm{BJ}} > \hat{\lambda}_{\mathrm{HC}})$ is quite large, $\hat{\lambda}_{\mathrm{BJ}}$ may not be as precise as $\hat{\lambda}_{\mathrm{HC}}$ in l_2 (e.g., $\lambda = 0.1$, $\mu = 2$ and $\lambda = 0.2$, $\mu = 1.5$). Therefore when λ is small $\hat{\lambda}_{\mathrm{HC}}$ is better (in probability and/or in l_2), while as λ exceeds some critical value λ^*, $\hat{\lambda}_{\mathrm{BJ}}$ becomes a tighter lower bound. As can be seen in Table 3, the borderline value of λ^* for the probability comparison seems to be rather stable for different values of λ, whereas the analogous λ^* for l_2 distance decreases slowly from above 2.0 to below 1.5 as λ varies from 0.1 to 0.5. Moreover, when the individual signal strength is weak ($\mu = 1, 1.5$), neither $\hat{\lambda}_{\mathrm{HC}}$ nor $\hat{\lambda}_{\mathrm{BJ}}$ works well unless λ is about 0.3 or even larger; and in this case the difference

TABLE 3
Comparison between $\hat{\lambda}_{\mathrm{HC}}$ and $\hat{\lambda}_{\mathrm{BJ}}$, $N = 400$

λ	μ	$P(\hat{\lambda}_{\mathbf{HC}} > \hat{\lambda}_{\mathbf{BJ}})$	$P(\hat{\lambda}_{\mathbf{HC}} < \hat{\lambda}_{\mathbf{BJ}})$	$\Vert\lambda - \hat{\lambda}_{\mathbf{HC}}\Vert_2/\lambda$	$\Vert\lambda - \hat{\lambda}_{\mathbf{BJ}}\Vert_2/\lambda$
0.1	1.0	0.36	0.28	0.93	0.94
0.1	1.5	0.43	0.55	0.79	0.89
0.1	2.0	0.29	0.71	0.61	0.61
0.1	2.5	0.26	0.74	0.46	0.45
0.1	3.0	0.37	0.63	0.33	0.32
0.2	1.0	0.59	0.40	0.80	0.83
0.2	1.5	0.40	0.60	0.60	0.62
0.2	2.0	0.17	0.83	0.44	0.44
0.2	2.5	0.06	0.94	0.32	0.31
0.2	3.0	0.03	0.97	0.23	0.21
0.3	1.0	0.74	0.26	0.69	0.73
0.3	1.5	0.45	0.55	0.50	0.51
0.3	2.0	0.14	0.86	0.36	0.35
0.3	2.5	0.03	0.97	0.26	0.24
0.3	3.0	0.001	0.99	0.18	0.16
0.4	1.0	0.82	0.18	0.62	0.66
0.4	1.5	0.42	0.58	0.43	0.44
0.4	2.0	0.09	0.91	0.31	0.29
0.4	2.5	0.02	0.99	0.22	0.19
0.4	3.0	0.003	0.99	0.15	0.13
0.5	1.0	0.84	0.16	0.57	0.61
0.5	1.5	0.34	0.66	0.39	0.39
0.5	2.0	0.04	0.96	0.27	0.25
0.5	2.5	0.003	1.0	0.19	0.16
0.5	3.0	0.0003	1.0	0.14	0.11

between the two lower bounds does not seem important compared to the gap between the confidence bounds and the true λ. This numerical behavior suggests $\hat{\lambda}_{\mathrm{BJ}}$ is preferable unless either prior information indicates a weak individual signal in the data or the worst case scenario is of primary concern.

4.2. *A more complex example.* Motivated by the problem of detecting intervals of copy number variation (CNV) occurring at the same location in a (usually small) fraction of aligned DNA sequences, Jeng, Cai and Li (2013) suggest use of a higher criticism based analysis as an alternative to the method suggested by Zhang et al. (2010) and Siegmund, Yakir and Zhang (2011). In brief, for each $n = 1, \ldots, N$, observations $y_{n,t}$, $t = 1, \ldots, T$ are independently and normally distributed with constant known variances σ_n^2 and means that under the null hypothesis are unknown constants μ_n, but are different by an increment $\delta_{n,I}$ in aligned short subintervals $I \subset \{1, \ldots, T\}$. The subset of $1, \ldots, N$ that exhibit changes in mean value in any particular interval I is usually relatively small. Jeng, Cai and

TABLE 4
Comparison between $\hat{\lambda}_{\mathrm{HC}}$ and $\hat{\lambda}_{\mathrm{BJ}}$, $\lambda = 0.2$

λ	μ	$P(\hat{\lambda}_{\mathbf{HC}} > \hat{\lambda}_{\mathbf{BJ}})$	$P(\hat{\lambda}_{\mathbf{HC}} < \hat{\lambda}_{\mathbf{BJ}})$	$\Vert\lambda - \hat{\lambda}_{\mathbf{HC}}\Vert_2/\lambda$	$\Vert\lambda - \hat{\lambda}_{\mathbf{BJ}}\Vert_2/\lambda$
400	1.0	0.59	0.40	0.80	0.83
400	1.5	0.40	0.60	0.60	0.62
400	2.0	0.17	0.83	0.44	0.44
400	2.5	0.06	0.94	0.32	0.31
400	3.0	0.03	0.97	0.23	0.21
800	1.0	0.70	0.30	0.71	0.75
800	1.5	0.49	0.51	0.51	0.53
800	2.0	0.20	0.80	0.37	0.37
800	2.5	0.06	0.94	0.27	0.25
800	3.0	0.02	0.98	0.19	0.17
1200	1.0	0.81	0.19	0.66	0.71
1200	1.5	0.60	0.40	0.47	0.49
1200	2.0	0.25	0.75	0.33	0.33
1200	2.5	0.07	0.93	0.24	0.22
1200	3.0	0.03	0.97	0.17	0.15

Li's method is, roughly speaking, to consider an interval $I \subset \{1, \dots, T\}$ having length at most L. They then apply a higher criticism based analysis across the N sequences to a statistic (in this case the sample mean) defined on the interval I. Large values of the higher criticism on various intervals is interpreted as evidence that those intervals contain CNV.

To control the false positive error rate, they suggest using the approximation of Jaeschke referenced above for each candidate interval in conjunction with a Bonferroni bound (multiplication by TL) to account for multiple comparisons involving overlapping candidate intervals of different lengths. For their actual analysis they use simulations. The number of repetitions of their simulation experiments is 100 for a small set of data set and 50 for a larger set of data.

For this problem, N is often in the hundreds, L is usually relatively small while T can be in the tens or hundreds of thousands.

Here we present a different simulation to compare a higher criticism based procedure, along the lines suggested by Jeng, Cai and Li (2013), a modified higher criticism based procedure, and its modified Berk–Jones counterpart. The type I error is set to be approximately 0.05. The other parameters are $N = 674$, $L = 20$ and $T = 40{,}929$. Since the higher criticism statistic is extremely sensitive to the value of k_0 in (1), we follow the suggestion of Jeng, Cai and Li (2013) and set $k_0 = 4$.

Although Jeng, Cai and Li (2013) use this example to illustrate their methods on real data, for our comparative numerical experiment, the data are similar in structure, but are artificially generated. The number of intervals $I = [\tau_1, \tau_2]$ that contain signals is 155, of which 75 have a length of 3, 50 have a length of 4, 25 have a length of 7 and 5 have a length of 10. The model has two variable parameters:

TABLE 5
Power comparison

μ	p	Power using HC ($k_0 = 4$)	Power using MHC	Power using MBJ
1	0.05	0.11	0.16	0.17
1	0.06	0.14	0.21	0.22
1	0.07	0.17	0.25	0.27
1	0.08	0.19	0.29	0.34
1	0.09	0.21	0.35	0.42
1.5	0.01	0.12	0.03	0.12
1.5	0.02	0.26	0.19	0.28
1.5	0.03	0.39	0.39	0.47
1.5	0.04	0.53	0.60	0.67
2	0.01	0.41	0.10	0.42
2	0.02	0.79	0.55	0.81

given that a particular interval contains at least one signal, p is the fraction of the N intervals that contain the signal, and μ is the change in mean values of the observed Gaussian random variables in the interval and sequence that contain the signal. The thresholds of the (modified) higher criticism based procedure and the (modified) Berk–Jones statistic are determined by simulations repeated 900 times, and are compared with our approximations. The number of repetitions in the power computation is 625.

The significance thresholds obtained by simulation are as follows (with theoretically calculated thresholds in parentheses): (a) for a global false positive error rate of 0.05, HC 20.0(21.5), MHC 9.3(9.1), MBJ 5.98(5.98); (b) for a global false positive error rate of 0.01, HC 24.1(26.0), MHC 9.84(9.79), MBJ 6.29(6.24). Even though our theoretically determined thresholds are in principle conservative because of an inclusion of a Bonferroni bound in the argument, the approximations in these examples appear to be very good. This may not continue to be the case for larger values of L.

Table 5 shows the power of the three procedures under different data configurations. Here power is taken to be the fraction of intervals containing signals that are detected. Generally speaking the two higher criticism statistics have poor performance for certain parameter combinations, small p for MHC and not so small p but small μ for HC, while the MBJ statistic maintains good power througout the table.

As a final example, we compare the method of Jeng, Cai and Li (2013), based on the modified Berk–Jones statistic, and a method suggested without further study by Siegmund, Yakir and Zhang (2011), here denoted SYZ. Their method contains a free parameter p_0, which can be loosely interpreted as a prior expectation of the fraction of the N intervals that contain a signal whenever one is present. Jeng, Cai and Li claim that their method is better at detecting *both* rare and common signals

TABLE 6
Power comparison

ξ	p	**Power using SYZ**	**Power using MBJ**
5.0	0.005	0.62	0.52
4.0	0.01	0.62	0.47
2.5	0.05	0.77	0.53
2.0	0.10	0.75	0.56
1.5	0.2	0.95	0.75
1.0	0.4	0.75	0.56
0.9	0.5	0.79	0.62

than a fixed value of p_0. SYZ's suggestion for making their method more robust against an incorrect choice of p_0 was to use two different values, at say significance level 0.025, so that by Bonferroni the overall significance level is $2 \times 0.025 = 0.05$. See also Xie and Siegmund (2013), who tested this suggestion in a somewhat different context. Here we consider the case where there is only a single interval I containing a signal, which has various expected frequencies p and noncentrality parameters $\xi = |I|\mu$ chosen so that the power is intermediate between 0 and 1. The parameters p_0 of the SYZ procedure are chosen to equal 0.005 and 0.2, for which 0.025 significance thresholds are 25.3 and 179.2, respectively. Table 6 gives Monte Carlo estimates of the marginal power, that is, the probability that the statistic computed from observations from the correct interval I exceed the appropriate significance threshold.

5. Discussion. We have derived an approximation to the significance level of higher criticism like statistics that appears to be sufficiently accurate for use in practice and for theoretical comparisons of the power of different statistics. As an alternative to the two higher criticism statistics suggested by Donoho and Jin (2004), we have also studied two statistics motivated by the goodness-of-fit procedure suggested by Berk and Jones (1979). In a normal mixture model, the Berk–Jones statistics have more power than the higher criticism statistic, except when the mixing fraction is very small, and more power than the modified higher criticism statistic when the mixing fraction is small. Even in cases where the Berk–Jones statistics have less power than one of the higher criticism statistics, the differences are only a few percent. The advantages of the Berk–Jones statistics are larger at smaller significance levels. Since the significance threshold of the original higher criticism statistic is extremely sensitive to the significance level, when the test is an intermediate part of a large multiple comparison problem (cf. Section 4.2) and hence involves a very small significance level, its power can be much less than that of the other statistics. This problem can be mitigated by taking a value $k_0 > 1$ in definition (1), but this deletes the capacity of the higher criticism statistic to detect very rare mixtures. For the range of parameter values we have studied, the two Berk–Jones statistics seem to be unequivocally better.

The statistics we have studied are related to goodness-of-fit tests based on the empirical distribution function; but for the higher criticism problem, as suggested by Donoho and Jin (2004) (and the applications discussed in Section 4), we have focused on one-sided statistics designed to detect an excess of small p-values. Jager and Wellner (2007) develop an elegant large sample theory for a large class of statistics, but they do not show how well their asymptotic theory predicts behavior for sample sizes of interest in practice.

One statistic that receives particular mention by Jager and Wellner as a perhaps reasonable compromise between statistics focusing on the center of the distribution and statistics focusing on the tails is (after modification to focus on an excess of small p-values) $\max_{k_0 \le j \le k_1} n^{1/2}\{[(j/n)^{1/2} - p_{(j)}]^+\}^{1/2}$. This statistic has the appealing feature that C and C' are given explicitly by $C(x, \xi) = [(x^{1/2} - \xi)^+]^2$ and $C'(x, \xi) = (1 - \xi/x^{1/2})^+$. Our methods apply and give good approximations (compared to simulations) for the significance threshold. For the examples in Table 2, we find that the statistic behaves well for values of p that are not too small. It is usually more powerful than the modified higher criticism statistic, but it has considerably less power than the original higher criticism statistic and both Berk–Jones statistics for small p. For example, for the third to fifth examples in Table 2, we find by summation of (6) that the threshold $b = 1.54$ gives the same level, 0.01 for $n = 1000$, as the examples given there, and we obtain as estimates of the power 0.84, 0.57 and 0.27, respectively. For the seventh to ninth examples, the appropriate threshold is 1.62, and the power is 0.34, 0.47 and 0.82.

It might be interesting to see more systematically whether our methods can be usefully applied in a goodness-of-fit context, for example, as they might be applied to give confidence bands for a distribution function, as in Owen (1995).

APPENDIX: PROOFS

The heuristic argument given above for our suggested approximations is based on the approach of Woodroofe (1976) and uses results obtained in a similar problem by Loader (1992). Although the heuristic is relatively simple, complete proofs are quite technical, and alternative approaches that have been proved successful in apparently similar problems do not seem to work here. The source of the difficulties is the requirement that we not impose a lower bound on k_0 and want k_1 to be of order n. In addition, different statistics require somewhat different techniques. Here we consider in detail the original higher criticism statistic and the modified Berk–Jones statistic.

We can obtain the analogous approximations for the original Berk–Jones statistic by similar methods (after some additional technical arguments to verify the general conditions stated below in Remark A.1) and for the Jager–Wellner statistic mentioned briefly in Section 5. For the modified higher criticism statistic, we obtain by similar calculations a slightly different approximation given explicitly at the end of this appendix.

Consider the following two functions:

$$f_1(x, y) = \frac{x - y}{[y(1 - y)]^{1/2}},$$

$$f_2(x, y) = 2\left[x \log \frac{x}{y} - (x - y)\right].$$

For each x, ξ let $C_i(x, \xi)$ denote the root $y \in (0, x)$ of $f_i(x, y) = \xi^i$. Then C_1 and C_2 correspond to the original higher criticism statistic and modified Berk–Jones statistic, respectively.

Although C_i is a function of two arguments, in most cases the second argument will be $b/n^{1/2}$. When this is the case we will simplify the notation by writing $C_i(k/n)$.

REMARK A.1. We will see the proofs below hold in general for a piecewise differentiable function $C(x)$, $x \in [0, 1]$ satisfying the following conditions:

(i) $C(x)$ is convex and $C(x) < x$ in the region of interest, $x = 0$ excluded;
(ii) for some $\alpha \in (0, 1)$, $C(k/n) \leq (k - 1)/(n - 1)$ for all $k \in [2, \alpha n]$ when n is large enough;
(iii) for some $\alpha \in (0, 1)$, $\sup_{0 \leq x \leq \alpha}(1 - x)C'(x)/[1 - C(x)] < 1$;
(iv) $\lim_{x \to 0^+} C(x)/x = 0$.

The probabilities of rejecting the global null hypothesis with higher criticism or Berk–Jones statistic have a similar expression.

$$\mathbb{P}\Big(\max_{k_0 \leq k \leq \beta n} n^{1/2} f_1(k/n, p_{(k)}) \geq b\Big) = \mathbb{P}\left(\bigcup_{k=1}^{\beta n} \{p_{(k)} \leq C_1(k/n)\}\right)$$

and

$$\mathbb{P}\Big(\max_{k_0 \leq k \leq \beta n, p_{(k)} \leq k/n} n f_2(k/n, p_{(k)}) \geq b^2\Big) = \mathbb{P}\left(\bigcup_{k=1}^{\beta n} \{p_{(k)} \leq C_2(k/n)\}\right).$$

If k_0 is proportional to n, the following division of the rejection region is unnecessary, and Proposition A.1 can be directly applied; otherwise the rejection region should be divided into two parts [equation (9)], and their probabilities are computed by different means.

There is an additional difficulty involving the value of k_1. For our purposes $k_1 = n/2$ is the primary case of interest, and in the following we take $k_1 = \beta n$ for a value $\beta < 1$. In some cases, it is possible to take $k_1 = n - 1$, and in others, this imposes additional constraints on k_0; for instance, a sufficient condition for higher criticism statistic is to have a constant k_0, and in still others the constraints are unclear.

$$\mathbb{P}\left(\bigcup_{k=k_0}^{\beta n}\{p_{(k)} \le C(k/n)\}\right)$$
$$= \mathbb{P}\left(\bigcup_{k=k_0}^{\alpha n-1}\{p_{(k)} \le C(k/n)\} \Big\backslash \left(\bigcup_{k=\alpha n}^{\beta n}\{p_{(k)} \le C(k/n)\}\right)\right) \tag{9}$$
$$+ \mathbb{P}\left(\bigcup_{k=\alpha n}^{\beta n}\{p_{(k)} \le C(k/n)\}\right).$$

The rejection regions of a large class of statistics, including the higher criticism and Berk–Jones statistics, correspond to a collection of curves $\{C(x)\}$, each of which satisfies $C(\beta) < \beta$ as well as $C(0) = 0$. Consequently there exists $\alpha \in (0, 1/2)$ such that $C(\beta) - C(\alpha) < \beta - \alpha$, and the α in (9) satisfies these conditions.

Let $\mathbb{P}_{\text{Bin}}(n, k, p) = \binom{n}{k} p^k (1-p)^{n-k}$ denote the binomial probability distribution, $C'(x) = \partial C(x, \xi)/\partial x$. Proposition A.1 below handles the second term of equation (9).

PROPOSITION A.1. *Suppose that for every $\xi > 0$, $C(x, \xi)$ is a convex and continuously differentiable function of x, $C(x, \xi)$ is increasing in x and $C(x, \xi) < x$ for all $x \in [\alpha, \beta]$. Then*

$$\mathbb{P}\left(\bigcup_{k=\alpha n}^{\beta n}\{p_{(k)} \le C(k/n)\}\right)$$
$$= (1 + o(1)) \sum_{k=\alpha n}^{\beta n}\left[1 - \frac{(n-k+1)C'(k/n)}{n - nC(k/n)}\right]\mathbb{P}_{\text{Bin}}(n, k, C(k/n)).$$

PROOF. Let $F_n(x)$ be the empirical distribution function associated with the independent p-values $p_1, p_2, \dots, p_n$, and let $D(x)$ be the inverse of $C(x, \xi)$ with respect to x, that is, $D(C(x, \xi)) = x$. Then

$$\mathbb{P}\left(\bigcup_{k=\alpha n}^{\beta n}\{p_{(k)} \le C(k/n)\}\right)$$
$$= \mathbb{P}(F_n(x) \ge D(x) \text{ for some } x \in \{p_{(\alpha n)}, p_{(\alpha n+1)}, \dots, p_{(\beta n)}\})$$
$$= \mathbb{P}(F_n(x) \ge D(x) \text{ for some } x \in [C(\alpha, \xi), C(\beta, \xi)])$$
$$= \mathbb{P}(F_n(x) \le 1 - D(1-x) \text{ for some } x \in [1 - C(\beta, \xi), 1 - C(\alpha, \xi)]).$$

The last equation results from the symmetry of $F_n(x)$, that is, $(\{F_n(x)\}_{x\in[0,1]} \overset{d}{=} \{1 - F_n(1-x)\}_{x\in[0,1]})$. The desired result now follows from the proof of Theorem 2.1 of Loader (1992). □

REMARK A.2. The summand in the formula in Theorem 2.1 of Loader (1992) converges uniformly, so when αn is not an integer, it could be replaced by $\lceil \alpha n \rceil$ or $\lfloor \alpha n \rfloor$, and the same goes for βn.

The rest of the proofs show the first term on the RHS of (9) has the identical expression. The event in this term decomposes into disjoint sub-events. Let $B_{n,k} = \{p_{(k)} \le C(k/n), p_{(k+j)} > C[(k+j)/n]\ \forall j = 1, 2, \dots, \beta n - k\}$. Then this term equals $\sum_{k=1}^{\alpha n-1} \mathbb{P}(B_{n,k})$.

Let $f_{np_{(k)}}(x)$ denote the density of $np_{(k)}$, $f_{n,k}(y)$ be $f_{np_{(k)}}(nC(k/n) - y)$ and $p_{n,k}(y)$ be $\mathbb{P}\{np_{(k+j)} > nC[(k+j)/n]\ \forall j = 1, 2, \dots, \beta n - k | np_{(k)} = nC(k/n) - y\}$.

CLAIM A.1. *If* $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{n+1}\}$ *is a sequence of i.i.d. exponentially distributed random variables with mean value of* 1, *and* $\Gamma_k = \sum_{i=1}^k \varepsilon_i$, *then* $p_{n,k}(y) = \mathbb{P}(\Gamma_j / \Gamma_{n+1-k} > \frac{y+nC[(k+j)/n]-nC(k/n)}{n+y-nC(k/n)}\ \forall j = 1, \dots, \beta n - k)$.

PROOF. The joint distribution of $(p_{(1)}, \dots, p_{(n)})$ is the same as that of $(\Gamma_1/\Gamma_{n+1}, \dots, \Gamma_n/\Gamma_{n+1})$. Conditional on $\Gamma_k/\Gamma_{n+1} = nC(k/n) - y$, $\{\Gamma_{k+j}/\Gamma_{n+1} > nC[(k+j)/n]\}$ is identical to $\{(\Gamma_{k+j} - \Gamma_k)/(\Gamma_{n+1} - \Gamma_k) > \frac{y+nC[(k+j)/n]-nC(k/n)}{n+y-nC(k/n)}\}$.

To complete the proof, we need to check the independence between Γ_k/Γ_{n+1} and $\{(\Gamma_{k+j} - \Gamma_k)/(\Gamma_{n+1} - \Gamma_k)\}_{j=1}^{n+1-k}$. Basu's theorem indicates $(\Gamma_{n+1} - \Gamma_k) \perp\!\!\!\perp \{(\Gamma_{k+j} - \Gamma_k)/(\Gamma_{n+1} - \Gamma_k)\}_{j=1}^{n-k}$. Besides, Γ_k is independent of $\{\Gamma_{k+j} - \Gamma_k\}_{j=1}^{n+1-k}$. Thus $(\Gamma_k, \Gamma_{n+1} - \Gamma_k) \perp\!\!\!\perp \{(\Gamma_{k+j} - \Gamma_k)/(\Gamma_{n+1} - \Gamma_k)\}_{j=1}^{n-k}$, which implies the desired independence. $\square$

We know $p_{n,k}(y)$ is decreasing in y for every pair of (n, k). The following claim shows $f_{n,k}(y)$ is also decreasing in y when n is large enough.

CLAIM A.2. *For all n large enough and all $k = 2, \dots, \alpha n - 1$, $f_{n,k}(y)$ is decreasing in y when $C(x)$ is C_1 or C_2* [*i.e., to check condition* (ii) *in Remark* A.1].

PROOF. Since $f_{np_{(k)}}$ is increasing on $[0, n(k-1)/(n-1)]$, it suffices to show $C_j(k/n) < (k-1)/(n-1)$ for $j = 1, 2$ when n is large enough. For any fixed x, $f_1(x, y)$ and $f_2(x, y)$ are decreasing in y when $y \le x$. Therefore the inequalities are equivalent to $f_j(k/n, (k-1)/(n-1)) < (b/n^{1/2})^j$ for $j = 1, 2$, which results from the following limit, which converges uniformly in k:

$$f_1\left(\frac{k}{n}, \frac{k-1}{n-1}\right) = \frac{1}{n}\sqrt{\frac{n-k}{k-1}} \to 0 \qquad (n \to \infty),$$

$$f_2\left(\frac{k}{n}, \frac{k-1}{n-1}\right) = \frac{k}{n}\log\frac{k(n-1)}{n(k-1)} - \frac{n-k}{n(n-1)} \to 0 \qquad (n \to \infty). \qquad \square$$

CLAIM A.3. *Assume that $b/n^{1/2} = \xi$ for some $\xi \in \mathbb{R}^+$, C is either C_1 or C_2 and let $\delta = \log n$. Then there exists $M = M(\xi) > 1$ such that*

$$\frac{\mathbb{P}(B_{n,k})}{\int_0^{nC(k/n)\wedge M\delta} f_{n,k}(y)p_{n,k}(y)\,dy} \to 1$$

as $n \to \infty$ uniformly in k.

PROOF. Consider $g(x) = [x - C(x)]/\{C(x)[1 - C(x)]\}$, g is continuous on $(0, \alpha]$; $\lim_{x\to 0^+} g(x) = +\infty$ if $\lim_{x\to 0^+} C(x)/x = 0$, a condition that C_1 and C_2 satisfy. Hence g achieves its minimum, denoted by m, on $(0, \alpha]$. Let $M = 1.1/m + 1.1 > 1$, and $I_{n,k} = \int_0^{nC(k/n)\wedge M\delta} f_{n,k}(y)p_{n,k}(y)\,dy$. The claim reduces to

$$\begin{aligned}\int_{nC(k/n)\wedge M\delta}^{nC(k/n)} & f_{n,k}(y)p_{n,k}(y)\,dy \\ &= o(1)\int_0^{nC(k/n)\wedge M\delta} f_{n,k}(y)p_{n,k}(y)\,dy.\end{aligned}$$

Claim A.2 indicates that when $2 \le k \le \alpha n - 1$,

$$\begin{aligned}&\frac{\mathbb{P}(B_{n,k}) - I_{n,k}}{I_{n,k}} \\ &\quad\le \begin{cases} 0, & \text{if } nC(k/n) \le M\delta, \\ \dfrac{[nC(k/n) - M\delta]^+ f_{n,k}(M\delta)p_{n,k}(M\delta)}{\delta f_{n,k}(\delta)p_{n,k}(\delta)}, & \text{otherwise.}\end{cases}\end{aligned}$$

When $n \ge 3$ and $nC(k/n) > M\delta$, we have $\delta = \log n > 1$, and hence

$$\begin{aligned}&\frac{[nC(k/n) - M\delta]^+ f_{n,k}(M\delta)p_{n,k}(M\delta)}{\delta f_{n,k}(\delta)p_{n,k}(\delta)} \\ &\quad= \frac{[nC(k/n) - M\delta]^+[C(k/n) - M\delta/n]^{k-1}[1 - C(k/n) + M\delta/n]^{n-k}}{\delta[C(k/n) - \delta/n]^{k-1}[1 - C(k/n) + \delta/n]^{n-k}}.\end{aligned} \tag{10}$$

Now consider the continuous version of the RHS of (10); that is, let $k = nx$, $x \in (0, \alpha]$, $y = C(x)$. Recall that α is less than $1/2$ in (9), so $y \le x < 1/2$. When x satisfies $C(x) > M\delta/n > \delta/n$, we have

$$\begin{aligned}&\text{RHS of (10)} \\ &\quad= \frac{(ny - M\delta)^+}{\delta}\exp\biggl\{(nx-1)\log\biggl(1 - \frac{(M-1)\delta}{ny - \delta}\biggr) \\ &\qquad\qquad + n(1-x)\log\biggl(1 + \frac{(M-1)\delta}{(1-y)n + \delta}\biggr)\biggr\}\end{aligned}$$

$$\le \frac{(ny - M\delta)^+}{\delta}\exp\biggl\{-(nx-1)\frac{(M-1)\delta}{ny - \delta} + n(1-x)\frac{(M-1)\delta}{(1-y)n + \delta}\biggr\} \tag{11}$$

$$= \frac{(ny - M\delta)^+}{\delta} \exp\left\{-(M-1)\delta \frac{(x-y) + \delta/n - \delta/n^2 - (1-y)/n}{(y-\delta/n)(1-y+\delta/n)}\right\}$$

$$\leq \frac{(ny - M\delta)^+}{\delta} \exp\left\{-(M-1)\delta \frac{(x-y)}{y(1-y)}\right\} \qquad \text{for } n \text{ large enough.}$$

The last inequality holds because $(y-\delta/n)(1-y+\delta/n) \leq y(1-y)$, and $\delta/n - \delta/n^2 - (1-y)/n$ is positive provided $\log n > 1 + \log n/n$. Then the RHS of (11) $\leq ny/(\delta n^{1.1}) \to 0$ $(n \to \infty)$. Hence the claim holds for all $k = 2, 3, \dots, \alpha n - 1$. For $k = 1$ the claim follows from $nC(1/n) \to 0$. $\square$

LEMMA A.1. *If* $\hat{\Gamma}_n = \sum_{i=1}^n (\xi_i - a)$ *where* ξ_i *are independent and exponentially distributed with mean value of* 1, *then* $\limsup[\log \mathbb{P}(\hat{\Gamma}_n/n \in F)]/n = -\inf_{x \in F} \Lambda_a^*(x)$ *for any interval* F *with positive length*, *where* $\Lambda_a^*(x) = a + x - 1 - \log(a + x)$ when $(a + x) > 0$ and $= +\infty$ otherwise.

PROOF. This lemma follows from the continuity of $\Lambda_a^*(x)$ and, for example, Theorem 2.2.3 in Dembo and Zeitouni (2010), page 27. $\square$

In what follows we continue to use the notation introduced above: $\delta = \log n$, as in the condition of Claim A.3, Λ_a^* is as described in Lemma A.1 and $\Gamma_k = \sum_{i=1}^k \varepsilon_i$, where the ε_i are independent exponential random variables with mean value 1.

PROPOSITION A.2. *If the boundary function* C *is* C_1 *or* C_2, *for all* $k = 1, 2, \dots, \alpha n - 1$, *then for any* $\varepsilon > 0$ *we have* $p_{n,k}(y) \leq (1 + R_{n,k}(\varepsilon))\{1 - \frac{(n-k+1)C'(k/n)}{(1+\varepsilon)[n-nC(k/n)+M\delta]}\}\exp\{-\frac{(n-k+1)y}{(1+\varepsilon)[n-nC(k/n)+M\delta]}\}$, *where* $R_{n,k}(\varepsilon) \to 0$ *as* $n \to \infty$ *uniformly in* k *and* $y \in [0, nC(k/n) \wedge M\delta]$. *As a result* $I_{n,k} \leq (1 + RR_{n,k}(\varepsilon)) \times \int_0^{nC(k/n)\wedge M\delta} f_{n,k}(y)\{1 - \frac{(n-k+1)C'(k/n)}{(1+\varepsilon)[n-nC(k/n)+M\delta]}\} \times \exp\{-\frac{(n-k+1)y}{(1+\varepsilon)[n-nC(k/n)+M\delta]}\}\, dy$ *with* $RR_{n,k}(\varepsilon) \to 0$ *uniformly in* k.

PROOF. Since C_1 and C_2 are convex functions in x for every ξ, C_j' is bounded above by 1 and bounded away from 1 when $x \leq \alpha < 1/2$. Hence

$$\begin{aligned}
&p_{n,k}(y) \\
&\quad = \mathbb{P}\left(\frac{\Gamma_j}{\Gamma_{n+1-k}} \geq \frac{y + nC((k+j)/n) - nC(k/n)}{n + y - nC(k/n)}\ \forall j \in [1, \beta n - k]\right) \\
&\quad \leq \mathbb{P}\left(\frac{\Gamma_j}{\Gamma_{n+1-k}} \geq \frac{y + jC'(k/n)}{n + M\delta - nC(k/n)}\ \forall j \in [1, \beta n - k]\right) \\
&\quad \leq \mathbb{P}\left(\Gamma_j \geq \frac{(n+1-k)(y + jC'(k/n))}{(1+\varepsilon)[n + M\delta - nC(k/n)]}\ \forall j \in [1, \beta n - k]\right) \qquad (12)
\end{aligned}$$

$$
\begin{aligned}
&+\mathbb{P}\left(\left|\frac{\Gamma_{n+1-k}}{n+1-k}\right| \in \left[(1+\varepsilon)^{-1}, 1+\varepsilon\right]^c\right) \\
&= \mathbb{P}\left(\Gamma_j \ge \frac{(n+1-k)(y+jC'(k/n))}{(1+\varepsilon)[n+M\delta-nC(k/n)]} \ \forall j \ge 1\right) + \mathrm{Res}_{\mathrm{up}} \\
&\quad + \mathbb{P}\left(\left|\frac{\Gamma_{n+1-k}}{n+1-k}\right| \in \left[(1+\varepsilon)^{-1}, 1+\varepsilon\right]^c\right) \\
&= \left\{1 - \frac{(n-k+1)C'(k/n)}{(1+\varepsilon)[n-nC(k/n)+M\delta]}\right\} \\
&\quad \times \exp\left\{-\frac{(n-k+1)y}{(1+\varepsilon)[n-nC(k/n)+M\delta]}\right\} + \mathrm{Res}_{\mathrm{up}} \\
&\quad + \mathbb{P}\left(\left|\frac{\Gamma_{n+1-k}}{n+1-k}\right| \in \left[(1+\varepsilon)^{-1}, 1+\varepsilon\right]^c\right).
\end{aligned}
$$

The first term on the RHS of (12) is due to 8.13 of Siegmund (1985), page 186.

Lemma A.1 indicates $\mathbb{P}(|\Gamma_{n+1-k}/(n+1-k)| \in [(1+\varepsilon)^{-1}, 1+\varepsilon]^c) \le A\exp\{-(1-\varepsilon)C_{\mathrm{up}}(\varepsilon)(n+1-k)\}$, where $C_{\mathrm{up}}(\varepsilon) = \min\{\Lambda_0^*(1+\varepsilon), \Lambda_0^*((1+\varepsilon)^{-1})\} > 0$. The union bound of $\mathrm{Res}_{\mathrm{up}}$ is

$$
\begin{aligned}
\mathrm{Res}_{\mathrm{up}} &\le \sum_{j=\beta n-k}^{+\infty} \mathbb{P}\left\{\Gamma_j < \frac{(n+1-k)(y+jC'(k/n))}{(1+\varepsilon)[n+M\delta-nC(k/n)]}\right\} \\
&= \sum_{j=\beta n-k}^{+\infty} \mathbb{P}\left\{\Gamma_j/j < \frac{[1-(k-1)/n]}{(1+\varepsilon)[1+M\delta/n-C(k/n)]}\left[y/j + C'(k/n)\right]\right\}.
\end{aligned}
$$

When n is large enough $[1-(k-1)/n]C'(k/n)/\{(1+\varepsilon)[1+M\delta/n-C(k/n)]\} < a^*$ where $a^* = \max_{x\in(0,\alpha]}(1-x)C'(x)/[1-C(x)] < 1$. Since $j \ge \beta n - k \ge (\beta-\alpha)n$ and $y \le M\delta = M\log n$, $[1-(k-1)/n]/\{(1+\varepsilon)[1+M\delta/n-C(k/n)]\}y/j \le \varepsilon' < 1-a^*$ for all n large enough. Let $C^*_{\mathrm{up}}(\varepsilon')$ denote $\Lambda^*_{a^*}(\varepsilon')$, and hence Lemma A.1 provides the upper bound of the summand,

$$
\begin{aligned}
\mathrm{Res}_{\mathrm{up}} &\le A \sum_{j=\beta n-k}^{+\infty} \exp\{-j(1-\varepsilon)C^*_{\mathrm{up}}(\varepsilon')\} \\
&= A\frac{\exp\{-(\beta-\alpha)n(1-\varepsilon)C^*_{\mathrm{up}}(\varepsilon')\}}{1-\exp\{-(1-\varepsilon)C^*_{\mathrm{up}}(\varepsilon')\}}.
\end{aligned}
$$

The second and third terms on the RHS of (12) decay uniformly faster than the first term, which tends to 0 more slowly than $O(1/n)$. □

PROPOSITION A.3. *With the same assumption of Proposition A.2, we have* $p_{n,k}(y) \ge (1+L_{n,k}(\varepsilon))[1 - \frac{(1+\varepsilon)(n-k+1)C'(k/n)}{n-nC(k/n)}]\exp\{-\frac{(1+\varepsilon)(n-k+1)y}{n-nC(k/n)}\}$, *where*

$L_{n,k}(\varepsilon) \to 0$ as $n \to \infty$ uniformly in k and $y \in [0, C_n(k/n) \wedge M\delta]$. As a result, $I_{n,k} \geq (1 + LL_{n,k}(\varepsilon)) \int_0^{nC(k/n)\wedge M\delta} f_{n,k}(y)\{1 - \frac{(1+\varepsilon)(n-k+1)C'(k/n)}{n-nC(k/n)}\} \times \exp\{-\frac{(1+\varepsilon)(n-k+1)y}{n-nC(k/n)}\}\, dy$ with $LL_{n,k}(\varepsilon) \to 0$ uniformly in k.

PROOF. Due to the convexity of C, we have

$$
\begin{aligned}
&p_{n,k}(y)\\
&\quad = \mathbb{P}\biggl(\frac{\Gamma_j}{\Gamma_{n+1-k}} \geq \frac{y + nC((k+j)/n) - nC(k/n)}{n + y - nC(k/n)}\ \forall j \in [1, \beta n - k]\biggr)\\
&\quad \geq \mathbb{P}\biggl(\Gamma_j \geq \frac{(1+\varepsilon)(n+1-k)}{n - nC(k/n)}\bigl[y + nC\bigl((k+j)/n\bigr) - nC(k/n)\bigr]\\
&\qquad\qquad\qquad\qquad\qquad\qquad \forall j \in [1, \beta n - k]\biggr)\\
&\qquad - \mathbb{P}\biggl(\biggl|\frac{\Gamma_{n+1-k}}{n+1-k}\biggr| \in \bigl[(1+\varepsilon)^{-1}, 1+\varepsilon\bigr]^c\biggr)\\
&\quad \geq \mathbb{P}\biggl(\Gamma_j \geq \frac{(1+\varepsilon)(n+1-k)}{n - nC(k/n)}\bigl[y + jC'\bigl(\bigl(k + \lceil n^{1/2}\rceil\bigr)/n\bigr)\bigr]\\
&\qquad\qquad\qquad\qquad\qquad\qquad \forall j \in \bigl[1, \lceil n^{1/2}\rceil\bigr]\biggr)\\
&\qquad - \mathrm{Res}_{\mathrm{down}} - \mathbb{P}\biggl(\biggl|\frac{\Gamma_{n+1-k}}{n+1-k}\biggr| \in \bigl[(1+\varepsilon)^{-1}, 1+\varepsilon\bigr]^c\biggr)\\
&\quad \geq \mathbb{P}\biggl(\Gamma_j \geq \frac{(1+\varepsilon)(n+1-k)}{n - nC(k/n)}\bigl[y + jC'\bigl(\bigl(k + \lceil n^{1/2}\rceil\bigr)/n\bigr)\bigr]\ \forall j \geq 1\biggr)\\
&\qquad - \mathrm{Res}_{\mathrm{down}} - \mathbb{P}\biggl(\biggl|\frac{\Gamma_{n+1-k}}{n+1-k}\biggr| \in \bigl[(1+\varepsilon)^{-1}, 1+\varepsilon\bigr]^c\biggr).
\end{aligned}
\tag{13}
$$

By using the same argument as in Proposition A.2 and the uniform continuity of C',

$$
\begin{aligned}
&\text{The first term on the RHS of (13)}\\
&\quad = \biggl[1 - \frac{(1+\varepsilon)(n+1-k)}{n - nC(k/n)} C'\bigl(\bigl(k + \lceil n^{1/2}\rceil\bigr)/n\bigr)\biggr]\\
&\qquad \times \exp\biggl\{-\frac{(1+\varepsilon)(n+1-k)}{n - nC(k/n)} y\biggr\}\\
&\quad = \bigl(1 + o(1)\bigr)\biggl[1 - \frac{(1+\varepsilon)(n+1-k)}{n - nC(k/n)} C'(k/n)\biggr]\\
&\qquad \times \exp\biggl\{-\frac{(1+\varepsilon)(n+1-k)}{n - nC(k/n)} y\biggr\}.
\end{aligned}
$$

Let $C'_{\max} = [C(\beta) - C(\alpha)]/(\beta - \alpha) < 1$ [see (9)], we obtain $C((k+j)/n) - C(k/n) \le jC'_{\max}/n$ for all $k < \alpha n$ and $k + j < \beta n$. Hence the union bound of $\mathrm{Res}_{\mathrm{down}}$ becomes

$$
\begin{aligned}
&\mathrm{Res}_{\mathrm{down}} \\
&\quad \le \sum_{j=\lceil n^{1/2}\rceil}^{\beta n-k} \mathbb{P}\biggl(\Gamma_j \le \frac{(1+\varepsilon)(n+1-k)}{n-nC(k/n)}\bigl[y + nC\bigl((k+j)/n\bigr) - nC(k/n)\bigr]\biggr) \\
&\quad \le \sum_{j=\lceil n^{1/2}\rceil}^{\infty} \mathbb{P}\biggl(\Gamma_j \le \frac{(1+\varepsilon)(n+1-k)}{n-nC(k/n)}\bigl(y + jC'_{\max}\bigr)\biggr) \\
&\quad \le A \sum_{j=\lceil n^{1/2}\rceil}^{\infty} \exp\{-j(1-\varepsilon)C^*_{\mathrm{down}}(\varepsilon)\} \\
&\quad \le A \frac{\exp\{-\lceil n^{1/2}\rceil(1-\varepsilon)C^*_{\mathrm{down}}(\varepsilon)\}}{1-\exp\{-(1-\varepsilon)C^*_{\mathrm{down}}(\varepsilon)\}},
\end{aligned}
$$

where $C^*_{\mathrm{down}}(\varepsilon) = \Lambda^*_{(1+\varepsilon)a^{**}}(\varepsilon + \varepsilon^2) > 0$ with $a^{**} = \max_{x\in(0,\alpha]}(1-x)C'_{\max}/[1 - C(x)]$ (so $a^{**} < 1$). By using an argument similar to that in Proposition A.2, it can be concluded that $p_{n,k}(y) \ge (1 + L_{n,k}(\varepsilon))[1 - \frac{(1+\varepsilon)(n-k+1)C'(k/n)}{n-nC(k/n)}] \times \exp\{-\frac{(1+\varepsilon)(n-k+1)y}{n-nC(k/n)}\}$ with $L_{n,k}(\varepsilon) \to 0$ uniformly in k. $\square$

CLAIM A.4. *Suppose that $f(y)$ is a nonnegative, nonincreasing function defined on $[0, +\infty)$ with $f(a) > 0$ for some $a > 0$. Then for any fixed $B > 0$ and any $\beta_1, \beta_2 \ge B$, there exists a continuous and increasing function, denoted by $\rho_B(x)$, defined on $\mathbb{R}^+ \cup \{0\}$ with $\rho_B(0) = 0$, such that $|\ln \int_{\mathbb{R}^+} f(y)e^{-\beta_1 y}\,dy - \ln \int_{\mathbb{R}^+} f(y)e^{-\beta_2 y}\,dy| \le \rho_B(|\beta_1 - \beta_2|)$ and that ρ_B does not depend on f.*

PROOF. Since

$$
\begin{aligned}
\int_{M_1}^{+\infty} f(y)e^{-\beta_1 y}\,dy &\le f(M_1)e^{-\beta_1 M_1}/\beta_1 \\
&= e^{-\beta_1 M_1}/\bigl(1 - e^{-\beta_1 M_1}\bigr)\int_0^{M_1} f(M_1)e^{-\beta_1 y}\,dy \\
&\le e^{-BM_1}/\bigl(1 - e^{-BM_1}\bigr)\int_0^{M_1} f(y)e^{-\beta_1 y}\,dy,
\end{aligned}
$$

we have

$$
\begin{aligned}
\frac{\int_{\mathbb{R}^+} f(y)e^{-\beta_1 y}\,dy}{\int_{\mathbb{R}^+} f(y)e^{-\beta_2 y}\,dy} &\le \frac{1}{1-e^{-BM_1}} \frac{\int_0^{M_1} f(y)e^{-\beta_1 y}\,dy}{\int_0^{M_1} f(y)e^{-\beta_2 y}\,dy} \le \frac{e^{M_1|\beta_1-\beta_2|}}{1-e^{-BM_1}}, \\
\bigl(M_1 = |\beta_1 - \beta_2|^{-1/2}\bigr) &\le \frac{e^{|\beta_1-\beta_2|^{0.5}}}{1-e^{-B|\beta_1-\beta_2|^{-0.5}}}.
\end{aligned}
$$

Hence $\rho_B(x) = |x|^{0.5} - \log(1 - e^{-B|x|^{-0.5}})$ for $x \neq 0$, and = 0 otherwise is the desired function. $\square$

Let $P_{n,k}$ denote the integral $\int_0^{nC(k/n)\wedge M\delta} \{1 - (n - k + 1)C'(k/n)/[n - nC(k/n)]\} f_{n,k}(y) \exp\{-\frac{n+1-k}{n-nC(k/n)} y\}\, dy$.

PROPOSITION A.4. *As $n \to \infty$, $I_{n,k}/P_{n,k} \to 1$ uniformly in k, so consequently $\sum_{k=1}^{\alpha n-1} \mathbb{P}(B_{n,k}) \sim \sum_{k=1}^{\alpha n-1} I_{n,k} \sim \sum_{k=1}^{\alpha n-1} P_{n,k}$.*

PROOF. It is clear $1 - \alpha \le (n + 1 - k)/[n - nC(k/n)] \le 1$. Since $C'(x)$ is bounded away from 1 when $0 < x < \alpha$, $1 - \frac{(n-k+1)C'(k/n)}{(1+\varepsilon)[n-nC(k/n)+M\delta]} = (1 + O(\varepsilon) + O(M\delta/n))\{1 - \frac{(n-k+1)}{n-nC(k/n)} C'(k/n)\}$. For the same reason $1 - \frac{(1+\varepsilon)(n-k+1)}{n-nC(k/n)} \times C'(k/n) = (1 + O(\varepsilon))[1 - \frac{(n-k+1)}{n-nC(k/n)} C'(k/n)]$. So Claim A.4 indicates

$$
\begin{aligned}
&\int_0^{nC(k/n)\wedge M\delta} f_{n,k}(y) \exp\biggl\{-\frac{(n-k+1)y}{(1+\varepsilon)[n - nC(k/n) + M\delta]}\biggr\}\, dy \\
&\qquad \le e^{\rho(O(\varepsilon)+O(M\delta/n))} \int_0^{nC(k/n)\wedge M\delta} f_{n,k}(y) \exp\biggl\{-\frac{(n-k+1)y}{n - nC(k/n)}\biggr\}\, dy \\
&\qquad\quad \times \int_0^{nC(k/n)\wedge M\delta} f_{n,k}(y) \exp\biggl\{-\frac{(1+\varepsilon)(n-k+1)y}{n - nC(k/n)}\biggr\}\, dy \\
&\qquad \ge e^{\rho(O(\varepsilon))} \int_0^{nC(k/n)\wedge M\delta} f_{n,k}(y) \exp\biggl\{-\frac{(n-k+1)y}{n - nC(k/n)}\biggr\}\, dy.
\end{aligned}
$$

According to Propositions A.2, A.3, the previous two inequalities suggest $I_{n,k}/P_{n,k} \to 1$ uniformly in k if we send ε to 0 arbitrarily slowly. $\square$

PROPOSITION A.5. *The following convergence is uniform in k: $[1 - \frac{n+1-k}{n-nC(k/n)} C'(k/n)] \mathbb{P}_{\mathrm{Bin}}(n, k, C(k/n))/P_{n,k} \to 1$. As a result, $\sum_{k=k_0}^{\alpha n-1} P_{n,k} \sim \sum_{k=k_0}^{\alpha n-1} [1 - \frac{n+1-k}{n-nC(k/n)} C'(k/n)] \mathbb{P}_{\mathrm{Bin}}(n, k, C(k/n))$.*

PROOF. (i) When $y \le M\delta$ and $k \le \alpha n$, $\frac{y}{n-nC(k/n)} \le M\delta/[n(1-\alpha)] \to 0$, so consequentially

$$
\begin{aligned}
&\biggl[1 - C(k/n) + \frac{y}{n}\biggr]^{n-k} \exp\biggl\{-\frac{(n+1-k)y}{n - nC(k/n)}\biggr\} \\
&\qquad = [1 - C(k/n)]^{n-k} \exp\biggl\{(n-k)\log\biggl(1 + \frac{y}{n - nC(k/n)}\biggr) - \frac{(n+1-k)y}{n - nC(k/n)}\biggr\} \\
&\qquad = [1 - C(k/n)]^{n-k} \exp\biggl\{-\frac{y}{n - nC(k/n)} - O\biggl(\frac{(n-k)y^2}{n^2[1 - C(k/n)]^2}\biggr)\biggr\}.
\end{aligned}
$$

The remainder on the RHS of the last equation tends to 0 uniformly in k and $0 \le y \le M\delta$. So

$$
\begin{aligned}
P_{n,k} &\sim \left[1 - \frac{(n+1-k)C'(k/n)}{n - nC(k/n)}\right] k \binom{n}{k} [1 - C(k/n)]^{n-k} \\
&\quad \times \int_0^{nC(k/n)\wedge M\delta} \left[C(k/n) - \frac{y}{n}\right]^{k-1} \frac{1}{n}\, dy.
\end{aligned} \tag{14}
$$

(ii) The integral that appears on the RHS of (14) equals

$$
\frac{1}{k}\left\{C\left(\frac{k}{n}\right)^k - \left[C\left(\frac{k}{n}\right) - \frac{nC(k/n) \wedge M\delta}{n}\right]^k\right\}, \tag{15}
$$

$$
\begin{aligned}
(15)\Big/\left[\frac{C(k/n)^k}{k}\right] &= 1 - \left[1 - \frac{nC(k/n) \wedge M\delta}{nC(k/n)}\right]^k \\
&\ge \begin{cases} 1, & \text{if } nC(k/n) < M\delta, \\ 1 - \exp\left\{-\dfrac{kM\delta}{nC(k/n)}\right\}, & \text{otherwise.} \end{cases}
\end{aligned}
$$

Since $k/[nC(k/n)] \ge 1$, $(15) = (1 + o(1))C(k/n)^k/k$ where the infinitesimal tends to 0 uniformly in k. $\square$

Propositions A.1, A.5 together lead to the main result of this part.

THEOREM A.1 (Approximate formula for p-values of higher criticism and Berk–Jones statistics). *If the curve C is C_i $(i = 1, 2)$ and $\beta \in (0, 1)$, then under the overall null hypothesis*

$$
\begin{aligned}
&\mathbb{P}\left(\bigcup_{k=k_0}^{\beta n} \{p_{(k)} \le C(k/n)\}\right) \\
&\quad = (1 + o(1)) \sum_{k=k_0}^{\beta n} \left\{1 - \frac{(n-k+1)C'(k/n)}{n[1 - C(k/n)]}\right\} \mathbb{P}_{\mathrm{Bin}}(n, k, C(k/n)).
\end{aligned}
$$

REMARK. For the modified higher criticism statistic, the approximation takes a slightly different form. The binomial probability is replaced by $\mathbb{P}_{\mathrm{Bin}}\{n, k, \max[1/n, C(k/n)]\} - \mathbb{P}_{\mathrm{Bin}}(n, k, 1/n) \times \max(nC(k/n), 1)$.

Acknowledgments. The authors would like to thank two referees and the Associate Editor for their careful reading and helpful suggestions to improve the exposition of the paper.

REFERENCES

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **57** 289–300. MR1325392

BERK, R. H. and JONES, D. H. (1979). Goodness-of-fit test statistics that dominate the Kolmogorov statistics. *Z. Wahrsch. Verw. Gebiete* **47** 47–59. MR0521531

DARLING, D. A. and ERDÖS, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J.* **23** 143–155. MR0074712

DEMBO, A. and ZEITOUNI, O. (2010). *Large Deviations Techniques and Applications*. *Stochastic Modelling and Applied Probability* **38**. Springer, Berlin. MR2571413

DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195

EIGER, A. M., NADLER, B. and SPIEGELMAN, C. (2013). The calibrated Kolmogorov–Smirnov test. Available at arXiv:1311.3190v1.

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

JAGER, L. and WELLNER, J. A. (2007). Goodness-of-fit tests via phi-divergences. *Ann. Statist.* **35** 2018–2053. MR2363962

JENG, X. J., CAI, T. T. and LI, H. (2013). Simultaneous discovery of rare and common segment variants. *Biometrika* **100** 157–172. MR3034330

LOADER, C. R. (1992). Boundary crossing probabilities for locally Poisson processes. *Ann. Appl. Probab.* **2** 199–228. MR1143400

MEINSHAUSEN, N. and RICE, J. (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.* **34** 373–393. MR2275246

NOÉ, M. (1972). The calculation of distributions of two-sided Kolmogorov–Smirnov type statistics. *Ann. Math. Statist.* **43** 58–64. MR0300379

OWEN, A. B. (1995). Nonparametric likelihood confidence bands for a distribution function. *J. Amer. Statist. Assoc.* **90** 516–521. MR1340504

SIEGMUND, D. (1985). *Sequential Analysis*: *Tests and Confidence Intervals*. Springer, New York. MR0799155

SIEGMUND, D., YAKIR, B. and ZHANG, N. R. (2011). Detecting simultaneous variant intervals in aligned sequences. *Ann. Appl. Stat.* **5** 645–668. MR2840169

WALTHER, G. (2013). 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*. *Inst. Math. Stat.* (*IMS*) *Collect.* **9** 317–326. IMS, Beachwood, OH. MR3202643

WELLNER, J. A. and KOLTCHINSKII, V. (2003). A note on the asymptotic distribution of Berk–Jones type statistics under the null hypothesis. In *High Dimensional Probability*, *III* (*Sandjberg*, 2002). *Progress in Probability* **55** 321–332. Birkhäuser, Basel. MR2033896

WOODROOFE, M. (1976). Frequentist properties of Bayesian sequential tests. *Biometrika* **63** 101–110. MR0415920

XIE, Y. and SIEGMUND, D. (2013). Sequential multi-sensor change-point detection. *Ann. Statist.* **41** 670–692. MR3099117

YAKIR, B. (2013). *Extremes in Random Fields*: *A Theory and Its Applications*. Wiley, Chichester. MR3241226

ZHANG, N. R., SIEGMUND, D. O., JI, H. and LI, J. Z. (2010). Detecting simultaneous changepoints in multiple sequences. *Biometrika* **97** 631–645. MR2672488

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065
USA
E-MAIL: jlijli@stanford.edu
siegmund@stanford.edu