Biometrika (2018), **xx**, x, *pp*. 1–13

Printed in Great Britain

Continuous testing for Poisson process intensities: A new perspective on scanning statistics

BY FRANCK PICARD

Centre National de la Recherche Scientifique,
Laboratoire de Biométrie et Biologie Evolutive,
43 boulevard du 11 novembre 1918 F-69622 Villeurbanne, France.
franck.picard@univ-lyon1.fr

AND PATRICIA REYNAUD-BOURET,

Centre National de la Recherche Scientifique,
Laboratoire Jean Alexandre Dieudonné,
Parc Valrose F-06108 Nice, France.
Patricia.Reynaud-Bouret@unice.fr

AND ETIENNE ROQUAIN

Sorbonne Université, 4 place Jussieu F-75005, Paris, France.
etienne.roquain@upmc.fr

S-1. KERNEL-BASED STATISTICS FOR SLIDING WINDOWS

S-1·1. *Two-sided test for the two-sample problem*

The kernel framework for testing proposed by Gretton et al. (2012) and Fromont et al. (2013) is based on the estimation of a distance between the functions ν_A and ν_B in a reproducing kernel Hilbert space, or in L^2. In particular, Fromont et al. (2013) show that these test statistics are minimax with respect to various smoothness classes, and in this sense, powerful against some alternatives that cannot be distinguished by count-based tests. The shape of the statistic is given by

$$S(x) = \sum_{T \neq T' \in N \cap I_x} K_h(T - T')\varepsilon_T \varepsilon'_T,$$

where N is the joint process and the ε_T's are the marks; see Section 2·3. In particular, it only depends on the composition of the window. To understand why this statistic is meaningful, let us compute its expectation. By classical computations for Poisson processes,

$$\begin{aligned} E_{\theta,\lambda}\left\{S_\eta(x)\right\} &= E_\lambda\left\{\sum_{T \neq T' \in N \cap I_x} K_h(T - T')\theta(T)\theta(T')\right\} \\ &= \int_{I_x^2} K_h(t - t')\theta(t)\theta(t')\lambda(t)\lambda(t')dtdt' \\ &= \int_{I_x^2} K_h(t - t')\left\{\nu_A(t) - \nu_B(t)\right\}\left\{\nu_A(t') - \nu_B(t')\right\}dtdt'. \end{aligned}$$

In particular, we obtain, under classical assumptions on the smoothness of $\nu_A - \nu_B$, that

$$E_{\theta,\lambda}\left\{S(x)\right\} \underset{h\to 0}{\longrightarrow} \int_{I_x} \left\{\nu_A(t) - \nu_B(t)\right\}^2 dt.$$

This short derivation illustrates the motivation to use $S(x)$ as a test statistic for two-sided tests in the two-sample case. Large values of $S(x)$ indicate large differences between the functions ν_A and ν_B. Note that from a computational point of view, the complexity for computing $S(x)$ is of the order of $N^2(I_x)$. However, thanks to the piece-wise constant property of $S(x)$ on partition τ, the continuum $\{S(x), x \in \mathcal{X}\}$ can be computed using the in/out points of each scanning window, which drastically reduces the global computational burden.

S-1·2. *Two-sided test for homogeneity detection*

Conditionally on the event $\{N([0,1]) = n\}$, the points in N form an n independent and identically distributed sample with density $f = 1 + \theta$. Hence,

$$\widehat{f}_h(s) = \frac{1}{n} \sum_{T \in N \cap I_x} K_h(s - T),$$

is an unbiased consistent estimate when n tends to infinity of

$$f_h(s) = \int_{I_x} K_h(s-t) f(t) dt.$$

Consequently, the corresponding U-statistics given by

$$\frac{1}{n(n-1)} \sum_{\substack{T' \in N \cap I_x \\ T' \neq T}} K_h(T - T'), \tag{S-1}$$

is an unbiased estimate of:

$$\int_{I_x^2} K_h(s-t) f(t) f(s) ds \underset{h\to 0}{\longrightarrow} \|f\|_{I_x}^2.$$

Note that in a testing framework, we do not want to estimate $\|f\|_{I_x}^2$ and we do not need to consider small values for h, we even recommend to take $h = \eta$ when it is unknown. This choice would not be appropriate for estimation, but seems sufficient for testing. Next, if $\mathcal{H}_0(I_x)$ holds, $f = 1$ on I_x and $\|f\|_{I_x}^2 = \eta$. So we propose the following statistic

$$S(x) = \left| \frac{1}{n(n-1)} \sum_{\substack{T' \in N \cap I_x \\ T' \neq T}} K_h(T - T') - \eta \right|,$$

and the local null hypothesis is rejected for high values of $S(x)$.

S-1·3. *One-sided test for homogeneity detection*

In this case we focus on what is happening when $f > 1$. Hence the target quantity to estimate is

$$\int_{I_x} f(s) \max\left\{f(s), 1\right\} ds.$$

Here we consider the following density estimator:

$$\widehat{f}_h(s) = \frac{1}{N([0,1]) - 1} \sum_{\substack{T' \in N \cap I_x \\ T' \neq s}} K_h\left(T' - s\right),$$

which leads to the following statistics:

$$S(x) = \frac{1}{N([0,1])} \sum_{T \in N \cap I_x} \max\left\{\widehat{f}_h(T), 1\right\}.$$

If there exists a t in I_x such that $f(t) > 1$ and if f is smooth enough such that this holds on a neighborhood, $E_\theta\left[\max\left\{f(T), 1\right\} 1_{\{T \in I_x\}}\right]$ should be larger than η, whereas its value should be smaller than η if there are no such t.

S-1·4. *One-sided test for the two-sample problem*

Applying a reasoning similar to above, the quantity

$$\frac{1}{N([0,1]) - 1} \sum_{\substack{T' \in N \cap I_x \\ T' \neq T}} K_h(T - T')\varepsilon_{T'},$$

is a good estimate of $\theta(T)\lambda(T)/\int \lambda$, for each T in N. If one wants to reject only when there exists a $t \in I_x$ such that $\theta(t) > 0$, then one needs to take the positive part of the previous quantity and integrate it with respect to T in $N \cap I_x$. We do not want to multiply again by ε_T because we do not want to multiply by θ twice. This leads to

$$S(x) = \frac{1}{N([0,1])} \sum_{T \in N \cap I_x} \max\left\{ \frac{1}{N([0,1]) - 1} \sum_{\substack{T' \in N \cap I_x \\ T' \neq T}} K_h(T - T')\varepsilon_{T'} \ , \ 0 \right\},$$

which estimates $E_\lambda\left[\max\left\{\theta(T)\lambda(T), 0\right\} 1_{\{T \in I_x\}}\right]$ for T random variable of density proportional to λ.

S-2. MONTE-CARLO PROCEDURES FOR p-VALUES ESTIMATION AND ADJUSTMENT

Depending on whether the distribution of the single test statistic is explicit or not, the number of Monte Carlo steps that are needed in practice differ. More precisely, for tests based on the count statistic the p-value process is explicit. Hence, no need for Monte Carlo in this case. Moreover, since our weighted Benjamini–Hochberg procedure (Section 4·4) is also completely explicit, $\mathcal{R}_\alpha^{wBH}$ can be easily computed in practice.

However, the $\min$-p procedure based on count-statistics is not explicit since it requires the conditional distribution of $\inf_{x \in \mathcal{X}} p(x)$ given $\mathfrak{N}$ under the full null hypothesis. Therefore, we need to use a Monte Carlo procedure to approach this distribution. This can be achieved while maintaining a valid family-wise error rate control by following Romano and Wolf (2005a), as detailed below.

When dealing with kernel-based tests the p-value process is not even explicit; see Section S-1. One can use again the estimated p-values developed by Romano and Wolf (2005b) since their result shows that they satisfy (2) and $\mathcal{R}_\alpha^{wBH}$ can be easily computed in practice and only requires one Monte Carlo step. However, there is no theoretical false discovery rate guarantee and we only show the false discovery rate control by simulations in Section 5.

As for the $\min$-p procedure based on kernel statistics, it requires an additional Monte Carlo step. To this end, a classical and sufficient method is to perform an approximation using two independent Monte Carlo samples, which we use for homogeneity test. For the two-sample case, the Monte Carlo procedure is quite demanding and is close in spirit to bootstrap methods. In this case, we develop a double Monte Carlo procedure, which only requires one Monte-Carlo sample and still guarantees a controlled family-wise error rate, Section S-2·3.

S-2·1. *Monte Carlo estimation of the kernel-based p-value process for the homogeneity test*

Given $\mathfrak{N} = N([0,1])$, the distribution of $S(x)$ under $\mathcal{H}_0(I_x)$ does not depend on x. Let us focus on the first window $I_{\eta/2} = (0, \eta]$ and simulate points inside this window under $\mathcal{H}_0(I_{\eta/2})$. Therefore we consider B random samples of the number of points inside $I_{\eta/2}$, n_b for all $b = 1, \dots, B$, with conditional distribution $\mathcal{B}(n, \eta)$ given $N([0,1]) = n$, and for each b, we draw N^b the point process restricted to $I_{\eta/2}$ as the realization of a n_b independent sample of uniform variables on $(0, \eta]$. For each b, one can compute thanks to N^b the value of the test statistic for the first window, that we denote S_b. But since all the statistics are distribution invariant by translation under the null, one can also say that this sample has the same distribution as $S(x)$ under $\mathcal{H}_0(I_x)$ for every window center x in $\mathcal{X}$.

Therefore, whatever B, one has easily access to $\big(S_1, \dots, S_B\big)$, B independent variables with the same distribution as $S(x)$ under $\mathcal{H}_0(I_x)$ for every x. We use as estimated p-value process:

$$\widehat{p}(x) = \frac{1}{B+1}\left[1 + \sum_{b=1}^{B} 1_{\{S_b \geq S(x)\}}\right],$$

which satisfies Property 2 by an application of Romano and Wolf (2005b).

S-2·2. *Monte Carlo estimation of the kernel-based p-value process for the two-sample case*

We label the observed marks $\varepsilon^0 = \varepsilon$ such that it constitutes the first term of a $(B+1)$-sample of marks, filled by B independent draws of independent Rademacher sets, $\varepsilon^b := (\varepsilon^b_T)_{T \in N}$, for $b = 1, \dots, B$, whose distribution is the one that the observed marks ε_T's should have under the full null hypothesis. As previously, the Monte Carlo approximated conditional p-value of window I_x is defined by:

$$\widehat{p}(x) = \frac{1}{B+1}\left[1 + \sum_{b=1}^{B} 1_{\{S_b(x) \geq S_0(x)\}}\right], \tag{S-2}$$

where $S_0(x) = S(x)$ is the observed statistic and where $S_b(x)$ is the value of the test statistic on I_x that is computed with the resampled marks ε^b and fixed joint process N. Let us underline that since in the two-sample case the partition τ only depend on N and do not vary with the resampling scheme ε^b, the processes $\{S_b(x)\}_{x \in \mathcal{X}}$, for $b = 0, \dots, B$, and therefore the process $\{\widehat{p}(x)\}_{x \in \mathcal{X}}$ are piece-wise constant on τ.

S-2·3. *Two-step Monte-Carlo method for the $\min$-p procedure.*

Recall that when using the $\min$-p procedure with a p-value process that is not explicit, we need two Monte-Carlo approximations.

In the homogeneity case, the distribution of the test statistic under the null does not depend on the window centers, while the distribution of the minimum of the p-values relies on the observations over all windows. Hence, it is quite natural and straightforward to perform two separated Monte Carlo schemes for these two steps in a direct manner.

In the two-sample case, the trick is that one can use the same sample of (N, ε^b)'s as the one used in (S-2) for both approximations. More precisely, the p-values $\widehat{p}_b(x)$, corresponding to the simulated process (N, ε^b), can be computed as

$$\widehat{p}_b(x) = \frac{1}{B+1} \sum_{b'=0}^{B} 1_{\{S_{b'}(x) \geq S_b(x)\}},$$

that is we do the same computation as in Eq. S-2 except that we replace the observed statistics $S_0(x)$ by the simulated one $S_b(x)$, hereby computing a "resampled" p-value process corresponding to ε^b. Then, let us define, for all $b = 0, \ldots, B$

$$m^b = \inf_{x \in \mathcal{X}_\eta} \widehat{p}_b(x).$$

The corresponding adjusted Monte-Carlo p-values are then given by

$$\widehat{q}(x) = \frac{1}{B+1} \left[1 + \sum_{b=1}^{B} 1_{\left\{m^b \leq \widehat{p}(x)\right\}} \right].$$

Finally the rejection set is given by

$$\widehat{\mathcal{R}}^{\inf} = \{x \in \mathcal{X}, \widehat{q}(x) \leq \alpha\}.$$

The previous double Monte-Carlo procedure is based on the use of the same random signs $\left(\varepsilon_T^b\right)_{T \in N}$, $b = 0, \ldots, B$ and not on two sets of different simulations. Indeed the ε^b's are used twice: a first time to estimate the p-values of the observed process and a second time to give the p-values even on the simulated realizations leading to the adjusted p-values . This procedure has the double advantage of sparing computational resource, while enjoying the same control property as shown hereafter.

THEOREM S-1. *Let $\alpha \in (0, 1)$. For the two-sample case and for all (θ, λ), the continuous testing procedure defined by* (S-2·3) *satisfies*

$$\mathrm{FWER}_{\theta,\lambda}(\widehat{\mathcal{R}}^{\inf}) \leq \alpha.$$

This result holds whatever the choice of the single test statistic that only depends on the composition of the window. It is proved in Section S-5·3.

S-3. PRACTICAL ISSUES

S-3·1. *Construction of the partition τ*

If the partition τ is obvious in the homogeneity case, the two-sample case is more difficult. In this case, occurrences from N_A and N_B are merged to form the joint process N and Rademacher marks $(\varepsilon_T)_T$ are introduced as labels, see Fig. S-1. Each occurrence $T \in N$ has a span η that is used to create the partition τ whose elements constitute the centers of the testing windows I_{τ_m}, $(\tau_m \in \tau)$. Such construction of the windows ensures that the composition of the observed point processes is constant between two centers τ_m and τ_{m+1} in terms of number and repartition of points. The p-value process is then a càdlàg process with jumps defined by partition τ.

S-3·2. *Description of the full procedure in practice*

In the applications, the p-value process is computed as well as its $\min$-p and weighted Benjamini–Hochberg adjusted versions. In order to propose a completely operational procedure,

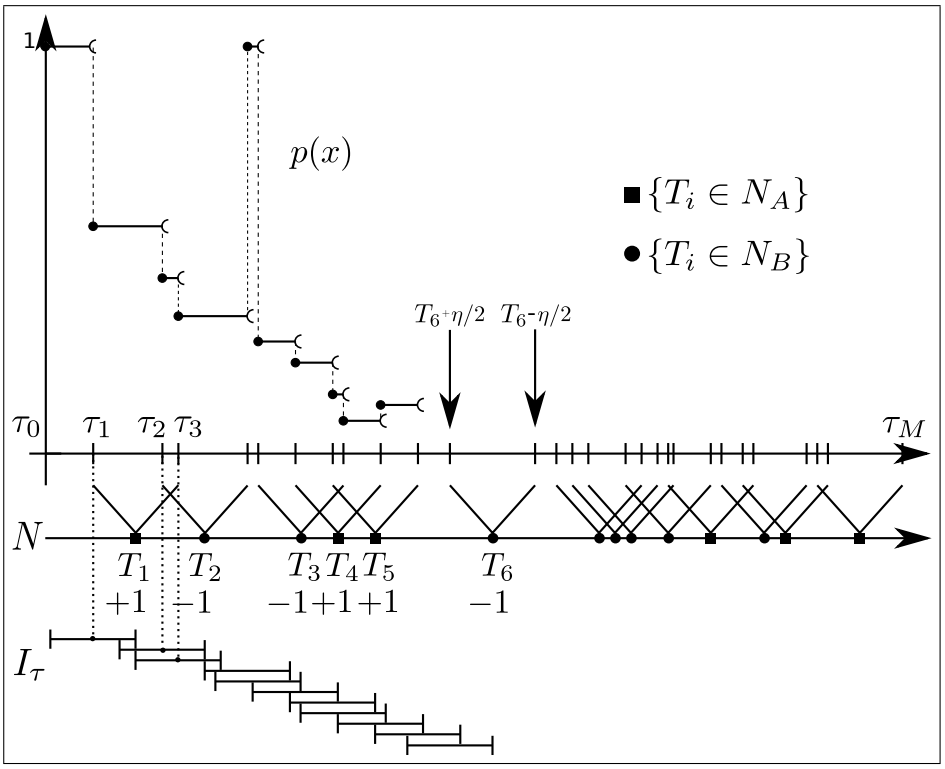

Fig. S-1. Construction of the partition τ for the two-sample test.

we need to clarify the distinction between the point space and the window space on which the decision is made. Indeed, if window I_x is accepted by our procedure without mistake, then one is sure that $H_{0,t}$ holds for all point t in I_x, whereas if a window I_x is rejected without mistake, it is yet not clear whether there is not a smaller interval included in I_x on which the null hypothesis holds. Consequently, we start by defining the set of accepted windows

$$\widehat{\mathcal{A}} = \{x \in \mathcal{X} \ : \ q(x) > \alpha\},$$

with $q(x)$ denoting the adjusted p-value process (possibly estimated). The graphical representation of our procedure is then based on the estimated set of points that only lie within rejected windows, namely:

$$\widehat{\mathcal{I}}_1 = [0,1] \backslash \widehat{\mathcal{I}}_0,$$

where $\widehat{\mathcal{I}}_0 = \Big\{ t \in [0,1] \ : \ \exists x \in \widehat{\mathcal{A}} \text{ s.t. } t \in I_x \Big\}$; see the gray rectangle in Fig. 2, main text.

S-3·3. *Simulations under the full null hypothesis for the two-sample test*

We consider the two sample test under the full null hypothesis, $\mathcal{H}_0([0,1])$ that corresponds to $\{\theta = \theta_0\}$ (Section 3·1). To proceed we simulate a homogeneous Poisson process with intensity λ, along with $\varepsilon = (\varepsilon_T)_{T \in N}$, a set of marks with values in $\{-1, 1\}$ that are sampled with a Bernoulli distribution or parameter $1/2$; see Section 2·3. Then our procedure is run with window size $\eta = 0.05$.

Table S-1. *Error rates control under the full null hypothesis for the two-sample test (two-sided). The level of the test is* $\alpha = 10\%$*, the window size is* $\eta = 0.05$.

	FWER min-p		FDR wBH	
λ	Count	Kernel	Count	Kernel
500	0.8	10.6	7.1	5.5
1000	0.4	9.5	4.7	4.2
5000	0.2	9.6	4.9	5.0

FWER, family-wise error rate; FDR, false discovery rate; CZ, Chan and Zhang method; SZY, Siegmund, Zhang and Yakir method with 5 or 50 windows; wBH, weighted Benjamini–Hochberg procedure.

S-3·4. *Exact expression of* θ *in the simulations*

For a fixed parameter r, we simulate a heterogeneous Poisson process corresponding to the following piecewise constant signal θ:

$$\theta(t) = \begin{cases} +\theta^*, & \text{for } t \in \mathcal{I}_1^+ \\ -\theta^*, & \text{for } t \in \mathcal{I}_1^- \\ 0, & \text{otherwise.} \end{cases}$$

with

$$\mathcal{I}_1^+ = \left[\frac{1}{4} - \frac{r}{4}, \frac{1}{4} + \frac{r}{4}\right] \cup \left[\frac{1}{2} - \frac{r}{4}, \frac{1}{2} + \frac{r}{4}\right] \cup \left[\frac{3}{4} - \frac{r}{4}, \frac{3}{4} + \frac{r}{4}\right],$$

and

$$\begin{aligned} \mathcal{I}_1^- = &\left[\frac{1}{4} - \frac{r}{2}, \frac{1}{4} - \frac{r}{4}\right] \cup \left[\frac{1}{4} + \frac{r}{4}, \frac{1}{4} + \frac{r}{2}\right] \\ &\cup \left[\frac{1}{2} - \frac{r}{2}, \frac{1}{2} - \frac{r}{4}\right] \cup \left[\frac{1}{2} + \frac{r}{4}, \frac{1}{2} + \frac{r}{2}\right] \\ &\cup \left[\frac{3}{4} - \frac{r}{2}, \frac{3}{4} - \frac{r}{4}\right] \cup \left[\frac{3}{4} + \frac{r}{4}, \frac{3}{4} + \frac{r}{2}\right]. \end{aligned}$$

Then $\mathcal{I}_1 = \mathcal{I}_1^+ \cup \mathcal{I}_1^-$, $\mathcal{I}_0 = [0,1] \backslash \mathcal{I}_1$, denote the positions on which θ is not null. We consider this particular shape of θ to ensure that $\int_0^1 \theta(t)dt = 0$. This function is displayed in Fig. S-2.

S-3·5. *Assessing the performance of methods based on true positives*

In order to also compare methods based on type-II errors, we propose two indicators that are the type-II analogues of the family-wise error rate and of the false discovery rate. To assess a family-wise type-II error rate we consider the following probability:

$$P_{\theta,\lambda}(J_0^c \cap \mathcal{R} \neq \emptyset), \tag{S-3}$$

with $J_0^c \cap \mathcal{R}$ being the set of true positive windows. To obtain a rate, we also consider the true positive rate defined such as:

$$\mathrm{TPR}_{\theta,\lambda}(\mathcal{R}) = E_{\theta,\lambda}\left\{\frac{\Lambda\left(J_0^c \cap \mathcal{R}\right)}{\Lambda\left(J_0^c \cap \mathcal{R}\right) + \Lambda\left(J_0^c \cap \mathcal{A}\right)}\right\} = E_{\theta,\lambda}\left\{\frac{\Lambda\left(J_0^c \cap \mathcal{R}\right)}{\Lambda\left(J_0^c\right)}\right\}, \tag{S-4}$$

with $J_0^c \cap \mathcal{A}$ being the set of false negative windows. This indicators provide a measure of the power of different methods while considering the multiplicity of the test. The comparison of these type-II error rates should be done cautiously since the type-I error rates are not comparable between different methods (Fig. 1, main text).

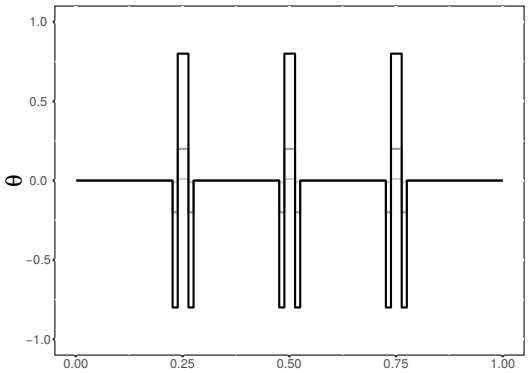

Fig. S-2. Simulation setting with an alternative for the homogeneity case. Signal function $\theta(.)$ that equals 0 under the null hypothesis, and θ^* elsewhere. Black line : $\theta^* = 0.8$, dark grey line $\theta^* = 0.2$, light grey line $\theta^* = 0.01$.

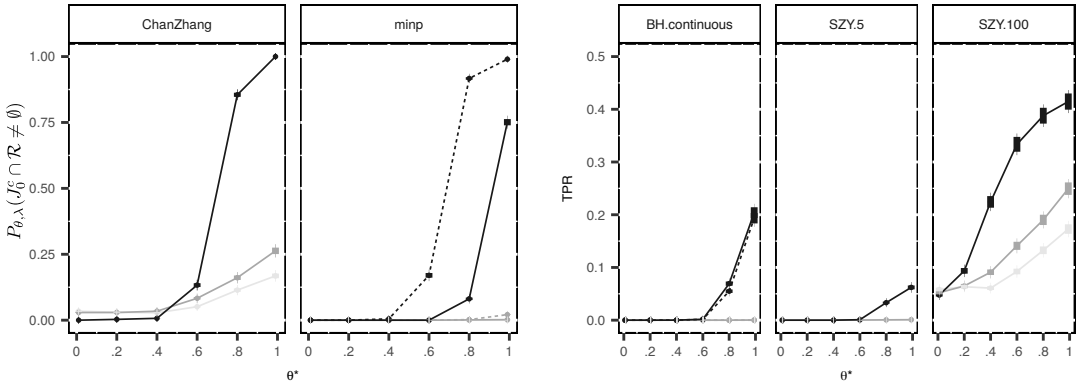

Fig. S-3. Comparing true positive performance between methods($\alpha = 10\%$). Black line: $\lambda = 5000$, dark grey line: $\lambda = 1000$, light grey line: $\lambda = 500$. Plain line: count statistic, dotted line: kernel statistic. ChanZhang: Chan and Zhang (2007) method, SYZ: Siegmund et al. (2011) method with 5 and 100 windows of identical length.

S-4. FURTHER EXTENSIONS FOR THE min-p PROCEDURE

S-4·1. *Step-down improvement*

We now provide the step-down version of our family-wise error rate controlling procedure. Let us first introduce some notation. For any fixed subset $\mathcal{C}$ of $\mathcal{X}$, denote $F^{\mathcal{C}}_{\theta_0,\mathfrak{N}}$ the conditional cumulative distribution function of $\inf_{x\in\mathcal{C}} p(x)$ given $\mathfrak{N}$. Then we define

$$\mathcal{N}(\mathcal{C}) = \left\{x \in \mathcal{X}, F^{\mathcal{C}}_{\theta_0,\mathfrak{N}}\{p(x)\} \leq \alpha\right\}.$$

The theoretical step-down algorithm can then be derived as follows: Step 0 : compute $\mathcal{R}^0 = \emptyset$; Step $j \geq 1$: compute $\mathcal{R}^j = \mathcal{N}\left\{\left(\mathcal{R}^{j-1}\right)^c\right\}$. The step-down rejection set $\mathcal{R}^{\text{step-down}}$ is then defined as the limit of $\mathcal{R}^j$ when j tends to infinity. Note that $\mathcal{R}^1 = \{x \in \mathcal{X} \,:\, q(x) \leq \alpha\}$ and that the sequence of $\mathcal{R}^j$ is increasing.

The fact that this algorithm ends almost surely in a finite number of steps is just an easy consequence of the fact that for a given realization, there is only a finite number of possible values for the p-value process, which is in fact upper bounded by the size of the partition τ. Indeed, at each round, $\mathcal{R}^j$ cannot decrease and if it increases, it is by absorbing sets of the type $\{x \in \mathcal{X}, p(x) = a\}$ for some a. We show that such an algorithm guarantees a controlled family-wise error rate below in Section S-5·4.

S-5. PROOFS

S-5·1. *Proof of Theorem 1*

By definition, we have

$$\begin{aligned}
\text{FWER}_{\theta,\lambda}(\mathcal{R}^{\inf}) &= E_\lambda\left[P_\theta\left\{\exists x \in J_0, q(x) \leq \alpha \;\middle|\; \mathfrak{N}\right\}\right] \\
&\leq E_\lambda\left(P_\theta\left[F^{\inf}_{\mathfrak{N}}\left\{\inf_{x\in J_0} p(x)\right\} \leq \alpha \;\middle|\; \mathfrak{N}\right]\right) \\
&= E_\lambda\left(P_{\theta_0}\left[F^{\inf}_{\mathfrak{N}}\left\{\inf_{x\in J_0} p(x)\right\} \leq \alpha \;\middle|\; \mathfrak{N}\right]\right) \\
&\leq E_\lambda\left(P_{\theta_0}\left[F^{\inf}_{\mathfrak{N}}\left\{\inf_{x\in \mathcal{X}} p(x)\right\} \leq \alpha \;\middle|\; \mathfrak{N}\right]\right),
\end{aligned}$$

where we used property $(\mathcal{P})$. The result comes from the fact that

$$P_{\theta_0}\left[F^{\inf}_{\mathfrak{N}}\left\{\inf_{x\in \mathcal{X}} p(x)\right\} \leq \alpha \;\middle|\; \mathfrak{N}\right] \leq \alpha,$$

because under θ_0, the quantity $F^{\inf}_{\mathfrak{N}}\left\{\inf_{x\in\mathcal{X}} p(x)\right\}$ can be seen as the p-value of the test based on the statistics $-\inf_{x\in\mathcal{X}} p(x)$, see for instance Lemma 1 of Fromont et al. (2016).

S-5·2. *Proof of Theorem 2*

We follow the methodology introduced in Blanchard et al. (2014) by applying their Theorem 4.1. While the measurability conditions can be easily checked because the p-value process is càdlàg, the result will be proved if we show that the p-value process is finite-dimensional strong Positive Regression Dependent on each one from a Subset (PRDS) on J_0, as defined in Blanchard et al. (2014).

For the homogeneity test, Lemma A.2 in Blanchard et al. (2014) shows that the p-value process is finite-dimensional strong PRDS on any subset, and thus also on J_0, which shows the desired false discovery rate control.

Let us now consider the two-sample case. In that situation, we prove in fact the stronger statement

$$E_\theta \left\{ \frac{\Lambda\left(J_0 \cap \mathcal{R}_\alpha^{wBH}\right)}{\Lambda\left(\mathcal{R}_\alpha^{wBH}\right)} \mid N \right\} \leq \alpha, \tag{S-5}$$

which can be seen as a "conditional false discovery rate control". The standard false discovery rate control easily follows from (S-5) by an integration. By Theorem 4.1 of Blanchard et al. (2014), (S-5) is proved as soon as we show that the p-value process is finite-dimensional strong PRDS on J_0 conditionally on N. This condition takes the following form: for any $q \geq 1$, for any $(x_j)_{j=1,\dots,q} \in \mathcal{X}^q$ such that $x_1 \in J_0$, for any nonincreasing set $D \subset \mathbb{N}^q$,

$$n \mapsto P_\theta \left[\{S(x_j)\}_{j=1,\dots,q} \in D \mid S(x_1) = n, N \right]. \tag{S-6}$$

is nonincreasing. Above, a nonincreasing set is defined as a subset $D \subset \mathbb{N}^q$ such that for all $(z, z') \in \mathbb{N}^q$ with $z_j \leq z'_j$ for $j = 1, \dots, q$, we have $z' \in D \Rightarrow z \in D$. Also, in (S-6), we adopt the convention that the conditional probability is equal to 0 whenever the event $S(x_1) = n$ is of probability 0. Equivalently, the nonincreasingness in (S-6) should be proven only for n in the range $\{0, \dots, m\}$ where $m = N(I_{x_1})$ is the maximum possible value for $S(x_1)$.

Now, let us denote $p = N([0,1]) \geq m$ the total number of points of N and write

$$N = (T_i)_{i=1,\dots,p},$$

where the T_i's are distinct points of N. To shorten the notation, we also assume that the labeling of the T_i's ensures that $T_1, \dots, T_m$ belongs to I_{x_1}. For $i = 1, \dots, p$, the mark of each T_i is denoted by Δ_i, so that the variables Δ_i, $i = 1, \dots, p$, are mutually independent and Δ_i is Bernoulli with parameter $\theta(T_i)$. Since $x_1 \in J_0$, note that Δ_i, $i = 1, \dots, m$, are independent and identically distributed Bernoulli variables with parameter $1/2$. Let us denote $\mathbb{P}_{\theta,p}$ the distribution of the whole vector $\Delta = (\Delta_i)_{i=1,\dots,p}$. We thus have

$$S(x_j) = \sum_{i : T_i \in I_{x_j}} \Delta_i, \quad j = 1, \dots, q \quad \text{with} \quad S(x_1) = \sum_{i=1}^{m} \Delta_i.$$

Let us now define

$$C = \left\{ b \in \{0,1\}^p \, : \, \left(\sum_{i : T_i \in I_{x_j}} b_i \right)_{j=1,\dots,q} \in D \right\}.$$

Since D is a nonincreasing set, so is C. It is easy to see that

$$P_\theta \left[\{S(x_j)\}_{j=1,\dots,q} \in D \mid S(x_1) = n, N \right] = \mathbb{P}_{\theta,p} \left(\Delta \in C \mid \sum_{i=1}^{m} \Delta_i = n \right).$$

Therefore it remains to prove that this last quantity decreases with n in the range $\{0, \dots, m\}$. For this, define for two given vectors $y \in \{0,1\}^m$ and $z \in \{0,1\}^{p-m}$, the vector $yz \in \{0,1\}^p$ as the concatenation of y and z. Denote also for all $z \in \{0,1\}^{p-m}$

$$C_z = \{ y \in \{0,1\}^m \mid yz \in C \}$$

and for all $b \in \{0,1\}^p$, $b^m = (b_i)_{i=1,\dots,m} \in \{0,1\}^m$ and $b^{-m} = (b_i)_{i=m+1,\dots,p} \in \{0,1\}^{p-m}$. Let us also denote $\mathbb{P}_{\theta,p,z}(.) = \mathbb{P}_{\theta,p}(. \mid \Delta^{-m} = z)$ the conditional distribution with respect to Δ^{-m}.

It is easy to see that

$$\mathbb{P}_{\theta,p}\left(\Delta \in C \mid \sum_{i=1}^{m}\Delta_i = n\right) = \mathbb{E}_{\theta,p}\left\{\mathbb{P}_{\theta,p,\Delta^{-m}}\left(\Delta \in C \mid \sum_{i=1}^{m}\Delta_i = n\right)\right\}$$
$$= \mathbb{E}_{\theta,p}\left\{\mathbb{P}_{\theta,p,\Delta^{-m}}\left(\Delta^m \in C_{\Delta^{-m}} \mid \sum_{i=1}^{m}\Delta_i = n\right)\right\}.$$

As a consequence, to finish the proof, it is sufficient to show that

$$n \in \{0,\dots,m\} \mapsto \mathbb{P}_{\theta,p,z}\left(\Delta^m \in C_z \mid \sum_{i=1}^{m}\Delta_i = n\right) \text{ is nonincreasing, for any } z \in \{0,1\}^{p-m}. \tag{S-7}$$

Let $z \in \{0,1\}^{p-m}$. First, C_z is a nonincreasing set since C is. Second, consider $J_1,\dots,J_m$ a sequence of indices that are taken uniformly at random without replacement in $\{1,\dots,m\}$ and denote by $\mathbb{P}_J$ the distribution of the m-tuple $(J_1,\dots,J_m)$. Since Δ^m is independent of Δ^{-m}, conditionally to $\sum_{i=1}^{m}\Delta_i = n$, the set of positions of the 1's in Δ_m has exactly the same distribution as the set $\{J_1,\dots,J_n\}$. This entails

$$\mathbb{P}_{\theta,p,z}\left(\Delta^m \in C_z \mid \sum_{i=1}^{m}\Delta_i = n\right) = \mathbb{P}_J\left(\delta_{\{J_1,\dots,J_n\}} \in C_z\right),$$

where for any subset A of $\{1,\dots,m\}$, δ_A is the vector of $\{0,1\}^m$ with 1's exactly at the positions i in A. Now, for $n < m$, $\delta_{\{J_1,\dots,J_n\}} \leq \delta_{\{J_1,\dots,J_{n+1}\}}$ and since C_z is a nonincreasing set, one has

$$\delta_{\{J_1,\dots,J_{n+1}\}} \in C_z \Rightarrow \delta_{\{J_1,\dots,J_n\}} \in C_z.$$

Hence, for $n < m$,

$$\mathbb{P}_J\left(\delta_{\{J_1,\dots,J_{n+1}\}} \in C_z\right) \leq \mathbb{P}_J\left(\delta_{\{J_1,\dots,J_n\}} \in C_z\right),$$

which shows (S-7) and thus the result.

S-5·3. *Proof of Theorem S*-1

By using (7) and thanks to Property 1, that is used for the observed sample and the fact that the distribution of the resampled ε^b for $b = 1,\dots,B$, is fixed whatever θ, we have that

$$\begin{aligned}
\mathrm{FWER}_{\theta,\lambda}(\widehat{\mathcal{R}}^{\inf}) &= E_\lambda\left[P_\theta\left\{\exists x \in J_0, \widehat{q}(x) \leq \alpha \mid \mathfrak{N}\right\}\right] \\
&= E_\lambda\left\{P_\theta\left(\frac{1}{B+1}\left[1 + \sum_{b=1}^{B} 1_{\left\{m^b \leq \inf_{x\in J_0}\widehat{p}(x)\right\}}\right] \leq \alpha \mid \mathfrak{N}\right)\right\} \\
&= E_\lambda\left\{P_{\theta_0}\left(\frac{1}{B+1}\left[1 + \sum_{b=1}^{B} 1_{\left\{m^b \leq \inf_{x\in J_0}\widehat{p}(x)\right\}}\right] \leq \alpha \mid \mathfrak{N}\right)\right\} \\
&\leq E_\lambda\left\{P_{\theta_0}\left(\frac{1}{B+1}\left[1 + \sum_{b=1}^{B} 1_{\left\{m^b \leq \inf_{x\in \mathcal{X}}\widehat{p}(x)\right\}}\right] \leq \alpha \mid \mathfrak{N}\right)\right\}.
\end{aligned}$$

As a consequence, by using the notation m^0, we have

$$\mathrm{FWER}_{\theta,\lambda}(\widehat{\mathcal{R}}^{\inf}) \leq E_\lambda\left[P_{\theta_0}\left\{\frac{1}{B+1}\left(1 + \sum_{b=1}^{B} 1_{\left\{m^b \leq m^0\right\}}\right) \leq \alpha \mid \mathfrak{N}\right\}\right].$$

Since, under P_{θ_0}, the random variables $\varepsilon^0, \ldots, \varepsilon^B$ are independent identically distributed Rademacher, the process vector $[\{\widehat{p}^b(x)\}_{x\in\mathcal{X}}]_{b=0,\ldots,B}$ is exchangeable under P_{θ_0} and conditionally on N, which in turn implies that the vector $(m^b)_{b=0,\ldots,B}$ is exchangeable under P_{θ_0} and conditionally on N. Therefore, applying Lemma 1 of Romano and Wolf (2005b), we obtain

$$P_{\theta_0}\left\{\frac{1}{B+1}\left(1+\sum_{b=1}^{B}1_{\{m^b\le m^0\}}\right)\le\alpha \mid \mathfrak{N}\right\}\le\alpha,$$

and it remains to integrate with respect to N to conclude.

S-5·4. *Control of the step-down algorithm*

We follow the approach proposed in Goeman and Solari (2010). Let us consider the event

$$\mathcal{E}=\{J_0\cap\mathcal{N}[J_0]=\emptyset\}.$$

Then, by (1), we have

$$\begin{aligned}P_{\theta,\lambda}(\mathcal{E}^c)&=P_{\theta,\lambda}\left[\exists x\in J_0, F^{J_0}_{\theta_0,\mathfrak{N}}\{p(x)\}\le\alpha\right]\\&=P_{\theta_0,\lambda}\left[\exists x\in J_0, F^{J_0}_{\theta_0,\mathfrak{N}}\{p(x)\}\le\alpha\right]\\&=E_\lambda\left(P_{\theta_0}\left[F^{J_0}_{\theta_0,\mathfrak{N}}\left\{\inf_{x\in J_0}p(x)\right\}\le\alpha \mid \mathfrak{N}\right]\right)\le\alpha.\end{aligned}$$

The last point comes from the fact that we have point-wise,

$$P_{\theta_0}\left[F^{J_0}_{\theta_0,\mathfrak{N}}\left\{\inf_{x\in J_0}p(x)\right\}\le\alpha \mid \mathfrak{N}\right]\le\alpha,$$

because

$$F^{J_0}_{\theta_0,\mathfrak{N}}\left\{\inf_{x\in J_0}p(x)\right\}$$

can be seen as the p-value of the test based on the statistics $-\inf_{x\in J_0}p(x)$ and described for instance in Lemma 1 of Fromont et al. (2016).

Following the methodology of Goeman and Solari (2010), we have proved the so-called single-step property. Now, the following monotony property should be used: if $\mathcal{C}\subset\mathcal{C}'$ then $F^{\mathcal{C}}_{\theta_0,\mathfrak{N}}\le F^{\mathcal{C}'}_{\theta_0,\mathfrak{N}}$ and $\mathcal{N}(\mathcal{C}')\subset\mathcal{N}(\mathcal{C})$. The end of the reasoning is then as follows: for all j, if $J_0\subset(\mathcal{R}^j)^c$, $R^{j+1}=\mathcal{N}\{(\mathcal{R}^j)^c\}\subset\mathcal{N}(J_0)$. Now, on $\mathcal{E}$, $\mathcal{N}(J_0)\cap J_0=\emptyset$ and therefore if $J_0\subset(\mathcal{R}^j)^c$, we have also $J_0\subset(\mathcal{R}^{j+1})^c$. Since $J_0\subset(\mathcal{R}^0)^c=\mathcal{X}$, one has with probability larger than $1-\alpha$ that $J_0\subset\lim_j\mathcal{R}^j$, which means that the procedure derived by the step-down algorithm is controlling the family-wise error rate at level α.

REFERENCES

Blanchard, G., S. Delattre, and E. Roquain (2014). Testing over a continuum of null hypotheses with False Discovery Rate control. *Bernoulli 20*(1), 304–333.

Chan, H. and N. Zhang (2007). Scan statistics with weighted observations. *Journal of the American Statistical Association 102*(478), 595–602.

Fromont, M., B. Laurent, and P. Reynaud-Bouret (2013). The two-sample problem for Poisson processes: adaptive tests with a non-asymptotic wild bootstrap approach. *Annals of Statistics 41*(3), 1431–1461.

Fromont, M., M. Lerasle, and P. Reynaud-Bouret (2016). Family wise separation rate for multiple testing. *to appear in the Annals of Statistics*.

Goeman, J. J. and A. Solari (2010). The sequential rejection principle of familywise error control. *The Annals of Statistics 38*(6), 3782–3810.

Gretton, A., K. Borgwardt, M. Rasch, B. Schoelkopf, and A. Smola (2012). A kernel two-sample test. *Journal of Machine Learning Research 13*, 723–773.

Romano, J. P. and M. Wolf (2005a). Exact and approximate stepdown methods for multiple hypothesis testing. *J. Amer. Statist. Assoc. 100*(469), 94–108.

Romano, J. P. and M. Wolf (2005b). Exact and approximate stepdown methods for multiple hypothesis testing. *J. Amer. Statist. Assoc. 100*(469), 94–108.

Siegmund, D. O., N. R. Zhang, and B. Yakir (2011). False discovery rate for scanning statistics. *Biometrika 98*(4), 979–985.