Scandinavian Journal of Statistics Theory and Applications

Scandinavian Journal of Statistics, Vol. 40: 752-769, 2013
doi: 10.1111/sjos.12027

Optimal detection of a jump in the intensity of a Poisson process or in a density with likelihood ratio statistics

CAMILO RIVERA and GUENTHER WALTHER

Stanford University

ABSTRACT. We consider the problem of detecting a 'bump' in the intensity of a Poisson process or in a density. We analyze two types of likelihood ratio-based statistics, which allow for exact finite sample inference and asymptotically optimal detection: The maximum of the penalized square root of log likelihood ratios ('penalized scan') evaluated over a certain sparse set of intervals and a certain average of log likelihood ratios ('condensed average likelihood ratio'). We show that penalizing the *square root* of the log likelihood ratio — rather than the log likelihood ratio itself — leads to a simple penalty term that yields optimal power. The thus derived penalty may prove useful for other problems that involve a Brownian bridge in the limit. The second key tool is an approximating set of intervals that is rich enough to allow for optimal detection, but which is also sparse enough to allow justifying the validity of the penalization scheme simply via the union bound. This results in a considerable simplification in the theoretical treatment compared with the usual approach for this type of penalization technique, which requires establishing an exponential inequality for the variation of the test statistic. Another advantage of using the sparse approximating set is that it allows fast computation in nearly linear time. We present a simulation study that illustrates the superior performance of the penalized scan and of the condensed average likelihood ratio compared with the standard scan statistic.

Key words: average likelihood ratio, fast computation, penalized log likelihood ratio, scan statistic

1. Introduction and overview of results

The paper is concerned with the following problem: One observes an inhomogeneous Poisson process $X_1, \dots, X_N$ on the real line with intensity

$$\lambda(x) = \begin{cases} p\mu(x), & x \in I \\ q\mu(x), & x \notin I \end{cases}$$

where $\mu(x) \geq 0$ is a known function with $\int \mu < \infty$, but $p, q > 0$ and the interval I are unknown. Hence, the intensity is known up to a multiplicative factor, and we want to test whether this factor is elevated on some interval I:

$$H_0: \ p = q, \quad H_A: p > q \ \text{ for some interval } I.$$

This setting arises in a number of applications involving the detection of a 'cluster', for example Glaz & Balakrishnan (1999), Loader (1991) and Kulldorff (1997). The latter two references also give extensions to the bivariate case, which is relevant for detecting spatial disease clusters while adjusting for the known population density μ. Because under H_0 the nuisance parameter $p = q$ is unknown, we follow Loader (1991) and analyze the problem conditional on $N = n$. Then $X_1, \dots, X_n$ are independent and identically distributed (i.i.d.) with density

$$f_{r,I}(x) = \frac{r1(x \in I) + 1(x \in I^c)}{rF_0(I) + F_0(I^c)} f_0(x), \quad \text{where } f_0(x) := \frac{\mu(x)}{\int \mu} \text{ and } r := \frac{p}{q}, \tag{1}$$

and the testing problem becomes $H_0 : r = 1$ versus $H_A : r > 1$, so we test whether the observations come from a known density f_0 (which we may assume without loss of generality to be the uniform density, (5)) versus the case where f_0 is elevated by a multiplicative factor over some interval I. Thus the methodology introduced in this paper may also be applied for certain 'bump-hunting' problems, for example Good & Gaskins (1980), Hartigan & Hartigan (1985), Müller & Sawitzki (1991), Minnotte & Scott (1993) or Polonik (1995).

Loader (1991) and Kulldorff (1997) address the previous problem with the scan statistic, that is the maximum of the log likelihood ratio statistic for varying I. Chan & Walther (2013) investigate a related problem in the abstract Gaussian white noise model. They show that the scan is generally suboptimal for this type of detection problem, but that optimal detection is possible by averaging likelihood ratios over a judiciously chosen collection of intervals. They also suggest that optimality can be restored for the scan either by modifying it with a penalty term that was introduced by Dümbgen & Spokoiny (2001) for kernel statistics in a different context or by using the blocked scan introduced by Walther (2010) and Rufibach & Walther (2010).

Here, we show how optimal detection can be achieved in the practically important case of intensities and densities with likelihood ratios as the principal tool for inference. The main problem in trying to adapt the penalization technique from the abstract Gaussian white noise model is that the form of the penalty term depends partly on the specifics of an exponential inequality that needs to be established for the variation of the local test statistic. This inequality has to be established anew in each setting, and this is a quite difficult theoretical exercise, Section 6.2.. Walther (2010) and Rufibach & Walther (2010) circumvent this problem by penalizing p-values rather than critical values, but at the cost of a more complex methodology and more complex computation.

One of the main contributions of this paper is to show how the conceptually simpler penalization of critical values can be implemented in the important case of log likelihood ratios, without having to establish an exponential inequality for its variation. Our main tool is to consider an appropriate subcollection of the collection of all intervals. It is possible to construct such an approximating set of intervals that on the one hand is rich enough to allow optimal detection and on the other hand is sparse enough to allow justifying the validity of the penalization scheme simply with the union bound. This approach was used in Walther (2010) in the Multivariate Bernoulli Model to penalize p-values when scanning with rectangles. Our key idea to make this approach work for penalizing critical values is to penalize the square root of twice the log likelihood ratio instead of the log likelihood ratio. This transformation results in a penalty that yields optimal detection. And because of the use of a sparse approximating set of intervals, the appropriate penalty term can be read off from the tail bound of the log likelihood ratio itself, which in this case is simply given by Hoeffding's inequality. As will become clear from the exposition, this methodology should also be applicable in a wide range of other contexts, such as those cited in this section.

We end up with a new penalty that is somewhat different from the one used in Dümbgen & Spokoiny (2001). The form of this new penalty derives from a different limiting process (Brownian bridge instead of Brownian motion), and simulations show that it results in a superior finite sample performance when compared with the Dümbgen-Spokoiny penalty.

In the second part of the paper, we show that averaging the likelihood ratios over a particular approximating set of intervals [the *condensed average likelihood ratio (ALR)*] also results in optimal detection. We note that the construction of an appropriate approximating set of intervals plays a crucial role for both methodologies, both in terms of statistical inference and for efficient computation: For the condensed ALR, the appropriate construction of an approximating set directly results in optimal detection, whereas for the penalized scan it justifies the use of the particular penalty term. In both cases, it results in efficient algorithms that run in almost linear time versus the quadratic algorithms required for evaluating all intervals. This computational aspect may well be the dominant concern for some users.

In Section 5, we provide a simulation study that shows that the penalized scan and the condensed ALR are clearly superior to the scan, with the condensed ALR having the overall best performance.

2. The scan statistic and the penalized scan

We will work in the density setting (1), that is, conditional on $N = n$. The main advantage of such a conditional analysis is that it eliminates the nuisance parameter p under the null hypothesis, and hence, this approach avoids the problematic performance of likelihood ratio tests when a parameter is misspecified. Another advantage of the conditional analysis is that it allows for exact finite sample inference as will be seen in the succeeding text. Finally, we note that the conditional analysis does not require the underlying point process to be a Poisson process, but it is also valid for certain other processes that are not Poisson processes or that do not even have independent increments.

A standard computation shows that for a given interval I the log likelihood ratio test statistic for testing $H_0 : r = 1$ versus $H_A : r > 1$ in (1) is given by

$$logLR_n(F_0(I), F_n(I)) := \begin{cases} nF_n(I)\log\left(\frac{F_n(I)}{F_0(I)}\right) + n(1-F_n(I))\log\left(\frac{1-F_n(I)}{1-F_0(I)}\right) & \text{if } F_n(I) > F_0(I) \\ 0 \quad \text{else,} \end{cases}$$

where F_n denotes the empirical cumulative distribution function. Because I is unknown, it is customary to assess the evidence against H_0 with the *scan statistic* (maximum likelihood ratio statistic)

$$M_n := \sup_{\text{intervals } I \subset \boldsymbol{R}} logLR_n(F_0(I), F_n(I)) = \max_{1 \le j < k \le n} logLR_n\left(F_0([X_{(j)}, X_{(k)}]), \frac{k-j+1}{n}\right) \tag{2}$$

where the equality follows from elementary considerations. Kulldorff (1997) gives a derivation of the maximum likelihood ratio without conditioning on N that results in the same formula for M_n. As observed experimentally by Neill (2009a) and Chan (2009), and explained theoretically by Chan & Walther (2013) in an abstract Gaussian regression setting, the scan will generally be suboptimal for detection. One way to rectify the situation is by adding a penalty term as introduced by Dümbgen & Spokoiny (2001) for kernel estimates. We propose to use the following form for a *penalized scan*:

$$P_n := \max_{[X_{(j)}, X_{(k)}] \in \mathcal{J}_{app}} \left(\sqrt{2 logLR_n\left(F_0([X_{(j)}, X_{(k)}]), \frac{k-j+1}{n}\right)} - \sqrt{2\log\frac{en^2}{(k-j)(n-k+j)}} \right),$$

where the data-dependent collection of intervals $\mathcal{J}_{app}$ is defined in the succeeding text. For some applications it may be more appropriate to use a collection of intervals that is not data-dependent, for example Neill (2009b). Therefore, we also analyze the variant

$$P_n^0 := \max_{I \in \mathcal{J}_{app}^0} \left(\sqrt{2logLR_n(F_0(I), F_n(I))} - \sqrt{2 \log \frac{e}{F_n(I)\left(1 - \min\left(F_n(I), \frac{1}{2}\right)\right)}} \right),$$

where $\mathcal{J}_{app}^0$ is defined in the succeeding text. Note that the structure of the penalty in P_n^0 is essentially the same as that in P_n, but a different notation is required because the intervals in $\mathcal{J}_{app}^0$ are not determined by the data. The null distributions of both P_n and P_n^0 are distribution free, which allows exact finite sample inference as detailed in Section 5. Penalizing the square root of $logLR_n$ instead of $logLR_n$ is crucial if one wants to use a simple, additive penalty term that yields optimal detection: Calculations show that an analogously derived penalty term for $logLR_n$ will not result in optimal detection, unless one is willing to work with an intricate non-additive penalty. The previous penalty is different from what one would expect from the work in the abstract Gaussian settings in Dümbgen & Spokoiny (2001) and Chan & Walther (2013). That work suggests to penalize the statistic pertaining to the interval I with $\sqrt{2\log e/F_n(I)}$. However, it will be seen in Section 6.2. that the relevant limiting process of $\sqrt{logLR_n}$ does not involve the increments of Brownian motion, but those of the Brownian bridge. Although a theoretical analysis shows that one can still employ the $\sqrt{2\log e/F_n(I)}$ penalty for the latter case (provided that $F_n(I)$ stays bounded away from 1), it also shows that there is some flexibility in designing the penalty. In fact, the theoretical analysis in Section 6.2. as well as simulations show that for a Brownian bridge it is much preferable to use the penalty $\sqrt{2\log \frac{e}{F_n(I)(1-\min(F_n(I),\frac{1}{2}))}}$, and this is essentially the penalty we used for P_n because we always have $F_n(I) \le \frac{1}{2}$ there.

As approximating set $\mathcal{J}_{app}$, we use the univariate version of the approximating set introduced in Walther (2010):

$$\mathcal{J}_{app} = \bigcup_{\ell=2}^{\ell_{max}} \mathcal{J}_{app}(\ell), \quad \text{where } \ell_{max} = \left\lfloor \log_2 \frac{n}{\log n} \right\rfloor \quad \text{and}$$

$$\mathcal{J}_{app}(\ell) = \{[X_{(j)}, X_{(k)}] : j, k \in \{1 + i d_\ell, i = 0, 1, \ldots\} \text{ and } m_\ell < k - j \le 2m_\ell\},$$

$$\text{where } m_\ell = n2^{-\ell},\ d_\ell = \left\lceil \frac{m_\ell}{6\sqrt{\ell}} \right\rceil.$$

$\mathcal{J}_{app}^0$ is defined[1] analogously with the endpoints of the intervals given by the corresponding quantiles of F_0 rather than those of F_n, that is we use $\left[F_0^{-1}\left(\frac{j}{n}\right), F_0^{-1}\left(\frac{k}{n}\right)\right]$ in place of $[X_{(j)}, X_{(k)}]$. A simple counting argument shows that $\#\mathcal{J}_{app}(\ell) \le 36\,\ell\,2^\ell$, hence both $\mathcal{J}_{app}$ and $\mathcal{J}_{app}^0$ have a cardinality that is bounded by $\sum_{\ell=1}^{\ell_{max}} 36\,\ell\,2^\ell = O(n)$. Thus both P_n and P_n^0 can be computed in $O(n \log n)$ steps, where the complexity is dominated by sorting the data. This advantage of efficient computation plays an important role in many applications.

By definition, $\mathcal{J}_{app}(\ell)$ contains intervals whose empirical measure is roughly the same, up to a factor of two. The 'largest' intervals at $\ell = 2$ have empirical measure up to $\frac{1}{2}$, there is no practical interest in considering larger intervals, and this upper bound can be changed as appropriate. The 'smallest' intervals at $\ell = \ell_{max}$ have empirical measure of about $\log n/n$ because in a density setting it is not possible to obtain consistent inference with fewer observations.

[1] $\log_2$ and log denote the logarithm with base 2 and e, respectively.

This particular choice of $\ell = \ell_{max}$ was also found to work well for the finite sample sizes used in the simulation study in Section 5. The key parameter of the approximating set is d_ℓ, which describes how finely the endpoints are spaced as a function of the length of the interval: Small intervals require a fine spacing for a good approximation, whereas for large intervals a coarser spacing is sufficient. The particular formula given by d_ℓ ensures that intervals of all sizes are approximated sufficiently well to guarantee optimal detection, as shown in Theorem 2, whereas at the same time the approximating set is sparse enough that one can control P_n simply with the union bound (this property does not hold, e.g. for the approximating set given in Rufibach & Walther (2010)):

Proposition 1. *Both P_n and P_n^0 are $O_p(1)$ under H_0.*

Before proving Proposition 1, we note that the second key ingredient besides the sparse approximating set is the 'standardization' of $F_n(I)$ in terms of the transformation $\sqrt{2logLR_n(F_0(I), F_n(I))}$ instead of the usual way to standardize a binomial random variable. The latter case results in one tail that is not subgaussian and which is heavier than the other tail, Shorack and Wellner (1986,Ch.11.1), a problematic fact for the multiple testing set-up considered here. In contrast, the 'standardization' via the previous likelihood ratio transformation leads to clean subgaussian tails: For a fixed interval I and $t > 0$

$$\mathbb{P}\left(\sqrt{2logLR_n(F_0(I), F_n(I))} > t\right) \le \exp\left(-\frac{1}{2}t^2\right). \tag{3}$$

Although we could not find a statement of this result in the literature, it is implicit in the proof of the Chernoff-Hoeffding theorem, Hoeffding (1963): That proof establishes $\mathbb{P}(F_n(I) \ge v) \le \exp(-logLR_n(F_0(I), v))$ for $v \in (F_0(I), 1]$, A.6.1 in van der Vaart & Wellner (1996). Because it is easily seen that the function $v \to logLR_n(F_0(I), v)$ is strictly increasing for $v > F_0(I)$, we obtain $\mathbb{P}(logLR_n(F_0(I), F_n(I)) > t) \le \exp(-t)$ and (3) follows. We note that (3) also holds for the two-sided version of the likelihood ratio provided one adds the factor 2 on the right side.

Because $\#\mathcal{J}^0_{app}(\ell) \le 36\,\ell\,2^\ell$ we obtain for $\kappa > 2$:

$$\begin{aligned}
&\mathbb{P}\left(\max_{I \in \mathcal{J}^0_{app}}\left(\sqrt{2logLR_n(F_0(I), F_n(I))} - \sqrt{2\log\frac{e}{F_0(I)\left(1-\min(F_n(I), \frac{1}{2})\right)}}\right) > \kappa\right) \\
&\le \sum_{\ell=2}^{\ell_{max}} \#\mathcal{J}^0_{app}(\ell) \max_{I \in \mathcal{J}^0_{app}(\ell)} \exp\left(-\frac{1}{2}\left(\sqrt{2\log\frac{e}{F_0(I)}} + \kappa\right)^2\right) \\
&\le \sum_{\ell=2}^{\ell_{max}} 36\ell\, \exp(-\kappa\sqrt{\ell} - \kappa^2/2) \quad \text{since } F_0(I) \le 2 \times 2^{-\ell} \\
&< C\exp(-\kappa^2/2)
\end{aligned}$$

for some universal $C > 0$, proving Proposition 1 for P_n^0, but with $F_0(I)$ instead of $F_n(I)$ in the penalty term. Using this result and (6) one readily finds uniform bounds on the ratios $F_n(I)/F_0(I)$, which allow to replace F_0 by F_n in the penalty term.

The proof of $P_n = O_p(1)$ is analogous, the main difference being that the intervals I are now random. Because by construction of all intervals $I \in \mathcal{J}_{app}$ have empirical measure at least $\log n/n$, Lemma 2 in Section 6 shows that the tails of $\sqrt{2logLR_n}$ are close enough to subgaussian that the previous argument goes through, concluding the proof of Proposition 1.

Finally, we will also consider the direct penalization of the scan (2), that is without approximating the set of all intervals:

$$P_n^{all} := \max_{\substack{1 \le j < k \le n \\ \log n \le k-j \le n/2}} \left(\sqrt{2 logLR_n \left(F_0([X_{(j)}, X_{(k)}]), \frac{k-j+1}{n} \right)} - \sqrt{2 \log \frac{en^2}{(k-j)(n-k+j)}} \right)$$

Our main reason for investigating P_n^{all} is that we need the following result for our theoretical analysis of the average likelihood ratio in Section 3:

Theorem 1. *$P_n^{all} = O_p(1)$ under H_0.*

The restriction of $k - j \geq \log n$ is necessary for this result to hold because for very small intervals, the tail of the test statistic is far from subgaussian, causing the null distribution to blow up, Lemma 2. Of course, those small intervals are not required for optimal detection, and $\mathcal{J}_{app}$ does not employ them either.

The proof of Theorem 1 is neither short nor straightforward, using the Hungarian construction. In contrast, the short proof of Proposition 1, given previously, is essentially an application of Boole's inequality together with a simple counting argument. This is one of the two main advantages of using the approximating set $\mathcal{J}_{app}$, the other being the computational complexity of $O(n \log n)$, whereas P_n^{all} requires to loop over $O(n^2)$ intervals.

Note that all versions of the scan introduced in this section are distribution free and thus allow exact finite sample inference. The availability of algorithms with complexity close to $O(n)$ is crucial for performing such a finite sample inference, Section 5 for details.

The procedures in this section require the specification of F_0. If F_0 is unknown, then these procedures can be viewed as goodness of fit tests for some hypothesized F_0, with optimal power properties against alternatives that concentrate more mass on some interval of unknown location and length. It may also be possible to use these procedures to construct confidence intervals for a distribution function, which improve on, for example, Kolmogorov-Smirnov bands.

3. The condensed average likelihood ratio

Chan & Walther (2013) introduce the *condensed average likelihood ratio* in a regression setting and show that it allows optimal detetion of a bump in a regression function. Here, we investigate its performance in a density context. Define

$$A_n^{cond} := \frac{1}{\#\mathcal{I}_{app}} \sum_{I \in \mathcal{I}_{app}} LR_n(F_0(I), F_n(I)),$$

which is the average of the likelihood ratios $LR_n = \exp(logLR_n)$ over the approximating set of intervals

$$\mathcal{I}_{app} = \bigcup_{\ell=2}^{\ell_{max}} \mathcal{I}_{app}(\ell), \quad \text{where } \ell_{max} = \left\lfloor \log_2 \frac{n}{\log n} \right\rfloor \quad \text{and}$$

$$\mathcal{I}_{app}(\ell) = \{(X_{(j)}, X_{(k)}] : j, k \in \{1 + i d_\ell, i = 0, 1, \ldots\} \text{ and } m_\ell < k - j \le 2m_\ell\},$$

$$\text{where } m_\ell = n2^{-\ell},\ d_\ell = \left\lceil \frac{\sqrt{m_\ell}\, \ell^{4/5}}{\log n} \right\rceil.$$

Note that $\mathcal{I}_{app}$ differs from $\mathcal{J}_{app}$ used previously for the scan in the choice of d_ℓ. The different choice of d_ℓ is necessary to guarantee optimal detection, but it still allows computation in almost linear time because it is readily checked that $\#\mathcal{I}_{app} = O(n \log^2 n)$. A second difference is that $\mathcal{I}_{app}$ uses half-open intervals $(X_{(j)}, X_{(k)}]$ rather than closed intervals with a corresponding empirical measure $\frac{k-j}{n}$ instead of $\frac{k-j+1}{n}$. These changes guarantee that A_n^{cond} will stay bounded under H_0:

Proposition 2. *$A_n^{cond} = O_p(1)$ under H_0.*

The density setting investigated here requires a proof that is more involved than the one in the regression setting considered in Chan & Walther (2013). Further, in the density setting there is no need to consider small intervals with empirical measure less than about $\log n/n$, and $\mathcal{I}_{app}$ is defined accordingly.

A_n^{cond} is also distribution free and thus allows exact finite sample inference.

4. Optimality

Next, we investigate whether the penalized scans P_n and P_n^0 and the condensed average likelihood ratio A_n^{cond} allow optimal detection, that is whether they are able to detect alternatives (1) that satisfy

$$\sqrt{n}\frac{F_{r,I}(I) - F_0(I)}{\sqrt{F_{r,I}(I)}} \geq \sqrt{2\log\frac{e}{F_{r,I}(I)}}\,(1+\epsilon_n), \tag{4}$$

with $\epsilon_n \sqrt{2\log\frac{e}{F_{r,I}(I)}} \to \infty$. Note that both r and I may depend on n, but for simplicity we will not include this in our notation. Using arguments as in Dümbgen & Spokoiny (2001) and in Walther (2010), one can show that no procedure can reliably detect alternatives $F_{r,I}$ that satisfy (4) when $(1+\epsilon_n)$ is replaced by $(1-\epsilon_n)$. Thus (4) does indeed describe a condition for optimal detection. We note that while in the regression context the 'scale' of the effect is given by the spatial extent $|I|$, in the density context this role is played by the probability $F_{r,I}(I)$.

Theorem 2. *The penalized scans P_n and P_n^{all} and the condensed average likelihood ratio A_n^{cond} provide optimal detection, that is they have asymptotic power one uniformly in signals satisfying (4). This result also holds for P_n^0 provided $F_0(I) > 2^{-\ell_{max}}$.*

Thus the optimality of P_n^0 comes with a proviso because of the fact that the approximating set $\mathcal{J}_{app}^0$ is built from the null model and not from the observed data: If the interval I supporting the bump is very small, then the approximating set $\mathcal{J}_{app}^0$ is not fine enough to allow optimal detection. Although this can be remedied by increasing ℓ_{max}, such a step will severely affect the computational complexity, and it is not clear a priori what an appropriate choice for ℓ_{max} would be. P_n and A_n^{cond} avoid this problem by using data-dependent approximating sets. One of the consequences of Theorem 2 is that these approximating sets are rich enough for optimal detection and there is no need to look over all intervals as in P_n^{all}. This has obvious computational advantages as discussed previously, and it allows for a much simplified theoretical analysis: Compare the proofs of Proposition 1 and Theorem 1. An interesting distinction between the scan and the average likelihood ratio is the fact that the approximating set will automatically lead to optimal detection for the latter, but not for the former: Evaluating the unpenalized scan M_n on $\mathcal{J}_{app}$ or on the approximating sets given in Neill & Moore (2004) or Arias-Casto *et al.* (2005) will result in optimal detection only on the smallest scales, that is for $F_{r,I}(I) \approx 2\frac{\log n}{n}$. Optimal detection on all scales seems to require the use of scale-dependent critical values, such as via a penalty term as in P_n or via the *blocked scan* introduced in Rufibach & Walther (2010) and Walther (2010).

5. A simulation study

We illustrate the theoretical results given previously with a simulation study that compares the performance of the scan, the penalized scan P_n and the condensed average likelihood ratio A_n^{cond}. In order to arrive at a fair comparison, we evaluate the scan M_n only over intervals that contain between $\log n$ and $n/2$ observations. This increases the power of the scan compared with the original definition (2) and provides the same a priori assumptions about the length of the cluster for all three methods.

Note that because F_0 is known, we may assume that F_0 is the $U[0,1]$ distribution: Applying the transformation $Y = F_0(X)$ transforms the model (1) into

$$f_{r,I}(y) = \frac{r1(y \in I) + 1(y \in I^c)}{r|I| + 1 - |I|} 1(y \in [0,1]), \tag{5}$$

where the interval I is the image of the original interval I under the map F_0. Moreover, all the statistics $M_n, P_n, P_n^0, P_n^{all}$ and A_n^{cond} are seen to be distribution free. Hence, we may simulate the null distributions of these statistics by drawing $X_1, \ldots, X_n$ independent identically distributed U[0,1] (say), thus allowing for exact (up to Monte Carlo simulation error) finite sample inference.

Tables 1 and 2 list the power at the 5% significance level for sample sizes $n = 10^4$ and $n = 10^6$, respectively. Each case considers the range for the effect ratio r where detection starts

Table 1. *Power in percent for detecting clusters (1) for various values of r and two different lengths of I for sample size* $\mathbf{n} = 10^4$

$\lvert I\rvert = 10^{-3}$				$\lvert I\rvert = 0.3$			
r	scan	pen.scan	cond.ALR	r	scan	pen.scan	cond.ALR
1.8	09	07	05	1.01	05	06	08
2.1	15	14	11	1.03	06	10	18
2.4	31	24	22	1.05	09	23	39
2.7	46	48	36	1.07	17	47	70
3.0	67	65	60	1.09	37	79	90
3.3	82	79	74	1.11	66	92	97
3.6	92	92	85	1.13	89	99	100
3.9	97	97	94	1.15	97	100	100
4.2	99	99	98				

Table 2. *Power in percent for detecting clusters(1) for various values of r and two different lengths of I for sample size* $\mathbf{n} = 10^6$

$\lvert I\rvert = 10^{-4}$				$\lvert I\rvert = 0.3$			
r	scan	pen.scan	cond.ALR	r	scan	pen.scan	cond.ALR
1.25	06	06	05	1.002	06	07	10
1.35	07	08	07	1.004	05	14	23
1.45	14	16	15	1.006	05	38	52
1.55	35	40	34	1.008	09	69	80
1.65	61	66	62	1.010	14	91	96
1.75	83	86	85	1.012	39	99	99
1.85	96	97	95	1.014	71	100	100
1.95	99	99	99	1.016	92	100	100
				1.018	99	100	100

to become possible, for a small interval and for a large interval I. These simulations illustrate how the optimality result of Section 4 about P_n and A_n^{cond} sets in. In contrast, one sees that the scan M_n is competitive only for signals on the smallest scales and it is inferior to P_n and A_n^{cond} otherwise. In the context of regression, the inferiority of the scan at larger scales was expounded theoretically by Chan & Walther (2013). Note that unlike in the regression context, 'scale' is not given by the length $|I|$ but by $F_{r,I}(I)$, which is of the order $rF_0(I)$ as long as the latter quantity stays bounded.

The simulations show that the condensed average likelihood ratio A_n^{cond} has arguably the best overall performance among the three procedures considered.

In the previous simulations the finite sample exact critical values and the power were approximated with 10^5 and 10^3 simulations, respectively. The location of the interval I was randomized in each simulation to avoid confounding the results with the particular construction of the approximating sets $\mathcal{I}_{app}$ and $\mathcal{J}_{app}$. In the case of the sample size $n = 10^6$, the scan M_n was evaluated on the approximation set $\mathcal{J}_{app}$, that is the same approximation set used for P_n, because looking at all intervals was computationally prohibitive. Conversely, for sample size $n = 10^4$ we examined the effect of the approximating set by running the simulation with the penalized scan, and the condensed average likelihood ratio evaluated over all intervals containing between $\log n$ and $n/2$ observations rather than evaluating them over an approximating set. We observed only a very small changes in power, mostly decreases, and the computation time was much longer. This confirms the theoretical finding from Section 4 that it suffices the evaluation of these statistics over an approximating set, which yields tremendous computational advantages without sacrificing detection power.

6. Proofs

6.1. Preliminary results

(1.) Using $\log x \leq x - 1$ and a two term Taylor expansion respectively gives for $a, b \in (0, 1)$:

$$
\begin{aligned}
n\frac{(b-a)^2}{a(1-a)} \;\geq\; logLR_n(a,b) \;&=\; \frac{n}{2\xi(1-\xi)}(b-a)^2 1(a<b) \quad \text{for } \xi \in (a,b) \\
&\geq\; \frac{n}{2b}(b-a)^2 1(a<b)
\end{aligned}
\tag{6}
$$

(2.) Let I be an interval satisfying $\ell := \lfloor \log_2 1/F_n(I) \rfloor + 1 \leq \ell_{max}$, so $m_\ell < nF_n(I) \leq 2m_\ell$. Then by construction of $\mathcal{J}_{app}(\ell)$ there exists $\tilde{I} \in \mathcal{J}_{app}(\ell)$ such that

$$
F_n(I \triangle \tilde{I}) \;\leq\; 2\frac{d_\ell - 1}{n} \;\leq\; \frac{F_n(I)}{3\sqrt{\ell}},
\tag{7}
$$

and the same result holds for $\mathcal{J}^0_{app}$ with F_n replaced by F_0 in the previous text.

Lemma 1. *Let J be an interval, and $F_{r,I}$ be the distribution given in* (1) *with $r \geq 1$. Then for $G \in \{F_0, F_{r,I}\}$:*

$$
(F_{r,I} - F_0)(J) \geq (F_{r,I} - F_0)(I)\left(1 - \frac{G(I \triangle J)}{G(I)}\right) \quad \text{if } G(I) \leq \frac{1}{2}, \text{ and}
$$

$$
1 - \frac{F_0(I \setminus J)}{F_0(I)} \leq \frac{F_{r,I}(J)}{F_{r,I}(I)} \;\leq\; 1 + \frac{F_0(J \setminus I)}{F_0(I)}.
$$

For a proof of Lemma 1 note that

$$f_{r,I}(x) - f_0(x) = \begin{cases} d_{r,I}\, f_0(x) & \text{if } x \in I \\ -\frac{F_0(I)}{F_0(I^c)} d_{r,I}\, f_0(x) & \text{if } x \in I^c \end{cases}$$

where $d_{r,I} := r/(rF_0(I) + F_0(I^c)) - 1 \geq 0$ because $r \geq 1$. Hence,

$$\begin{aligned}(F_{r,I} - F_0)(J) &= d_{r,I}\, F_0(I \cap J) - \frac{F_0(I)}{F_0(I^c)} d_{r,I}\, F_0(J \setminus I) \\ &= (F_{r,I} - F_0)(I)\left(\frac{F_0(I \cap J)}{F_0(I)} - \frac{F_0(J \setminus I)}{F_0(I^c)}\right) \\ &\geq (F_{r,I} - F_0)(I)\frac{F_0(I \cap J) - F_0(J \setminus I)}{F_0(I)} \quad \text{if } F_0(I) \leq \frac{1}{2}\end{aligned}$$

and the claim for $G = F_0$ follows by writing $F_0(I \cap J) = F_0(I) - F_0(I \setminus J)$. The claim for $G = F_{r,I}$ follows because $\frac{F_0(I\cap J)}{F_0(I)} - \frac{F_0(J\setminus I)}{F_0(I^c)} = \frac{F_{r,I}(I\cap J)}{F_{r,I}(I)} - \frac{F_{r,I}(J\setminus I)}{F_{r,I}(I^c)}$ by the definition of $f_{r,I}$. The second claim follows from dividing the inequality $F_{r,I}(I) - F_{r,I}(I \setminus J) \leq F_{r,I}(J) \leq F_{r,I}(I) + F_{r,I}((J \setminus I)$ by $F_{r,I}(I)$ and observing $\frac{F_{r,I}(I\setminus J)}{F_{r,I}(I)} = \frac{F_0(I\setminus J)}{F_0(I)}$ and $\frac{F_{r,I}(J\setminus I)}{F_{r,I}(I)} = \frac{1}{r}\frac{F_0(J\setminus I)}{F_0(I)}$ by the definition of $f_{r,I}$.

The following lemma is an extension of Proposition 2.1 in Dümbgen (1998):

Lemma 2. *Denote the two-sided log likelihood ratio statistic by* $logLR_n^{two}(a,b) := nb\log\frac{b}{a} + n(1-b)\log\frac{1-b}{1-a}$*, and the one-sided versions by* $logLR_n^{left}(a,b) := logLR_n^{two}(a,b)1(a < b)$ *and* $logLR_n^{right}(a,b) := logLR_n^{two}(a,b)1(a > b)$*. Hence,* $logLR_n^{left}$ *equals* $logLR_n$ *used previously. Let* $U_1, \ldots, U_n$ *be* i.i.d. $U[0,1]$, *so* $U_{(k)} - U_{(j)}$ ∼beta$(k-j, n+1-k+j)$ *for* $1 \leq j < k \leq n$*. Set* $p_{jk} := \frac{k-j}{n+1}$*. Then for* $p \in (0,1)$ *and* $t > 0$:

$$\begin{aligned}&\mathbb{P}\left(\sqrt{2logLR_n^{two}(U_{(k)} - U_{(j)}, p)} > t\right) \\ &\quad\leq 2\exp\left\{-\min\left(\frac{p_{jk}}{p}, \frac{1-p_{jk}}{1-p}\right)\frac{(n+1)}{n}\frac{t^2}{2} + n\frac{p - p_{jk}}{1(p > p_{jk}) - p_{jk}}\right\} \\ &\quad\leq \begin{cases} 2\exp\left(-\frac{t^2}{2}\right) & \text{if } p := p_{jk} \\ 2\exp\left(-\frac{(k-j)}{(k-j+1)}\frac{t^2}{2} + 3\right) & \text{if } p := \frac{k-j+1}{n} \leq \frac{1}{2}. \end{cases}\end{aligned}$$

In more detail:

$$\mathbb{P}\left(logLR_n^{left}(U_{(k)} - U_{(j)}, p) > t\right) \leq \exp\left\{-\frac{p_{jk}}{p}\frac{(n+1)}{n}\left(t - n\frac{(p - p_{jk})(p - p_{jk}1(p_{jk} > p))}{p_{jk}(1 - p_{jk})}\right)\right\}$$

$$\begin{aligned}\mathbb{P}\left(logLR_n^{right}(U_{(k)} - U_{(j)}, p) > t\right) &\leq \exp\left\{-\frac{(1 - p_{jk})}{(1-p)}\frac{(n+1)}{n}\right. \\ &\quad\left.\times\left(t - n\frac{(p_{jk} - p)[1 - p - (1 - p_{jk})1(p_{jk} < p)]}{p_{jk}(1 - p_{jk})}\right)\right\}\end{aligned}$$

Hence, in the case of random intervals whose lengths follow the beta distribution, $\sqrt{2logLR_n^{two}}$ has subgaussian tails for $p = p_{jk}$. For p close to p_{jk}, the tails are still subgaussian, but with a scale factor that is smaller in one tail and larger in the other.

Proof of Lemma 2. For $u \in (0, p)$:

$$\begin{aligned}
logLR_n^{left}(u,p) &= logLR_n^{two}(u,p) \\
&= \frac{p}{p_{jk}} logLR_n^{two}(u,p_{jk}) + logLR_n^{two}(p_{jk},p) \\
&\quad + n\frac{p_{jk}-p}{p_{jk}} \log\frac{1-p_{jk}}{1-u} \\
&\le \frac{p}{p_{jk}} logLR_n^{two}(u,p_{jk}) + n\frac{(p_{jk}-p)^2}{p_{jk}(1-p_{jk})} \\
&\quad + n\frac{p_{jk}-p}{p_{jk}} \log(1-p_{jk})\, 1(p_{jk}<p) \quad \text{by (6)} \\
&\le \frac{p}{p_{jk}} logLR_n^{left}(u,p_{jk}) + n\frac{(p-p_{jk})(p-p_{jk}1(p_{jk}>p))}{p_{jk}(1-p_{jk})}
\end{aligned}$$

because $-(1-p_{jk})\log(1-p_{jk}) \le p_{jk}$ and $logLR_n^{left}(u,p)$ is non increasing in u, whereas $logLR_n^{two}(u,p_{jk})$ is increasing for $u > p_{jk}$. Hence, the previously mentioned inequality also holds with $logLR_n^{two}(u,p_{jk})$ replaced by $\min\left(logLR_n^{two}(u,p_{jk}), logLR_n^{two}(p_{jk},p_{jk})\right) = logLR_n^{left}(u,p_{jk})$. Now the probability inequality for $logLR_n^{left}$ follows from the previous inequality together with $\mathbb{P}\left(logLR_n^{left}(U_{(k)}-U_{(j)},p_{jk})>t\right) \le \exp\left(-\frac{n+1}{n}t\right)$, which is a consequence of Proposition 2.1 in Dümbgen (1998). The inequality for the right tail follows analogously, and the tail bound for $\sqrt{2logLR_n^{two}}$ obtains as a consequence. □

6.2. *Proof of Theorem 1*

Under H_0 and $(F_0(X_1),\dots,F_0(X_n)) \stackrel{d}{=} (U_1,\dots,U_n)$, where $U_1,\dots,U_n$ are i.i.d. U[0,1]. We divide the collection of intervals under consideration into a collection of small intervals

$$\mathcal{S}_n := \left\{(j,k):\ 1 \le j < k \le n,\ \log n \le k-j \le \log^2 n\right\}$$

and the collection of the remaining intervals

$$\mathcal{I}_n := \{(j,k);\ 1 \le j < k \le n,\ \log^2 n < k-j \le n/2\}.$$

The cardinality of $\mathcal{S}_n$ is small enough that we can use the union bound to show

$$\max_{(j,k)\in\mathcal{S}_n}\left(\sqrt{2logLR_n\left(U_{(k)}-U_{(j)},\frac{k-j+1}{n}\right)} - \sqrt{2\log\frac{en^2}{(k-j)(n-k+j)}}\right)^+ = o_p(1) \tag{8}$$

For the larger intervals we approximate $\sqrt{2logLR_n}$ by the normalized increment of the uniform quantile process:

$$\max_{(j,k)\in\mathcal{I}_n}\left|\sqrt{2logLR_n\left(U_{(k)}-U(j),\frac{k-j+1}{n}\right)}-\sqrt{n}\frac{\left|\frac{k-j}{n}-(U_{(k)}-U(j))\right|}{\sqrt{\frac{k-j}{n}\left(1-\frac{k-j}{n}\right)}}\right|=O_p(1) \tag{9}$$

The normalized increments of the uniform quantile process can in turn be approximated on $\mathcal{I}_n$ by the normalized increments of a Brownian bridge B:

$$\max_{(j,k)\in\mathcal{I}_n}\left|\sqrt{n}\frac{\left|\frac{k-j}{n}-(U_{(k)}-U(j))\right|}{\sqrt{\frac{k-j}{n}\left(1-\frac{k-j}{n}\right)}}-\frac{\left|B\left(\frac{k}{n}\right)-B\left(\frac{j}{n}\right)\right|}{\sqrt{\frac{k-j}{n}\left(1-\frac{k-j}{n}\right)}}\right|=O_p(1) \tag{10}$$

Theorem 1 follows from (8–10) together with

$$\sup_{0<s<t<1}\left(\frac{|B(t)-B(s)|}{\sqrt{(t-s)(1-(t-s))}}-\sqrt{2\log\frac{e}{(t-s)(1-(t-s))}}\right)<\infty \; a.s. \tag{11}$$

The previous results also show how one might design an appropriate penalty if one wishes to scan over a very large intervals, that is $(k-j)/n$ close to one. Indeed, it is well-known that the normalized uniform quantile process blows up both at zero and one, Ch. 16 in Shorack & Wellner (1986).

Proof of (8). Clearly $\#\mathcal{S}_n \le n\log^2 n$. Hence, the tail inequality for $\sqrt{2logLR_n}$ given in Lemma 2 yields for $C>0$:

$$\begin{aligned}
&\mathbb{P}\left(\max_{(j,k)\in\mathcal{S}_n}\left(\sqrt{2logLR_n\left(U_{(k)}-U_{(j)},\frac{k-j+1}{n}\right)}\right.\right.\\
&\qquad\left.\left.-\sqrt{2\log\frac{en^2}{(k-j)(n-k+j)}}\right)>C\right)\\
&\le n(\log^2 n)\max_{(j,k)\in\mathcal{S}_n}2e^3\exp\left\{-\frac{k-j}{2(k-j+1)}\left(C+\sqrt{2\log\frac{en^2}{(k-j)(n-k+j)}}\right)^2\right\}\\
&\le 2e^3 n(\log^2 n)\exp\left\{-\left(1-\frac{1}{\log n}\right)\left(\frac{C^2}{2}+\log\frac{en}{\log^2 n}+C\sqrt{2\log\frac{en}{\log^2 n}}\right)\right\}\\
&\quad\text{since } (k-j)(n-k+j)\le n\log^2 n \text{ on } \mathcal{S}_n\\
&\le 2e^3(\log^4 n)\exp\left\{-\left(1-\frac{1}{\log n}\right)\left(\frac{C^2}{2}+C\sqrt{2\log\frac{en}{\log^2 n}}\right)\right\}\ \to 0
\end{aligned}$$

□

Proof of (9). By (6)

$$\left| \sqrt{2logLR_n\left(U_{(k)}-U_{(j)},\frac{k-j+1}{n}\right)} - \sqrt{n}\frac{\left|\frac{k-j+1}{n}-(U_{(k)}-U_{(j)})\right|}{\sqrt{\frac{k-j}{n}\left(1-\frac{k-j}{n}\right)}}\right|$$
$$= \sqrt{n}\left|\frac{k-j+1}{n}-(U_{(k)}-U_{(j)})\right|\left|\sqrt{\frac{1}{\xi(1-\xi)}}-\sqrt{\frac{1}{\frac{k-j}{n}\left(1-\frac{k-j}{n}\right)}}\right| \quad (12)$$

for ξ between $\frac{k-j+1}{n}$ and $U_{(k)}-U_{(j)}$. On the event

$$\mathcal{A}_n(C) := \left\{\left|U_{(k)}-U_{(j)}-\frac{k-j}{n}\right| \le \left(C+\sqrt{2\log\frac{en^2}{(k-j)(n-k+j)}}\right) \times \sqrt{\frac{k-j}{n^2}\left(1-\frac{k-j}{n}\right)} \text{ for all } (j,k)\in\mathcal{I}_n\right\}$$

we have $\left|U_{(k)}-U_{(j)}-\frac{k-j}{n}\right| \le \frac{(C+\sqrt{2\log n})}{\sqrt{k-j}}\frac{(k-j)}{n} \le \frac{2}{\sqrt{\log n}}\frac{(k-j)}{n}$ eventually. Hence,

$$\left|\frac{1}{\xi(1-\xi)}-\frac{1}{\frac{k-j}{n}\left(1-\frac{k-j}{n}\right)}\right| \le \frac{1}{\xi\frac{k-j}{n}}\left|\xi-\frac{k-j}{n}\right| + \frac{1}{(1-\xi)\left(1-\frac{k-j}{n}\right)}\left|\xi-\frac{k-j}{n}\right|$$
$$\le \frac{4n}{(k-j)\sqrt{\log n}}+\frac{4n}{(n-k+j)\sqrt{\log n}}$$
$$= \frac{4}{\frac{k-j}{n}\left(1-\frac{k-j}{n}\right)\sqrt{\log n}}.$$

Because $0<b-a<b/2$ for $a,b>0$ implies $|\sqrt{b}-\sqrt{a}| \le (b-a)/\sqrt{b}$, (12) is not larger than

$$\frac{4\sqrt{n}\left|\frac{k-j+1}{n}-(U_{(k)}-U_{(j)})\right|}{\sqrt{\frac{k-j}{n}\left(1-\frac{k-j}{n}\right)\log n}} \le \frac{4\left(C+\sqrt{2\log\frac{en^2}{(k-j)(n-k+j)}}\right)}{\sqrt{\log n}}$$
$$\times\frac{4}{\sqrt{n}\sqrt{\frac{k-j}{n}\left(1-\frac{k-j}{n}\right)\log n}} \le 4\frac{C+\sqrt{2\log n}}{\sqrt{\log n}}+\frac{8}{(\log n)^{3/2}} \quad \text{for } (j,k)\in\mathcal{I}_n.$$

Equation (9) follows because $\lim_{C\to\infty}\liminf_{n\to\infty}\mathbb{P}(\mathcal{A}_n(C)) = 1$ by (10) and (11), and replacing $\frac{k-j}{n}$ with $\frac{k-j+1}{n}$ in the numerator of the second term in (9) incurs a difference bounded by $8/\log n$ as seen previously. □

Proof of (10). By the Hungarian construction, Theorem 12.2.2 in Shorack & Wellner (1986), there exists a version B_n of the Brownian bridge on [0, 1] such that

$$\begin{aligned}
&\limsup_n \max_{(j,k)\in\mathcal{I}_n} \left| \sqrt{n} \left| \frac{k-j}{n} - (U_{(k)} - U_{(j)}) \right| - \left| B_n\left(\frac{k}{n}\right) - B_n\left(\frac{k}{n}\right) \right| \right| \\
&\quad \le \limsup_n \max_{(j,k)\in\mathcal{I}_n} \left(\left| \sqrt{n}\left(U_{(k)} - \frac{k}{n}\right) - B_n\left(\frac{k}{n}\right) \right| + \left| \sqrt{n}\left(U_{(j)} - \frac{j}{n}\right) - B_n\left(\frac{j}{n}\right) \right| \right) \\
&\quad \le 2M \frac{\log n}{\sqrt{n}} \quad \text{a.s. form some } M < \infty
\end{aligned}$$

The claim follows because $\sqrt{\frac{k-j}{n}\left(1 - \frac{k-j}{n}\right)} \ge \frac{\log n}{2\sqrt{n}}$ for $(j,k) \in \mathcal{I}_n$. □

Proof of (11). This can be proved using Theorem 6.1 in Dümbgen & Spokoiny (2001): On the set of all intervals $\mathcal{T} := \{(s,t] : 0 < s < t < 1\}$ define the metric ρ via $\rho^2((s,t],(s',t']) := |s-s'| + |t-t'|$, and the stochastic process $X((s,t]) := B(t) - B(s)$. With $\sigma^2((s,t]) := (t-s)(1-t+s)$ one readily checks that $\sigma^2((s,t]) \le \sigma^2((s',t']) + \rho^2((s,t],(s',t'])$. Because $X((s,t])/\sigma((s,t]) \sim$N(0,1), the subgaussian tail condition (i) of said theorem holds. To prove the subgaussian tail condition (ii) for the variation of X, write $B(t) = W(t) - tW(1)$ for a Brownian motion W. Then

$$\begin{aligned}
\frac{X((s,t]) - X((s',t'])}{\rho((s,t],(s',t'])} &= \frac{W((s,t] \setminus (s',t']) - W((s',t'] \setminus (s,t])}{\sqrt{|s-s'| + |t-t'|}} \\
&\quad - W(1)\frac{(t-s) - (t'-s')}{\sqrt{|s-s'| + |t-t'|}} \\
&\stackrel{d}{=} N\left(0, \frac{\text{Leb}((s,t]\triangle(s',t'])}{|s-s'| + |t-t'|} + \frac{((t-s)-(t'-s'))^2}{|s-s'| + |t-t'|} \right. \\
&\quad \left. - 2\,\frac{\text{Leb}((s,t]\triangle(s',t'])\,((t-s)-(t'-s'))}{|s-s'| + |t-t'|}\right)
\end{aligned}$$

The latter variance is not more than four, hence, condition (ii) holds with $L = 1$ and $M = 8$. Finally, a calculation similar as in Dümbgen & Spokoiny (2001) shows that $V = 1$. (11) follows.

Checking condition (iii), that is establishing an exponential inequality for the variation of the process under consideration, is the technically most challenging aspect in applying said Theorem 6.1, for example Dümbgen & Walther (2008). Here, we approached this problem via the Hungarian construction, which leads to the simpler task of establishing an exponential inequality for the variation of the increments of a Brownian bridge. □

6.3. Proof of Proposition 2

We use $F_0\left((X_{(j)}, X_{(k)}]\right) \stackrel{d}{=} U_{(k)} - U_{(j)}$ for $U_1, \ldots, U_n$ i.i.d. U[0,1] and define the event

$$\begin{aligned}
\mathcal{B}_{m,n} := \Bigg\{ &\sqrt{2logLR_n\left(U_{(k)} - U_{(j)}, \frac{k-j}{n}\right)} \\
&\le \sqrt{2\log\frac{en^2}{(k-j)(n-k+j)}} + m \text{ for all } (j,k) \in \boldsymbol{\mathcal{I}}_{app} \Bigg\}.
\end{aligned}$$

We will show that for $(j,k) \in \boldsymbol{\mathcal{I}}_{app}(\ell)$

$$\mathbb{E}\left(LR_n\left(U_{(k)}-U_{(j)},\frac{k-j}{n}\right)1_{\mathcal{B}_{m,n}}\right) \le 14\left(\sqrt{2\ell}+m\right) \quad \text{eventually, uniformly in} \tag{13}$$

(j,k) and ℓ. Then $A_n^{cond} = O_p(1)$ can be shown as in the proof of Theorem 3 in Chan & Walther (2013) because $\lim_{m\to\infty}\liminf_{n\to\infty}\mathbb{P}(\mathcal{B}_{m,n}) = 1$ by Theorem 1, which is readily seen to hold also with $\frac{k-j}{n}$ in place of $\frac{k-j+1}{n}$ in the definition of P_n^{all}.

To prove (13) fix $(j,k) \in \mathcal{I}_{app}(\ell)$. We will show in the succeeding text that on the event $\mathcal{B}_{m,n}$ for $n \ge n_0(m)$

(a) $u := U_{(k)} - U_{(j)}$ falls in an interval B of length at most

$$4\sqrt{\frac{\frac{k-j}{n}\left(1-\frac{k-j}{n}\right)}{n}}\left(C\left(\frac{k-j}{n}\right)+m\right), \quad \text{where } C(\delta) := \sqrt{2\log\frac{1}{\delta}},$$

and

(b) $u \ge \frac{k-j}{8n}$.

Using the fact that $U_{(k)} - U_{(j)} \sim \text{beta}(k-j, n+1-k+j)$, we can then compute

$$\begin{aligned}
&\mathbb{E}\left(LR_n\left(U_{(k)}-U_{(j)},\frac{k-j}{n}\right)1_{\mathcal{B}_{m,n}}\right)\\
&= \int_B \left(\frac{k-j}{nu}\right)^{k-j}\left(\frac{1-\frac{k-j}{n}}{1-u}\right)^{n-k+j} u^{k+j-1}(1-u)^{n-k+j}\\
&\quad\times \frac{n!}{(k-j-1)!(n-k+j)!}\,du\\
&\le \frac{e}{2\pi}\int_B \frac{k-j}{nu}\sqrt{\frac{n}{\frac{k-j}{n}\left(1-\frac{k-j}{n}\right)}}\,du \quad \text{by Stirling's formula}\\
&\le \frac{e}{2\pi}32\left(C\left(\frac{k-j}{n}\right)+m\right)
\end{aligned}$$

by (a) and (b). (13) follows because $(j,k) \in \mathcal{I}_{app}(\ell)$ implies $\frac{k-j}{n} > 2^{-\ell}$.

(a) follows for $n \ge n_0(m)$ from the inequality given in Proposition 2.1 in Dümbgen (1998) together with the fact that $k-j \ge \log n$ by the construction of $\mathcal{I}_{app}$. The said inequality yields in particular

$$u \ge \frac{k-j}{n} - \sqrt{\frac{\frac{k-j}{n}\left(1-\frac{k-j}{n}\right)}{n}}\left(C\left(\frac{k-j}{n}\right)+m\right) \ge \frac{k-j}{n}\left(1-\frac{C\left(\frac{k-j}{n}\right)+m}{\sqrt{k-j}}\right).$$

Thus in the case $k-j \ge 4\log n$, (b) follows because $\sqrt{k-j} \ge \frac{8}{7}\left(C\left(\frac{k-j}{n}\right)+m\right)$ for $n \ge n_0(m)$. In the case of $k-j = b\log n$ with $b \in [1,4)$, consider $u := \frac{k-j}{8n}$. Then a standard calculation shows that $logLR_n\left(u,\frac{k-j}{n}\right) \ge \left(\frac{9}{8}+o(1)\right)\log n$, where the $o(1)$ term is uniform in b. Thus this choice of u violates the inequality defining $\mathcal{B}_{m,n}$ for $n \ge n_0(m)$. Because $logLR_n\left(u,\frac{k-j}{n}\right)$ increases as u moves away from $\frac{k-j}{n}$, this implies that we must have $u > \frac{k-j}{8n}$ for $n \ge n_0(m)$, completing the proof of (b).

6.4. Proof of Theorem 2

We first prove the claim about P_n^0. Consider an alternative (1) that satisfies (4) and also $F_0(I) > 2^{-\ell_{max}}$. Then $\ell := \lfloor \log_2 1/F_0(I) \rfloor + 1 \le \ell_{max}$, so by (7) there exists $\tilde{I} \in \mathcal{J}^0_{app}(\ell)$ with $F_0(I \triangle \tilde{I}) \le \frac{F_0(I)}{3\sqrt{\ell}}$. Set $b_n := \epsilon_n \sqrt{2 \log \frac{e}{F_{r,I}(I)}}$, so $b_n \to \infty$ by assumption (4). On the event $\mathcal{A}_n := \left\{ F_n(\tilde{I}) \ge F_{r,I}(\tilde{I}) - \sqrt{\frac{F_{r,I}(\tilde{I}) b_n}{n}} \right\}$ condition (4) implies $F_n(\tilde{I}) \ge F_0(\tilde{I})$ and hence,

$$
\begin{aligned}
\sqrt{2logLR_n(F_0(\tilde{I}), F_n(\tilde{I}))} &\ge \sqrt{n} \frac{F_n(\tilde{I}) - F_0(\tilde{I})}{\sqrt{F_n(\tilde{I})}} \quad \text{by (6)} \\
&\ge \sqrt{n} \frac{F_{r,I}(\tilde{I}) - F_0(\tilde{I})}{\sqrt{F_{r,I}(\tilde{I})}} - \sqrt{b_n} \quad \text{on } \mathcal{A}_n \text{ since } x \to \frac{x - F_0(\tilde{I})}{x} \nearrow \\
&\ge \sqrt{n} \frac{F_{r,I}(I) - F_0(I)}{\sqrt{F_{r,I}(I)}} \left(1 - \frac{1}{3\sqrt{\ell}}\right)^2 - \sqrt{b_n} \quad \text{by Lemma 1} \\
&\ge \left(\sqrt{2 \log \frac{e}{F_{r,I}(I)}} + b_n \right) \left(1 - \frac{2}{3\sqrt{\log \frac{e}{F_{r,I}(I)}}} \right) - \sqrt{b_n} \\
&\ge \sqrt{2 \log \frac{e}{3F_n(\tilde{I})}} + \frac{1}{3} b_n - 1 - \sqrt{b_n}
\end{aligned}
$$

where the last inequality holds by Lemma 1 and on the event $\mathcal{B}_n := \{F_{r,I}(\tilde{I}) \le 2F_n(\tilde{I})\}$. Chebyshev's inequality gives $\mathbb{P}(\mathcal{A}_n) \ge 1 - \frac{1}{b_n}$ and $\mathbb{P}(\mathcal{B}_n) \ge 1 - \frac{4}{nF_{r,I}(\tilde{I})} \ge 1 - \frac{3}{\log n}$, where the last inequality follows with Lemma 1 from $F_{r,I}(I) \ge 2 \log n/n$, which in turn is a consequence of (4). Hence, $P_n^0 \to \infty$ with probability converging to 1, uniformly in alternatives satisfying (4). On the other hand, the critical value of P_n^0 stays bounded by Proposition 1.

To prove the claim for P_n note that by (7), we can find $\tilde{I} \in \mathcal{J}_{app}(\ell)$ such that $F_n(I \triangle \tilde{I}) \le \frac{F_n(I)}{3\sqrt{\ell}}$ by taking $\ell := \lfloor \log_2 1/F_n(I) \rfloor + 1$. This index satisfies $\ell \le \ell_{max}$: It is readily seen that (4) implies $F_{r,I}(I) \ge \frac{2 \log n + b_n \sqrt{\log n}}{n}$ for n large enough, hence $\mathbb{P}\left(|F_n(I) - F_{r,I}(I)| \le |\frac{2 \log n}{n} - F_{r,I}(I)| \right) \ge 1 - \frac{3}{b_n}$ by Chebyshev. This implies that with probability converging to one, we can now guarantee firstly that $F_n(I) \ge \frac{2 \log n}{n}$, and hence, $\ell \le \ell_{max}$, secondly, $F_n(I) \le 2F_{r,I}(I)$, hence, $F_n(I \triangle \tilde{I}) \le \frac{F_{r,I}(I)}{\sqrt{\ell}}$. Note that $\tilde{I}$ is a random interval because $\mathcal{J}_{app}$ is constructed w.r.t. F_n. Hence, the previous proof for fixed $\tilde{I}$ does not go through any more, but the claim can be established as in the proof for A_n^{cond} in the succeeding text. There we consider $\tilde{I} \subset I$, which can be enforced previously because still guaranteeing $\ell \le \ell_{max}$. Alternatively, (15) can be readily extended to cover the case $\tilde{I} \not\subset I$. The approximating set $\mathcal{I}_{app}$ used for A_n^{cond} differs from $\mathcal{J}_{app}$ used for P_n in the spacing parameter d_ℓ, but that is not relevant for the part of the proof in the succeeding text that establishes $\sqrt{2logLR_n(F_0(\tilde{I}), F_n(\tilde{I}))} \ge \sqrt{2 \log \frac{e}{F_{r,I}(I)}} + B_n$.

To prove the claim for A_n^{cond}, we consider the collection of all intervals in the approximating set whose endpoints are close to those of I:

$$\mathcal{A}(I) := \{ \tilde{I} \in \mathcal{I}_{app}(\ell) : \tilde{I} \subset I \text{ and } F_n(\tilde{I}) \ge F_n(I)(1 - \eta_n/2) \}$$

where $\eta_n := \min\left(1, \frac{b_n}{2\sqrt{\log e/F_n(I)}}\right)$ and $\ell := \lfloor \log_2 \frac{1}{F_n(I)(1-\eta_n/4)} \rfloor + 1$. Hence, $m_\ell < nF_n(I)(1 - \eta_n/4) \le 2m_\ell$. As mentioned previously one can show that $\ell \in \{2, \ldots, \ell_{max}\}$ with probability converging to one. As in Lemma 2 of Chan & Walther (2013) one finds

$$\frac{\#\mathcal{A}(I)}{\#\mathcal{I}_{app}} \geq C \frac{\eta_n^2 F_n(I)}{(\log_2 e/F_n(I))^{8/5}} \tag{14}$$

Standard considerations using Lemma 1 and (15) show that the event $\left\{\inf_{\tilde{I}\in\mathcal{A}(I)} 1\left(F_n(\tilde{I}) > F_0(\tilde{I})\right) = 1\right\}$ has probability converging to one, hence, on this event

$$\begin{aligned}
\inf_{\tilde{I}\in\mathcal{A}(I)} \sqrt{2logLR_n\left(F_0(\tilde{I}), F_n(\tilde{I})\right)} &\geq \inf_{\tilde{I}\in\mathcal{A}(I)} \sqrt{n}\frac{F_n(\tilde{I}) - F_0(\tilde{I})}{\sqrt{F_0(\tilde{I}) \vee F_n(\tilde{I})}} \quad \text{by (6)} \\
&\geq \inf_{\tilde{I}\in\mathcal{A}(I)} \sqrt{n}\frac{F_{r,I}(\tilde{I}) - F_0(\tilde{I})}{\sqrt{F_{r,I}(\tilde{I}) \vee F_n(\tilde{I})}} - \sup_{\tilde{I}\in\mathcal{A}(I)} \sqrt{n}\frac{F_n(\tilde{I}) - F_{r,I}(\tilde{I})}{\sqrt{F_{r,I}(\tilde{I}) \vee F_n(\tilde{I})}} \\
&\geq \left(\inf_{\tilde{I}\in\mathcal{A}(I)} \sqrt{n}\frac{F_{r,I}(\tilde{I}) - F_0(\tilde{I})}{\sqrt{F_{r,I}(\tilde{I})}}\right)\left(1 - O_p\left(\frac{1}{\sqrt{\log n}}\right)\right) - O_p(1) \quad \text{by (15)} \\
&\geq \sqrt{n}\frac{F_{r,I}(I) - F_0(I)}{\sqrt{F_{r,I}(I)}}(1-\eta_n/2)\left(1 - O_p\left(\frac{1}{\sqrt{\log n}}\right)\right) - O_p(1) \\
&\geq \sqrt{2\log\frac{e}{F_{r,I}(I)(1-F_{r,I}(I))}} + B_n \quad \text{where } B_n := b_n/9 + O_P(1)
\end{aligned}$$

and where the second to last inequality follows from Lemma 1 because

$$\begin{aligned}
1 - \frac{F_{r,I}(I\triangle\tilde{I})}{F_{r,I}(I)} = \frac{F_{r,I}(\tilde{I})}{F_{r,I}(I)} &= \frac{F_n(\tilde{I})}{F_n(I)}\left(1 + O_p\left(\frac{1}{\sqrt{\log n}}\right)\right) \quad \text{by (15)} \\
&\geq (1-\eta_n/2)\left(1 + O_p\left(\frac{1}{\sqrt{\log n}}\right)\right)^2 \quad \text{by the definition of } \mathcal{A}(I).
\end{aligned}$$

Hence, $\inf_{\tilde{I}\in\mathcal{A}(I)} LR_n\left(F_0(\tilde{I}), F_n(\tilde{I})\right) \geq \frac{1}{F_{r,I}(I)(1-F_{r,I}(I))}\exp\left\{B_n\left(B_n/2 + \sqrt{2\log\frac{e}{F_{r,I}(I)}}\right)\right\}$ and so $A_n^{cond} \overset{P}{\to} \infty$ as in the proof of Theorem 3 in Chan & Walther (2013), using (15). Because the critical value of A_n^{cond} stays bounded by Proposition 2, the claim follows.

It remains to show

$$\sup_{\tilde{I}\in\mathcal{A}(I)} \left|1 - \frac{F_{r,I}(\tilde{I})}{F_n(\tilde{I})}\right| = O_p\left(\frac{1}{\sqrt{\log n}}\right) \tag{15}$$

Denote by $X_{(a)}$ the smallest and by $X_{(b)}$ the largest observation in I. Writing $d := b - a$ and $U_i = F_{r,I}(X_i)$:

$$\sup_{\tilde{I}\in\mathcal{A}(I)} \sqrt{n}\frac{\left|F_n(\tilde{I}) - F_{r,I}(\tilde{I})\right|}{\sqrt{F_n(\tilde{I})}} \leq 2 \max_{j=a,\ldots,a+d} \sqrt{n}\frac{\left|U_{(j)} - U_{(a)} - \frac{j-a}{n}\right|}{\sqrt{\frac{d}{2n}}} = O_p(1)$$

by well known facts. Together with $F_n(\tilde{I}) \geq \frac{\log n}{n}$ for $\tilde{I} \in \mathcal{I}_{app}$, this implies (15).

Acknowledgement

Work supported by NSF grant DMS-1007722 and NIH grant AI09851901.

References

Arias-Casto, E., Donoho, D.L. & Huo, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Th.* **51**, 2402–2425.

Chan, H.P. (2009). Detection of spatial clustering with average likelihood ratio test statistics. *Ann. Statist.* **37**, 3985–4010.

Chan, H.P. & Walther, G. (2013). Detection with the scan and the average likelihood ratio. *Statist. Sinica* **23**, 409–428.

Dümbgen, L. (1998). New goodness-of-fit tests and their application to nonparametric confidence sets. *Ann. Statist.* **26**, 288–314.

Dümbgen, L. & Spokoiny, V.G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29**, 124–152.

Dümbgen, L. & Walther, G. (2008). Multiscale inference about a density. *Ann. Statist.* **36**, 1758–1785.

Glaz, J. & Balakrishnan, N. (eds). (1999). *Scan statistics and applications*, Birkhäuser, Boston.

Good, I.J. & Gaskins, R.A. (1980). Density estimation and bump hunting by the penalized maximum likelihood method exemplified by scattering and meteorite data (with discussion). *J. Amer. Statist. Assoc.* **75**, 42–73.

Hartigan, J.A. & Hartigan, P.M. (1985). The DIP test of multimodality. *Ann. Statist.* **13**, 70–84.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30.

Kulldorff, M. (1997). A spatial scan statistic. *Commun. Statist. - Theory Meth.* **26**, 1481–1496.

Loader, C.R. (1991). Large-deviation approximations to the distribution of scan statistics. *Adv. Appl. Prob.* **23**, 751–771.

Minnotte, M.C. & Scott, D.W. (1993). The mode tree: a tool for visualization of nonparametric density features. *J. Comp. Graph. Statist.* **2**, 51–68.

Müller, D.W. & Sawitzki, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.* **86**, 738–746.

Neill, D. & Moore, A. (2004). A fast multi-resolution method for detection of significant spatial disease clusters. *Adv. Neur. Info. Proc. Sys.* **10**, 651–658.

Neill, D.B. (2009a). An empirical comparison of spatial scan statistics for outbreak detection. *Int. J. Health Geogr.* **8**, 1–16.

Neill, D.B. (2009b). Expectation-based scan statistics for monitoring spatial time series data. *Int. J. Forecasting* **25**, 498–517.

Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters - An excess mass approach. *Ann. Statist* **23**, 855–881.

Rufibach, K. & Walther, G. (2010). The block criterion for multiscale inference about a density, with applications to other a density, with applications to other multiscale problems. *J. Comput. Graph. Stat.* **19**, 175–190.

Shorack, G.R. & Wellner, J.A. (1986). *Empirical processes with applications to statistics*, Wiley, New York.

van der Vaart, A.W. & Wellner, J.A. (1996). *Weak convergence and empirical processes with applications to statistics*, Springer, New York.

Walther, G. (2010). Optimal and fast detection of spatial clusters with scan statistics. *Ann. Statist* **38**, 1010–1033.

Received November 2012, in final form April 2013

Guenther Walther, Stanford University.
E-mail: walther@stat.stanford.edu